

GLI

OPEN

DATA

IN AMBITO PARLAMENTARE



Nexa Center
for Internet & Society



Gli Open Data in ambito parlamentare

Report

Versione 1.0 Beta

28 Maggio 2015

Hanno collaborato per il Centro Nexa: Mauro Alovisio, Claudio Artusio, Eleonora Bassi, Lorenzo Benussi, Giulio De Petra, Raimondo Iemma, Stefano Leucci, Caterina Lupo, Federico Morando.

Hanno collaborato per il Senato della Repubblica: Mauro Fioroni, Carlo Marchetti, Giovanni Lalle, Roberta Rapaccini.

Hanno collaborato per la Camera dei deputati: Marco D'Orta, Elena Flavia Candia, Massimiliano Del Ferraro.

Redazione a cura di: Claudio Artusio, Giulio De Petra, Raimondo Iemma e Federico Morando (per il Centro Nexa), Carlo Marchetti (per il Senato della Repubblica), Elena Flavia Candia (per la Camera dei deputati).

Hanno partecipato alle audizioni informali (in ordine di audizione e riportando la sola affiliazione relativa all'invito a partecipare all'audizione):

Francesco Marcelli e Simone Bonanni (Servizio Studi del Senato della Repubblica), Annibale Ferrari (Servizio Studi della Camera dei deputati), Vittorio Alvino (Associazione Openpolis), Aldo Gangemi e Giorgia Lodi (CNR-ISTC), Guido Vetere (Associazione Senso Comune), Andrea Marchetti, Oreste Signore e Fabio Valsecchi (W3C Italia), Matteo Brunati (ePSIPlatform e Spaghetti Open Data), Monica Palmirani (Spaghetti Open Data e CIRSIFID), Marco Fossati (Associazione DBPedia e Spaghetti Open Data), Maurizio Napolitano (Open Data Institute – nodo italiano e Spaghetti Open Data), Alessandra Donnini e Stefano Penge (Istituto Italiano Open Data).

Il Capitolo I del presente Report incorpora materiali tratti da “Il Libro bianco per il riutilizzo dell'informazione del settore pubblico”, versione 1.1, disponibile con licenza Creative Commons Attribuzione 3.0 Italia e redatto a cura di Federico Morando, Raimondo Iemma e Claudio Artusio, ed aggiornato a cura di Claudio Artusio e Alberto Segà. Tra gli autori del Libro bianco (oltre ad alcuni dei co-autori del presente report) si ricordano anche: Juan Carlos De Martin, Alessandro Mantelero, Marco Ricolfi, Angelo Maria Rovati, Margherita Salvadori e Cristiana Sappa. Il Libro bianco contiene sezioni liberamente ispirate all'Open Data Handbook (<http://opendatahandbook.org/en/>), a sua volta pubblicato come contenuto aperto a cura della Open Knowledge Foundation.



Il Report “Gli Open Data in ambito parlamentare” è distribuito con Licenza [Creative Commons Attribuzione 4.0 Internazionale](https://creativecommons.org/licenses/by/4.0/).

L'opera è disponibile all'indirizzo <http://nexa.polito.it/open-data-parlamentari>.

Gli Open Data in ambito parlamentare – indice generale

Gli Open Data in ambito parlamentare.....	1
Premessa.....	1
Cap. 1. Dati aperti: movente, opportunità e metodo.....	5
1.1 Obiettivi tradizionali.....	5
1.1.1 I dati aperti come opportunità.....	5
1.1.2 Dati aperti: una definizione.....	6
1.1.3 Quanto valgono i dati aperti?.....	8
1.2 ...e nuove opportunità.....	9
1.2.1 I dati aperti come piattaforma di innovazione tecnologica e sociale.....	9
1.2.2 Alcuni esempi concreti.....	9
1.2.3 Il ruolo della società civile.....	11
1.3. Il contesto normativo.....	13
1.3.1 In Europa.....	13
1.3.2 In Italia.....	15
1.3.3 A livello delle autonomie locali.....	17
1.4 La pratica dell'apertura dei dati: strumenti e processi.....	18
1.4.1 Fare open data nella pratica.....	18
1.4.2 Costi, profitti e tariffe.....	21
1.4.3 Le possibilità di riutilizzo e la tutela dei dati personali.....	23
1.4.4 Formati e formalismi.....	27
1.4.5 Piattaforme Open Data.....	34
1.4.6 Licenziare i dati.....	37
Cap. 2. L'offerta di dati aperti parlamentari.....	41
2.1 Cosa si intende per Open Data parlamentari.....	41
2.2 Lo scenario italiano.....	41
2.3 La roadmap verso i Linked Open Data parlamentari.....	42
2.3.1 La scelta tecnologica, l'ontologia OCD e dati.camera.it.....	43
2.3.2 I Linked Open Data in ambito Intranet con PAD e il progetto Senato.....	44
2.3.3 L'estensione della ontologia OCD, l'ontologia OSR e dati.senato.it.....	44
2.3.4 Gli Open Data su votazioni e dati bibliografici/documentali.....	44
2.3.5 Il riutilizzo degli Open Data parlamentari: l'Hackathon Montecitorio 2014.....	45
2.4 La Camera dei deputati.....	46
2.4.1 Introduzione.....	46
2.4.2 L'ontologia OCD Ontologia Camera deputati.....	47
2.4.3 Il Portale Open Data dati.camera.it.....	49
2.4.4 Le licenze d'uso.....	57
2.5 Il Senato della Repubblica.....	57
2.5.1 Introduzione.....	57
2.5.2 I dati pubblicati.....	58
2.5.3 La licenza d'uso.....	60
2.5.4 Il portale.....	61
2.5.5 Gli atti parlamentari in formato aperto.....	63
2.5.6 Evoluzioni future.....	64
2.6 I dati aperti sulle norme.....	64
2.6.1 Le leggi: dato aperto per eccellenza.....	64
2.6.2 La rete delle leggi.....	65
2.6.3 Gli standard di identificazione e rappresentazione degli atti normativi.....	65
2.6.4 L'accesso agli atti normativi.....	69
2.6.5 Open data e Linked Open Data normativi.....	70

2.7 Un confronto internazionale.....	72
2.7.1 Nota metodologica sull'analisi dell'offerta dei parlamenti.....	72
2.7.2 Differenti modalità di pubblicazione dei dati.....	73
2.7.3 Lo status giuridico dei dati pubblicati: un'esigenza di certezza.....	74
2.7.4 Reperibilità e fruibilità dei dati.....	75
2.7.5 Buone pratiche.....	75
2.7.6 Approfondimenti utili.....	77
2.7.8 Tabella riassuntiva.....	77
Cap. 3. La domanda di dati aperti parlamentari.....	81
3.1 Il riutilizzo di dati parlamentari: opportunità e aspetti critici.....	81
3.2 “Open Community Data”.....	81
3.3 Le priorità per i riutilizzatori specializzati in dati parlamentari.....	82
3.4 Le indicazioni di ricercatori e "communities".....	83
3.4.1 Interoperabilità semantica.....	83
3.4.2 Formati e qualità dei dati.....	84
3.4.3 Apertura di ulteriori tipologie di dati.....	86
3.5 La domanda di dati (aperti) da parte del Parlamento.....	86
3.5.1 Le attività dei Servizi di documentazione parlamentari.....	87
3.5.2 L'approntamento della documentazione al servizio delle attività parlamentari.....	88
Cap. 4. A way forward.....	92
Annex.....	95
Glossario.....	95

Premessa

La possibilità di accedere e di **riutilizzare il patrimonio informativo prodotto e gestito dal settore pubblico** è diventata uno degli obiettivi fondamentali dell'agenda digitale italiana, così come di quella europea.

I termini “Open Data” o “Dati Aperti” sono ormai largamente utilizzati nella formulazione delle politiche di innovazione digitale, per descrivere una moltitudine di iniziative, progetti, aspettative che coinvolgono non più soltanto gli addetti ai lavori, ma anche settori sempre più vasti di cittadinanza attiva.

Fa ormai parte del senso comune che i dati prodotti o raccolti per finalità pubblica in qualsiasi ambito e da chiunque - siano organizzazioni pubbliche o private, siano società o singoli individui - debbano essere **resi disponibili all'accesso, alla condivisione e al riutilizzo in modalità gratuita e senza vincoli** di accesso a tutti i possibili interessati attraverso la rete Internet.

Libero accesso, ma anche libero riutilizzo.

È in questo contesto che viene promossa la ricerca che ha prodotto il presente Rapporto e che ha come oggetto la particolare declinazione delle politiche Open Data nei due rami del Parlamento italiano: il Senato e la Camera. L'obiettivo non è soltanto quello di censire e analizzare le specifiche pratiche di apertura dei dati in proprio possesso da parte delle amministrazioni del Senato e della Camera, ma soprattutto **valorizzare la specifica valenza che il tema Open Data assume in ambito parlamentare.**

In primo luogo per il **valore esemplare che la qualità delle soluzioni realizzate dagli uffici parlamentari può assumere per le amministrazioni pubbliche** che sono attualmente impegnate nella progettazione o nella concreta attuazione di progetti di apertura del proprio patrimonio informativo, e che si confrontano con le oggettive difficoltà che questi progetti incontrano dal punto di vista organizzativo, normativo o tecnologico.

In secondo luogo per il valore particolare che una larga disponibilità di informazioni assume proprio per l'attività legislativa, che costituisce l'attività peculiare dell'istituzione parlamentare. Questo vale non solo per ciò che tradizionalmente avviene all'interno del parlamento, ma si estende oggi alle nuove forme di coinvolgimento dei cittadini fuori dal parlamento, rese possibili proprio grazie alla disponibilità e alla diffusione delle tecnologie digitali. È infatti evidente che la **materia prima indispensabile per ogni esperienza di partecipazione civica**, anche e soprattutto per quelle che si realizzano su piattaforme digitali, sia una larga, adeguata ed efficace disponibilità di dati relativi ai temi oggetto di partecipazione.

In terzo luogo, infine, per la stretta relazione che oggi esiste tra i contenuti della produzione legislativa e la possibilità di rimuovere i vincoli e di adeguare le norme che consentono di favorire il processo di apertura del patrimonio informativo pubblico. A questo fine la speranza è che i contenuti di questo Rapporto possano **favorire una maggiore conoscenza del tema “open data” presso tutti coloro che operano in Parlamento, generando una adeguata consapevolezza degli effetti che specifici contenuti delle leggi approvate possono avere nel favorire, o nel ritardare, la piena accessibilità e riutilizzabilità del patrimonio informativo.** Questa considerazione riguarda certamente le leggi che disciplinano esplicitamente materie inerenti al digitale, ma anche quelle che, pur trattando materie diverse, possono comunque esercitare una influenza anche indiretta sui percorsi di “apertura” dei dati pubblici.

È anche a tale scopo che **il primo capitolo del Rapporto è dedicato a una descrizione sintetica, ma il più possibile aggiornata e completa, delle principali caratteristiche del vasto movimento a cui generalmente ci si riferisce con il termine “open data”.** Si tratta non solo di caratteristiche organizzative, tecnologiche, ma anche della vasta normativa di riferimento, sia

Il “movimento” Open Data.

italiana che europea. Una specifica attenzione è stata dedicata alle relazioni tra disponibilità di dati aperti e garanzie per la privacy. La descrizione, corredata da numerosi esempi concreti, non presuppone conoscenze specifiche da addetti ai lavori, ed in questo senso può essere considerata una lunga introduzione che mette il lettore in grado di comprendere meglio i successivi capitoli. Il primo capitolo si conclude infine con una sezione dedicata ad alcuni pratici suggerimenti metodologici per quelle amministrazioni che vogliono avviare effettive esperienze di apertura dei propri dati.

L'offerta di dati aperti da parte del Parlamento.

Il secondo capitolo del Rapporto è esplicitamente dedicato alla offerta di dati aperti da parte del Parlamento italiano, con una

accurata descrizione dei dati attualmente disponibili, delle tecnologie utilizzate, degli strumenti forniti per la loro interpretazione e per il loro riutilizzo. Gli uffici informatici dei due rami del Parlamento italiano hanno infatti avviato da tempo efficaci esperienze di apertura dei dati in loro possesso, iniziate quando ancora il paradigma "open data" non aveva la visibilità che lo contraddistingue attualmente, ma era praticato solo da pochi addetti ai lavori che univano alla competenza tecnica anche sensibilità istituzionale e lungimiranza professionale. Negli anni più recenti il tema "open data" è venuto assumendo anche rilevanza politica, ed è stato pertanto oggetto di un forte impulso istituzionale sia presso il Senato che presso la Camera. Esempio eclatante di questa forte attenzione politica è stato l'Hackathon sugli open data parlamentari che si è svolto nel maggio 2014 nei locali della Camera dei deputati, sicuramente il primo mai svolto in Italia e probabilmente il primo anche nel panorama dei parlamenti europei.

Questo forte impulso politico ha potuto produrre risultati concreti solo perché il contesto tecnico degli uffici parlamentari era già da tempo preparato e attivo su questi temi.

È questa una considerazione più generale, che deriva dal caso parlamentare, ma si estende a tutte le amministrazioni, centrali e locali, coinvolte in iniziative di apertura del proprio patrimonio informativo: senza forte consapevolezza e impulso da parte della direzione politica di ogni amministrazione, nessuna iniziativa nata all'interno dell'amministrazione, sia pure tecnicamente pregevole e avanzata, ha possibilità di tradursi in progetti in grado di coinvolgere sia l'amministrazione che la comunità degli utenti. Ma analogamente ogni impulso politico, sia pure il più consapevole ed esplicito, non ha possibilità di successo senza una adeguata competenza e sensibilità professionale degli uffici interni dell'amministrazione, che deve essere di conseguenza valorizzata e sostenuta. Il secondo capitolo si completa con una sezione dedicata a quello che può essere considerato il patrimonio informativo specifico dell'istituzione parlamentare: il corpo normativo prodotto nel tempo dalla attività parlamentare, che è da tempo oggetto di alcuni significativi progetti che hanno come obiettivo (e come risultato, già in parte raggiunto) di rendere disponibile l'intero patrimonio legislativo, sia nazionale che locale, ad una platea di utilizzatori non limitata ai soli addetti ai lavori. Al capitolo dedicato al Parlamento italiano segue una **analisi comparativa sulla situazione in termini di "open data" di numerosi parlamenti, sia europei che extraeuropei**, che descrive l'ampiezza e la varietà delle esperienze attualmente in corso e conferma la qualità della esperienza italiana.

La domanda di dati parlamentari aperti.

Il terzo capitolo è dedicato all'analisi della domanda di dati parlamentari da parte di quella crescente schiera di riutilizzatori, nella quale sono particolarmente attivi sia associazioni e imprese che hanno

come finalità della loro azione il monitoraggio della attività parlamentare e/o il sostegno alle iniziative di partecipazione civica, sia istituti di ricerca e comunità professionali che operano sul tema degli open data dal punto di vista tecnologico, organizzativo e metodologico.

Dal confronto con alcuni di questi soggetti sono emerse alcune importanti indicazioni, la più significativa delle quali riguarda l'utilità di una **relazione continua e strutturata tra riutilizzatori e produttori di open data pubblici**. È infatti dalla continuità di questa relazione che può derivare non solo una effettiva capacità di utilizzare correttamente ed efficacemente i dati resi disponibili, ma anche l'effetto positivo che il feedback da parte di chi utilizza i dati pubblici può

avere sulla qualità e la completezza dei dati gestiti dall'amministrazione, con il miglioramento quindi anche della efficienza dei sistemi informativi interni dell'amministrazione.

Aprire i propri dati diventa quindi per un'amministrazione pubblica non solo un obbligo, ma una preziosa opportunità di miglioramento della qualità dei propri servizi. Oltre a utili indicazioni su quali dati parlamentari potrebbero essere ulteriormente messi a disposizione del pubblico utilizzo, l'analisi della domanda mette anche in evidenza la necessità di prevedere una adeguata attività di formazione dell'utenza potenziale, sia riguardo agli aspetti tecnici del riutilizzo, sia riguardo alla natura e alle possibili interpretazioni dei dati grezzi.

Il terzo capitolo è arricchito dalla prospettiva del soggetto pubblico come riutilizzatore di dati aperti, soffermandosi in particolare sull'uso di Open Data da parte dei Servizi di Camera e Senato.

È proprio dal confronto tra domanda e offerta che derivano alcune delle indicazioni per il futuro che sono contenute nel **quarto capitolo del** [A way forward.](#)

Rapporto, il quale si propone di suggerire alcune concrete direzioni di lavoro, con l'obiettivo di proseguire, sviluppare e migliorare un'esperienza già largamente positiva. Tra queste indicazioni, quella che sembra emergere con più evidenza, come suggerito anche dal confronto con gli attuali utilizzatori degli open data parlamentari, è che ogni iniziativa di "open data" in ambito pubblico non può che essere un processo continuo e iterativo di collaborazione tra amministrazione, nel nostro caso l'istituzione parlamentare, e soggetti interessati al riutilizzo del patrimonio informativo pubblico. Un processo continuo che non può accontentarsi di misurare la quantità di dataset disponibili, ma deve aprirsi alla collaborazione e alla verifica di riutilizzatori da non considerarsi più esterni, ma integrati nel percorso stesso di valorizzazione del patrimonio informativo pubblico. È anche attraverso questa cooperazione tra istituzione e associazioni, cittadini e imprese che il tema degli open data può fornire un suo contributo peculiare e concreto al rinnovamento delle istituzioni e delle forme della politica.

Il testo del presente Rapporto è organizzato in capitoli e paragrafi, per ognuno dei quali viene fornito, all'inizio, un breve riassunto in forma di "glossa". Ove possibile, per i concetti più generali, non si è provveduto a fornire una descrizione ad hoc, rimandando a, o incorporando, pagine già esistenti nella versione italiana di Wikipedia o facendo riferimento ad altri materiali informativi disponibili online. [Guida alla lettura.](#)

Per la maggior parte, i contenuti qui presentati sono originali, e derivano da attività di studio, analisi, e confronto tra esperti. Si segnala che il primo capitolo "Dati aperti: movente, opportunità e metodo", ha l'obiettivo consentire di avvicinarsi al mondo degli open data da un punto di vista generale anche a chi non possiede competenze in merito. Questo capitolo fa riferimento, aggiornandoli e ampliandoli, ad alcuni dei contenuti del "Libro Bianco per il riutilizzo dell'informazione del settore pubblico" prodotto nell'ambito del progetto EVPSI. Si consiglia dunque al lettore che già conoscesse il tema degli open data di scorrere i riassunti dei singoli paragrafi per individuare eventuali approfondimenti di suo interesse, tenendo presente che probabilmente dispone già degli strumenti per affrontare proficuamente la lettura dei capitoli seguenti, incentrati sull'offerta e sulla domanda di dati aperti in ambito parlamentare.

Cap. I. Dati aperti: movente, opportunità e metodo

1.1 Obiettivi tradizionali...

1.1.1 I dati aperti come opportunità

Viviamo oggi nella società dell'informazione, in cui la conoscenza e la comunicazione sono risorse e fattori strategici. Chiunque è potenziale produttore di informazione (e quindi di valore), tanto in ambito privato quanto in ambito pubblico. D'altro canto, chiunque, potrebbe, e – nella logica egualitaria e cooperativa dell'“open” – dovrebbe essere messo in condizione di usufruirne. Non basta, inoltre, mettere a disposizione sul Web le informazioni perché se ne possa prendere visione: per sfruttare appieno le possibilità offerte dalle tecnologie digitali occorre che queste informazioni siano rese disponibili come dati strutturati e liberamente riutilizzabili, affinché costituiscano un'opportunità di partecipazione e crescita per l'intera società. Questo è il principio sotteso agli Open Data.

In questo scenario diventa cruciale la posizione delle Pubbliche Amministrazioni (PA) e delle istituzioni, le quali - tutte - **accumulano dati**. Li accumulano per poter funzionare e per dimostrare di aver funzionato correttamente. Negli ultimi decenni, tuttavia, qualcosa è cambiato nella natura di questa prodigiosa mole di informazione: si tratta, sempre più spesso, di informazioni in formato digitale (e, come vedremo, questa peculiare caratteristica determina conseguenze dirompenti, perché rende i dati più facilmente distribuibili e riutilizzabili).

A volte tali informazioni sono generate direttamente come risultato primario dell'attività di un'amministrazione: pensiamo ai dati cartografici, al catasto, alle informazioni meteorologiche, o all'archivio delle emittenti radio-televisive pubbliche. In altri casi, invece, i dati sono acquisiti, organizzati e impiegati dall'amministrazione nell'adempimento dei propri obiettivi istituzionali, ma rappresentano un mezzo rispetto all'output principale – prodotto o servizio che sia. Questo è il caso, per esempio, della base di dati dei tempi medi di percorrenza dei mezzi pubblici in una grande città o della mappa geografica dei codici di avviamento postale.

Tali dati – espliciti e visibili nella prima tipologia di casi sopra delineata (catasto, meteo), e perlopiù invisibili al pubblico nella seconda (perché ad uso interno) – sono nel loro complesso denominati “**informazioni del settore pubblico**” o, in inglese, “**public sector information**” (da cui l'acronimo “**PSI**”). Si noti che la PSI viene generata o raccolta dalle pubbliche amministrazioni in senso stretto, nello svolgimento dei loro compiti istituzionali, ma anche da molti enti pubblici o finanziati da risorse pubbliche. Considerando che, in Italia, le entrate dello Stato sfiorano il 50% del PIL, è ragionevole stimare che anche la proporzione di informazione pubblica rispetto al totale dell'informazione rappresenti una percentuale considerevole.

Le potenzialità dell'Informazione del Settore Pubblico (PSI).

Bus Trento

Un esempio può introdurre al tema dati aperti meglio di molte parole. L'app [Bus Trento](#) è stata realizzata grazie ai dati rilasciati dalla Provincia Autonoma di Trento e da Trentino Trasporti, l'azienda dei trasporti locali.

Il processo di liberazione del dataset non è stato semplice in quanto il dato era detenuto in compartecipazione dai due enti. Una volta risolta la problematica giuridica tramite un accordo, il dato ha seguito la normale procedura interna per il controllo giuridico e tecnico. Sistemati questi accorgimenti ed applicata la licenza, il dato è stato pubblicato e metadatatato nel Catalogo dei dati aperti del Trentino. Da qui, il riutilizzatore del dato ha scaricato e rimodellato il dataset per renderlo più utile ai suoi fini. Il dato, espresso in formato sostanzialmente testuale secondo lo standard *de facto* GTFS, definisce le linee degli autobus e delle funivie

locali, gli orari, e i calendari. Alla mera visualizzazione del dato, l'App ha aggiunto altri servizi, come il monitoraggio del bus preferito, gli hotel, i ristoranti nelle vicinanze e i parcheggi: i dati relativi a questi servizi sono ugualmente disponibili nel Catalogo dei dati aperti del Trentino.

Tutti questi dati, detenuti dal settore pubblico, hanno anche un valore particolare, che normalmente manca ai dati raccolti dal settore privato: si tratta di dati raccolti nell'arco di molti anni e di regola sottoposti a verifiche (ad esempio nel corso dell'attività che essi supportano), ma soprattutto si tratta di dati di interesse generale per la collettività e di dati tendenzialmente "neutrali" o comunque relativamente affidabili, in quanto meno soggetti a distorsioni generate da interessi particolari o commerciali diretti.

Spesso, la singola informazione presenta un'utilità limitata, ma la **raccolta organizzata** e la **combinazione di più informazioni** provenienti da fonti differenti è in grado di generare un valore aggiunto notevole. Così, 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. 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.

Come si può immaginare, la PSI è varia quanto e più delle stesse pubbliche amministrazioni che la producono e detengono: questi dati vanno dalle immatricolazioni delle automobili agli orari delle farmacie, dai piani regolatori alle mappe della protezione civile, dalle rilevazioni meteo dell'Aeronautica Militare agli orari di passaggio in fermata dei mezzi pubblici. Ovviamente, esiste una parte di dati pubblici che devono rimanere riservati o addirittura segreti (per ragioni di sicurezza nazionale o di privacy, ad esempio); non pochi dati, però, sono già teoricamente accessibili a chiunque e moltissimi altri potrebbero esserlo. Aprire effettivamente questi dati e chiarire che essi sono liberamente riutilizzabili da chiunque è l'obiettivo dell'approccio (e del "movimento") Open (Government) Data.

1.1.2 Dati aperti: una definizione

Con il termine [dati aperti](#) (Open Data) si fa riferimento ad un approccio o "movimento" volto a promuovere l'accessibilità e il libero **riutilizzo** dei dati; in particolare, di quei dati che sono creati e detenuti dagli enti pubblici e dalla pubblica amministrazione nel suo complesso.

Open Definition: la definizione di dati e contenuti aperti, accettata come standard a livello internazionale.

La **definizione standard** oggi generalmente riconosciuta, che si è largamente affermata presso le comunità promotrici del modello Open Data, è quella contenuta nell'[Open Definition](#), aggiornata a Settembre 2014, secondo la quale sono aperti quei dati "a cui è liberamente possibile accedere, e che possono essere liberamente utilizzati, modificati e

ridistribuiti da chiunque, per qualunque scopo", rispettando, al più, requisiti che preservino l'origine del dato e la sua apertura.

Da questa definizione discende che i dati sono aperti solo se si applica una delle seguenti condizioni: sono automaticamente liberi da qualsiasi tipo di diritto di proprietà intellettuale (ovvero sono in "pubblico dominio" per legge); vengono messi a disposizione sotto licenze/liberatorie, che escludono qualsiasi limitazione (ovvero vengono posti contrattualmente in una situazione equivalente al pubblico dominio); vengono licenziati a condizioni molto permissive, che richiedano solo l'attribuzione/citazione della fonte; eventualmente, vengono sottoposti a clausole c.d. share-alike (condividi-allo-stesso-modo, vedi p. 38). Per ulteriori approfondimenti circa il licenziamento dei dati, rimandiamo al paragrafo Licenziare i dati (vedi sotto, p. 37).

La Direttiva 2013/37/UE definisce il riutilizzo come *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.*

Parola chiave: Riutilizzo

Chi si occupa di informazione del settore pubblico parla spesso di “riutilizzo” (o “riuso”), anziché di “utilizzo” (o “uso”). La scelta del termine riutilizzo sottolinea il fatto che stiamo parlando di **usi diversi e ulteriori, rispetto all'uso istituzionale, per cui il dato è stato raccolto o generato dalla PA.**

Il riutilizzo va anche chiaramente distinto dal mero accesso. Ciò che interessa non è solo la possibilità di accedere ai dati; approcci tipo “guardare, ma non toccare”, infatti, non facilitano la vita degli sviluppatori e la creazione di servizi utili ai cittadini. Poter vedere i dati è solo il primo passo per poterli acquisire e poi finalmente riutilizzare, cioè modificare, mescolare e/o trasformare per renderli più utili ed interessanti (per alcuni specifici scopi, senza togliere ad altri la possibilità di fare altrettanto, in modo diverso, per i loro differenti obiettivi).

In Italia, alla luce della definizione di “dati di tipo aperto” fornita dall'**art. 68 del Codice dell'Amministrazione Digitale (CAD)**, sono oggi definiti come tali i dati accessibili e disponibili gratuitamente (o comunque non oltre i costi marginali di riproduzione e diffusione) attraverso le tecnologie dell'informazione e della comunicazione in formati aperti, ivi comprese le reti telematiche pubbliche e private, quando sussiste una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali.

Dati aperti secondo la legge italiana: in formati aperti e riutilizzabili senza ulteriori restrizioni diverse dall'obbligo di citare la fonte e di rispettarne l'integrità.

L'art. 7 del c.d. Decreto Trasparenza (richiamando a sua volta l'art. 68 del CAD) stabilisce che i dati oggetto di pubblicazione obbligatoria ai sensi della normativa vigente siano riutilizzabili senza ulteriori restrizioni diverse dall'obbligo di citare la fonte e di rispettarne l'integrità. Maggiori dettagli sulle normative citate sono reperibili all'interno del paragrafo Contesto normativo (p. 13 e ss.).

Dal momento che, come accennato sopra, **l'effettivo valore degli Open Data si sprigiona nel momento in cui dati di diversa provenienza - e quindi messi a disposizione da soggetti differenti - vengono combinati tra loro**, affinché sia possibile effettuare tale combinazione è necessario che i dati presentino caratteristiche tali da consentire l'**interoperabilità**. Con questo termine si esprime la capacità di diversi sistemi e organizzazioni di lavorare insieme (appunto, **Inter-Operare**) e che nel contesto di cui ci stiamo occupando può riassumersi nella capacità di combinare una base di dati con una o più basi di dati differenti.

L'interoperabilità può declinarsi secondo diverse direttrici, ognuna delle quali tesa a rimuovere gli ostacoli che impediscono la piena ed efficace interazione e la fruizione delle differenti sorgenti informative. Da un lato,

Interoperabilità tecnica, semantica e giuridica

l'interoperabilità tecnica è l'insieme delle condizioni che consentono a sistemi diversi di scambiare automaticamente informazioni; la scelta di un **formato** adeguato per la pubblicazione dei dati è uno degli elementi che concorrono al raggiungimento di questo tipo di interoperabilità. Non meno importante è **l'interoperabilità semantica**, ovvero l'insieme delle misure adottate affinché sia adeguatamente esplicitato e rappresentato il significato dei dati; l'uso di **vocabolari e ontologie standard** concorre a questo obiettivo. D'altro canto, **l'interoperabilità giuridica** permette di manipolare i dati nei modi desiderati, senza incorrere (o temere di incorrere) in violazioni di diritti di proprietà intellettuale o altre norme; l'adozione di una opportuna **licenza di diritto d'autore** rientra tra le prassi volte a favorire l'interoperabilità giuridica. Torneremo su questi concetti in seguito, all'interno dei paragrafi Formati e formalismi (p. 27) e Licenziare i dati (p. 37).

L'attributo dell'interoperabilità è stato espressamente enucleato all'interno dell'art. 68 del CAD, laddove stabilisce le caratteristiche essenziali dei dati aperti e rimarca gli speculari requisiti di **apertura tecnica** (formati aperti e machine-readable) e **apertura giuridica** (licenze che ne consenta l'utilizzo da parte di chiunque, anche per finalità commerciali).

Come si sarà probabilmente già intuito, il punto chiave consiste nel fatto che **l'Open Data riguarda la messa a disposizione dei dati**, perché altri possano riusarli. Open Data, quindi, non significa spingere la pubblica amministrazione a sviluppare nuovi servizi a partire dai dati. Il servizio che la pubblica amministrazione deve sviluppare è “semplicemente” un modo pratico ed efficace per pubblicare i dati in proprio possesso, preceduto da un razionale processo per scegliere quali dati possano essere messi a disposizione di tutti e corredato da una chiara affermazione relativa al fatto che il riutilizzo dei dati è libero. Il raggiungimento di questi scopi richiede una buona capacità di modellazione del proprio dominio informativo che consenta di esplicitare al meglio la **semantica dei dati** in esso presenti (vedi, in merito, p. 27 e ss.). I **vocabolari comuni**, le **ontologie condivise** e costruite utilizzando metodologie definite nonché il ricorso a schemi affermati aiutano ad ottenere quel livello di qualità del dato tale da produrre una loro efficace condivisione che consenta di realizzare servizi di qualità.

I molteplici dataset prodotti da differenti organizzazioni, la presenza di ontologie che ne descrivono adeguatamente la semantica, insieme con l'assunzione di mondo aperto (OWA – Open World Assumption), permette di realizzare dei servizi che sono di ausilio alla scoperta di conoscenza non esplicitamente rappresentata all'interno dei singoli dataset.

L'esperienza mostra che i cittadini riceveranno nuovi servizi, sia gratuiti che a pagamento; ma saranno i singoli sviluppatori, gli imprenditori, le aziende e la società civile a produrre questi servizi. E il processo di apertura dei dati avrà successo se la pubblica amministrazione potrà concentrarsi a far meglio le attività essenziali che è suo compito istituzionale svolgere.

La PSI può essere riutilizzata in molte prospettive:

1. nell'ambito di mercati già esistenti o nuovi, per creare **servizi a valore aggiunto**, con benefici di tipo economico. 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;
2. per contribuire ad un **miglioramento delle scelte pubbliche**, in un quadro di [Open Government](#) e trasparenza. Ad esempio, il semplice fatto di mettere a disposizione e permettere la rielaborazione di dati quali il numero, il salario o i giorni di assenza degli amministratori e dei dipendenti pubblici può stimolare l'efficienza della PA. In altre parole, condividendo informazioni su sé stessa, la PA è spinta a migliorarsi;
3. per consentire forme più sofisticate di **partecipazione dei cittadini** alle scelte pubbliche (*e-democracy*). Ad esempio, la combinazione di dati sanitari e ambientali può permettere a gruppi di persone di valutare l'impatto di diverse scelte politiche e può avere ricadute sulla partecipazione dei cittadini alla vita pubblica. Inoltre, la partecipazione della popolazione può anche passare per la creazione di servizi, simili a quelli che potrebbe offrire lo stato o il mercato, ma su base volontaria e/o associativa.

1.1.3 Quanto valgono i dati aperti?

Il valore degli open data? Difficile da stimare, ma considerevole e ben superiore al costo di metterli a disposizione!

A livello macroeconomico, gli studi internazionali compiuti in questo campo concordano nell'affermare che il **valore indotto da dati pubblici riutilizzabili**, per un paese simile all'Italia, sia **nell'ordine delle centinaia di milioni di Euro**. Di contro, il costo dell'apertura dei dati è minore per diversi ordini di grandezza. Si può stimare il costo di realizzare un progetto e un portale per la condivisione di dati aperti tra qualche decina ad alcune

centinaia di migliaia di Euro (con 200-300.000 Euro si può far partire un rispettabile progetto di livello nazionale; sul finire del 2011 la Commissione Europea ha aperto un bando per realizzare il suo portale open data a un costo massimo di 800.000 Euro). Esistono poi *communities* online, enti

non governativi e centri di ricerca ben disposti ad aiutare chi voglia aprire i propri dati; inoltre, molte amministrazioni hanno già intrapreso il processo di apertura dei propri dati, il che offre l'opportunità di emulare esperienze e soluzioni esistenti. Con determinazione e passione, è addirittura possibile avviare un progetto open data a costo praticamente nullo (in termini di specifici stanziamenti di budget e abbastanza contenuto anche in termini di ore/uomo).

Vi sono buone ragioni per non lasciarsi sfuggire le opportunità offerte dal riutilizzo dell'informazione del settore pubblico (non ultima è la circostanza che i cittadini hanno già pagato, attraverso il prelievo fiscale, per ottenere quello stesso servizio mediante il quale si seleziona, raccoglie o genera il dato). 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 riusare 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.

1.2 ...e nuove opportunità

1.2.1 I dati aperti come piattaforma di innovazione tecnologica e sociale

Se l'e-government si propone di generare miglioramenti per i cittadini tramite l'introduzione di strumenti digitali, spesso informatizzando i tradizionali schemi di interazione tra amministrazioni e cittadini (ed il c.d. government 2.0 innova questo schema sottolineando il ruolo centrale dell'interazione coi cittadini/utenti e degli strumenti di social networking), si può dire che l'Open Government riguarda meno la tecnologia e molto di più il **rapporto tra governanti e governati**, amministratori e cittadini. In particolare, anziché focalizzarsi sull'offerta di specifici servizi, si mettono i cittadini in grado di capire e partecipare, dotandosi essi stessi di nuovi strumenti. Segnatamente al tema open (government) data (che non coincide con l'open government, ma che ne è un aspetto caratterizzante), la rivoluzione copernicana consiste nel fornire al cittadino non un nuovo mezzo tecnologico per accedere ai servizi pubblici, ma i **dati necessari a supportare, affiancare o criticare l'operato dell'amministrazione pubblica** stessa. Al limite anche fornendo un servizio concorrente, se – con gli stessi dati – qualcuno pensa di poter far meglio della pubblica amministrazione (a scanso di equivoci, in un caso del genere, non è detto che la pubblica amministrazione debba far venir meno il proprio servizio — sono molti i campi in cui la coesistenza di un servizio pubblico e di più servizi privati, alcuni dei quali imprenditoriali e altri non-profit, risulta socialmente desiderabile, poiché ciascuno ha i propri pregi e i propri difetti, spesso mitigati dalla presenza di alternative. Ad esempio, un'azienda municipalizzata vorrà mettere a disposizione sul proprio sito un servizio accessibile via Web per calcolare percorsi urbani coi mezzi pubblici, ma sarà opportuno facilitare il lavoro di chi voglia realizzare un'App per l'iPhone o per telefoni Android che faccia la stessa cosa, in modo ottimizzato per queste piattaforme software).

Open Government Data

1.2.2 Alcuni esempi concreti

I dati detenuti dalle amministrazioni possono, molto semplicemente, essere **utili in modo diretto** ai cittadini, se qualcuno si fa carico di renderli più accessibili e user friendly: ad esempio, una donna danese ha realizzato <http://findtoilet.dk>, un

Esempi di riutilizzo dei dati

servizio per individuare bagni pubblici, prendendo spunto dalle necessità di suoi conoscenti con problemi di incontinenza.

Un campo complesso è quello della **trasparenza**. Un possibile scopo della trasparenza è il controllo dell'operato della PA. Maggiore trasparenza, però, può anche voler dire maggior coinvolgimento e maggiore partecipazione: sapere dove finiscono i soldi delle mie tasse, per esempio, può renderle il pagarle più accettabile. In questo senso, è interessante il progetto britannico <http://wheredoesmymoneygo.org> (“dove finiscono i miei soldi”). Secondo gli sviluppatori del servizio, molti utilizzatori sono rimasti favorevolmente sorpresi dall'importanza della macro voce di spesa “aiutare gli altri”, che raccoglie ad esempio le spese di assistenza ai soggetti più deboli della società. Allo stesso modo, alcuni sono rimasti stupiti dalla relativa esiguità di voci di spesa spesso criticate, come quelle militari.

Naturalmente, mescolando più fonti di informazioni differenti si possono aiutare i cittadini a prendere **decisioni più complesse**. Ad esempio, Mapumental (relativo al Regno Unito) e [Mapnificent](#) (creato in Germania, ma che fa uso di dati da tutto il mondo) aiutano i cittadini a scegliere l'area in cui stabilirsi, considerando il tempo necessario a raggiungere il proprio luogo di lavoro con i mezzi pubblici, il prezzo delle case e altri indicatori relativi alla qualità della vita nella zona in questione. Un servizio analogo è disponibile anche a Torino: [Scegli l'isolaTO](#).

Mapumental

[Mapumental](#) è una applicazione Web che offre agli utenti una mappa interattiva, combinando tra loro dati geografici, dati catastali, dati dei servizi di trasporto pubblico e dati generati dagli utenti del Web.

Sviluppato da [mySociety](#) congiuntamente a [Channel 4's 4IP](#), questo strumento permette di individuare il luogo ideale della Gran Bretagna in cui stabilirsi, in base alle proprie esigenze di tempo negli spostamenti, possibilità economiche e aspettative circa la gradevolezza urbanistica (“scenograficità”).

Selezionando un codice di avviamento postale corrispondente ad una destinazione di riferimento (ad esempio, il proprio luogo di lavoro abituale), e scorrendo le barre relative al tempo che si è disposti ad impiegare per raggiungerla entro le ore 9:00, al costo degli immobili e alla scenograficità del quartiere, è possibile circoscrivere e visualizzare le zone corrispondenti ai parametri scelti, individuando la propria dimora ideale.

Mapumental si basa, incrociandoli tra loro, su una serie di dati di provenienza diversa. Le mappe di sfondo sono generate sfruttando [OpenStreetMap](#), che a sua volta riusa informazioni geografiche pubbliche, oltre a quelle generate dagli utenti. I dati del trasporto pubblico provengono dalla banca dati National Public Transport Data Repository ([NPTDR](#)). I prezzi degli immobili sono stati ricavati acquistando una lista di tutte le vendite di proprietà immobiliari ad uso abitativo registrate presso il Land Registry di Inghilterra e Galles tra gennaio 2008 e marzo 2009 (i dati relativi alla Scozia non sono stati inclusi perché troppo costosi). La “scenograficità” è stata stimata sfruttando le valutazioni sulla gradevolezza estetica di diverse aree dell'Inghilterra, della Scozia e del Galles, espresse dagli utenti del sito [ScenicOrNot](#), creato appositamente per accumulare i dati necessari per Mapumental. Le foto presenti su ScenicOrNot provengono dal sito [Geograph](#).

Il “dovere” di condividere i dati nasce anche dal costo irrisorio dalla condivisione, grazie alle tecnologie digitali.

Le ragioni che rendono opportuno aprire e condividere i dati – proprio adesso – sono fondamentalmente legate a **sviluppi tecnologici**. Oggi, infatti, condividere dati digitali costa poco, quasi nulla. Ed elaborarli è facile e veloce. In questo scenario, diventano possibili – anche solo nella logica del “Perché no?” – cose che negli ultimi anni erano semplicemente infattibili o

così costose da essere proibitive. Un'argomentazione forte a favore dell'apertura dei dati pubblici è proprio quella del “**Perché no?**”. Perché, infatti, si dovrebbe sottrarre alla creatività, alla passione, allo spirito d'iniziativa, alla fantasia e alle tante altre risorse dei cittadini e degli imprenditori una ricca dote, che la PA ha comunque a propria disposizione. Perché no, se il costo è ridotto? Perché no, se l'esperienza insegna che gli utenti della Rete riescono a portare avanti progetti ambiziosi del calibro di un'enciclopedia online come Wikipedia? Infatti, **la tecnologia non ha solo ridotto i**

costi per distribuire e processare i dati, ma è uno strumento con cui sviluppare piattaforme adatte ad esercitare forme di collaborazione di massa, capaci di generare una sorta di intelligenza collettiva. Uno dei principi fondamentali della new economy, che sembra applicarsi bene anche all'approccio open data è, appunto, che "l'intelligenza è là fuori": ci sarà sempre (o quasi) qualcuno capace di usare i dati custoditi dalla PA in modi che queste non avrebbe potuto immaginare. E questo non perché la PA non abbia persone capaci e volenterose al proprio interno, ma semplicemente perché le persone all'esterno di ogni organizzazione – dal Comune di Moncenisio a Google – sono sempre molte di più, hanno interessi disparati e punti di vista differenti. E possono collaborare tra loro utilizzando le piattaforme più varie.

1.2.3 Il ruolo della società civile

1.2.3.1 Una rivoluzione digitale

La valorizzazione del patrimonio informativo pubblico, attraverso il modello open data, è il frutto di **condizioni socio-economiche, politiche e tecnologiche** tipicamente contemporanee che riguardano la digitalizzazione della realtà, fisica e sociale, e la pervasiva disintermediazione delle attività umane. Viviamo in sistemi socio-tecnici complessi, ambienti ibridi in cui territori fisici e digitali convergono per costituire una realtà specchio dove il digitale descrive il fisico fino a determinarne o meno l'esistenza. Evolve il rapporto tra pubblico e privato e la tecnologia abilita nuove forme di organizzazione di gruppi con interessi omogenei o convergenti. L'utilizzo, diffuso ormai da più di 20 anni, dei navigatori satellitari è un esempio semplice ma efficace di come la tecnologia stia diventando la porta del reale, lo strumento principale di navigazione, così pervasivo da determinare quasi l'esistenza o meno degli elementi fisici. Quando visitiamo una nuova città, ad esempio, se un ristorante, un museo o una piazza non esistono sulle mappe digitali scompaiono dal nostro itinerario, diventano razionalmente introvabili e in ultima istanza svaniscono. Allo stesso modo, un governo che non renda disponibili in rete informazioni (e servizi) utili scompare in parte, perché non è presente in una parte importante delle nostre vite, quella digitale. È perciò naturale per la società civile contemporanea richiedere con fermezza che le amministrazioni pubbliche, ma anche le organizzazioni private, abbiano una presenza in rete, si descrivano con la disponibilità dei loro dati e si mettano in discussione anche online, secondo una nuova dialettica. La richiesta di dati è espressione dell'esigenza di avere nuovi canali digitali, o meglio nuovi media digitali per conoscere la realtà, comprenderne le dinamiche e decidere come gestire la cosa pubblica. **I dati diventano elemento essenziale di cittadinanza** e la società civile non può che esserne il primo utente e beneficiario.

Un mondo di dati, che diventano elemento essenziale per la cittadinanza attiva.

1.2.3.2 Cittadini protagonisti

La società civile ha avuto ed ha un ruolo primario nell'idea e nella pratica del modello open data nel mondo, i singoli cittadini, le associazioni e le reti d'interesse contribuiscono ogni giorno a dare significato al processo in continua evoluzione di liberazione dei dati pubblici, costruendo progetti e valutando l'efficacia o meno delle soluzioni adottate dalle amministrazioni. L'attività di promozione svolta dell'[Open Knowledge Foundation](#) in Europa, che sviluppa strumenti e modelli di valorizzazione dei dati pubblici facendo lobby a livello nazionale e comunitario, o dal programma [Code For America](#) negli Stati Uniti, che aiuta direttamente le amministrazioni americane a sviluppare progetti open data, sono solo due esempi di come le **associazioni di cittadini** siano a tutti gli effetti protagoniste, anche più di governi e aziende, del movimento open data. Ancor più rilevante e caratteristico è l'apporto delle **reti di esperti**, ovvero delle mailing list e dei forum che aggregano competenze e buone pratiche in modo libero e informale e che hanno creato in tutte le nazioni un network di promotori delle tecnologie e dei processi open data. Molto interessante è il caso italiano della comunità [Spaghetti Open Data](#)

Il ruolo attivo dei cittadini

che è a tutti gli effetti uno dei centri nazionali più importanti per discutere e progettare iniziative open data; essa è diventata il nodo centrale di una rete che comprende semplici cittadini, imprenditori e amministratori pubblici. Merita un approfondimento, all'interno delle pratiche della società civile, il caso di [Monithon](#): progetto collegato all'iniziativa open data forse più famosa e meglio riuscita in Italia, [Open Coesione](#). L'idea di Monithon nasce da un gruppo di cittadini interessanti a rendere più efficace e partecipato il processo di messa a disposizione dei dati di [opencoesione.it](#) volto a raccogliere le informazioni sugli interventi finanziati con fondi europei. Il progetto colleziona delle piccole inchieste sulle attività finanziate nei vari territori, es. la costruzione di una strada, la creazione di un laboratorio, ecc. che comprendono delle visite ai siti, la raccolta di documenti, foto e filmati e la realizzazione di interviste con l'obiettivo di rendere i cittadini attivi. In questo caso possiamo dire che l'iniziativa della società civile diventa quasi più rilevante di quelle dell'amministrazione, o meglio, che le due si alimentano vicendevolmente, costruendo quel circolo virtuoso tra informazione e partecipazione che costituisce il vero valore del modello open data.

1.2.3.3 Verso una nuova dialettica

Conoscenza, collaborazione, partecipazione

In particolare, il ruolo della società civile si può comprendere a partire da due istanze chiare e molto forti: la prima è l'esigenza di **conoscere sempre meglio** - e con strumenti e linguaggi sempre più precisi e attendibili - i dati che riguardano il funzionamento dello stato, la seconda è la necessità di attivare **piattaforme di collaborazione e partecipazione** che coinvolgano i cittadini nella definizione dei problemi rilevanti e delle possibili soluzioni.

Informare il dibattito è la prima domanda che i cittadini rivolgono alle istituzioni quando chiedono i dati pubblici. Essi vogliono conoscere lo stato del sistema, i dati puntuali sulle politiche e l'evoluzione, possibilmente in tempo reale, degli interventi. Abituati alle metriche digitali - pensiamo ai siti e alle app sul meteo che forniscono una descrizione numerica dei fenomeni fisici - sempre più i cittadini chiedono una **nuova dialettica pubblica** fatta di numeri, prove, documenti. In questo senso, il ruolo della società civile è dunque quello di esprimere una nuova domanda che richiede un'offerta informativa tecnologicamente avanzata e capace di utilizzare interfacce adeguate all'esperienza dei cittadini con i più diffusi servizi digitali, dalle mappe alle applicazioni meteo agli strumenti di finanza personale.

Il secondo elemento da considerare nell'analisi dell'apporto della società civile alla diffusione del modello open data è la necessità dei cittadini di essere protagonisti, di **partecipare alle decisioni** pubbliche e alla gestione dei beni pubblici. Questo deriva da due cause endogene: l'emergere di un nuovo concetto di cittadinanza digitale e la naturale complessità dei sistemi socio-tecnici. L'esempio più chiaro è forse quello delle Comunità Intelligenti, le cosiddette Smart City. Da un lato il cittadino di una smart city vuole accedere ai dati del suo territorio, a quel cruscotto intelligente cui dovrebbero avere ormai accesso gli amministratori pubblici, ma che oggi può essere a disposizione di tutti. Lo chiede perché è necessario a costruire una sua nuova cittadinanza digitale ed è essenziale per rendere realmente produttive, o meglio per dare un senso alle tecnologie che stanno popolando le città. Dall'altro lato, condividere le informazioni e trovare delle piattaforme inclusive di discussione è necessario per migliorare la capacità di prendere decisioni in un ambiente complesso; collaborare e partecipare con i cittadini si rivela perciò una metodologia di lavoro necessaria e utile a limitare le possibilità di errori. I dati diventano uno strumento necessario per creare identità e la società civile ha il ruolo di protagonista perché formula le richieste e definisce le aspettative a cui le istituzioni sono tenute a rispondere. Quando un'amministrazione decide di adottare il modello open data, è necessario coinvolgere i cittadini o, meglio, considerare la società civile un partner di lavoro, un collaboratore indispensabile a indirizzare l'operato degli amministratori nel modo più efficace ed efficiente.

1.3. Il contesto normativo

Il contesto normativo in tema di apertura dei dati pubblici è oggi particolarmente ricco, a tutti i livelli normativi. Se, da un punto di vista strettamente giuridico, le norme risultanti possono a volte apparire ridondanti, da una prospettiva sociologica e organizzativa questa ridondanza è spesso utile a rassicurare i funzionari pubblici rispetto alla coerenza tra l'apertura dei dati e le politiche dell'amministrazione di cui sono al servizio.

1.3.1 In Europa

A livello comunitario, il tema dell'accesso e riutilizzo delle informazione del settore pubblico è stato oggetto di specifiche iniziative di studio e discussione già a partire dalla seconda metà degli anni novanta. In particolare, con la pubblicazione, nel 1999, del [Libro verde sull'informazione del settore pubblico nella società dell'informazione](#) della Commissione Europea - frutto di un processo di consultazione avviato nel 1996 - sono stati evidenziati per la prima volta, all'interno di un unico documento, i principali profili giuridici, economici e tecnici connessi alla fruizione di dati prodotti dal settore pubblico. La riflessione su tali profili è stata in seguito ulteriormente stimolata attraverso una [consultazione pubblica](#) aperta a tutti gli operatori interessati, volta ad ampliare il dibattito sulle questioni problematiche coinvolte (profili di diritto d'autore; tutela della sfera privata; politiche sul prezzo e relativo impatto su accessibilità ed uso dell'informazione pubblica; profili di concorrenza sleale connessi all'attività degli enti pubblici sul mercato dell'informazione; impiego di metadati per facilitare l'accesso all'informazione).

Il confronto politico generato dalle iniziative descritte ha portato, qualche anno più tardi, all'adozione della [Direttiva 2003/98/CE](#) del Parlamento europeo e del Consiglio, approvata il 17 novembre 2003 e pubblicata nella

Direttiva PSI: la norma europea di riferimento.

GUCE n. L 345 del 31 dicembre 2003 (c.d. **Direttiva PSI**), la quale rappresenta a tutt'oggi il testo normativo di riferimento in tema di riutilizzo dell'informazione del settore pubblico nell'Unione Europea. Questa Direttiva si pone l'obiettivo di agevolare la *“creazione di prodotti e servizi a contenuto informativo, basati su documenti del settore pubblico, estesi all'intera Comunità, nel promuovere un effettivo uso, oltre i confini nazionali, dei documenti del settore pubblico da parte delle imprese private, al fine di ricavarne prodotti e servizi a contenuto informativo a valore aggiunto e nel limitare le distorsioni della concorrenza sul mercato comunitario”* (Considerando #25).

Naturalmente, la Direttiva non si applica indiscriminatamente a ogni dato detenuto dalle pubbliche amministrazioni, evitando dunque di pregiudicare diritti di terzi, tutela della sicurezza nazionale, segreto statistico o tutela della privacy.

Segreto Statistico

Sono esclusi dall'applicazione della normativa sul riutilizzo i documenti connessi alla tutela del segreto statistico, quali disciplinati dal [Decreto Legislativo 6 settembre 1989, n. 322, Norme sul Sistema statistico nazionale e sulla riorganizzazione dell'Istituto nazionale di statistica](#), ai sensi dell'art. 24 della legge 23 agosto 1988, n. 400.

Si noti che, all'art 10, il sopracitato D.Lgs. 322/1989 chiarisce che *“[i] dati elaborati nell'ambito delle rilevazioni statistiche comprese nel programma statistico nazionale sono patrimonio della collettività e vengono distribuiti per fini di studio e di ricerca a coloro che li richiedono [...]”*. La normativa stessa, dunque, riconosce uno dei presupposti fondamentali dell'open data (il fatto cioè che i dati siano patrimonio della collettività), anche se limita lo scopo del riutilizzo ad alcuni fini specifici (studio e ricerca). Riguardo a tale limitazione di scopo, si potrebbe ritenere che la norma sia stata oggetto di abrogazione parziale implicita a seguito del D.Lgs. 36/2006, Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico, il quale, si ricorda, permette il riutilizzo dei dati per fini commerciali e non commerciali di qualsiasi tipo (salvo per quei documenti esclusi dall'applicazione del decreto 36/2006 stesso, il cui accesso sia inibito per motivi specifici come, ad esempio, per ragioni di tutela del segreto statistico). Giova anche considerare, del resto, come proprio l'Istituto nazionale di statistica abbia inaugurato, a partire dal 2011, una sezione Open Data,

rilasciata con licenze standard (Creative Commons, CC BY 3.0), permettendo quindi di riprodurre, distribuire, trasmettere e adattare liberamente i suoi dati e analisi anche a scopi commerciali, [a condizione che se ne citi la fonte](#).

Il D.Lgs. 322/1989 stabilisce anche alcune “Disposizioni per la tutela del segreto statistico” (art. 9), che sono state affiancate nel 2004 dal [Codice di deontologia e di buona condotta per i trattamenti di dati personali per scopi statistici e scientifici](#). L'opportunità di simili norme appare chiara, qualora si consideri l'esistenza di uno specifico obbligo a fornire, da parte di cittadini e imprese, alcune informazioni potenzialmente sensibili o tali da permettere la ricostruzione di informazioni sensibili, ad esempio in occasione di un censimento. Al fine di ottenere risposte veritiere da parte di cittadini e imprese, dunque, è ragionevole imporre limitazioni al riutilizzo delle informazioni stesse.

In particolare, “[i] dati raccolti nell'ambito di rilevazioni statistiche comprese nel programma statistico nazionale [...] non possono essere esternati se non in forma aggregata, in modo che non se ne possa trarre alcun riferimento relativamente a persone identificabili e possono essere utilizzati solo per scopi statistici.” (art. 9 comma 1, D.Lgs. 322/1989). Al fine di prevenire eventuali operazioni illecite di [data mining](#), la norma stabilisce anche che “[i]n ogni caso, i dati non possono essere utilizzati al fine di identificare nuovamente gli interessati.” D'altro canto, il segreto statistico non può essere arbitrariamente esteso e utilizzato come scusa: come stabilito dallo stesso articolo, infatti, “non rientrano tra i dati tutelati dal segreto statistico gli estremi identificativi di persone o di beni, o gli atti certificativi di rapporti, provenienti da pubblici registri, elenchi, atti o documenti conoscibili da chiunque” (art 9, comma 2, D.Lgs. 322/1989). Naturalmente, infine, l'aggregazione ha come obiettivo la tutela della riservatezza, non l'opacità o l'oscurità delle procedure seguite o del tipo di dati originariamente raccolti. Per questo, e sia pure solo in base ad una richiesta motivata, il medesimo decreto prevede che possano essere distribuite anche collezioni campionarie di dati elementari, opportunamente resi anonimi.

La Direttiva PSI del 2003 è stata **successivamente modificata** dalla Direttiva 2013/37/UE del Parlamento Europeo e del Consiglio, del 26 giugno 2013: come recita il suo Considerando #5, dal 2003 ad oggi si è assistito ad una crescita esponenziale della quantità di dati nel mondo e, soprattutto, ad un'evoluzione costante delle tecnologie per lo sfruttamento dei dati stessi. Un'evoluzione tale da rendere non più attuali, a distanza di dieci anni, le norme del 2003. Alla luce delle sue modifiche, la nuova direttiva dispone, come principio generale, che **gli Stati membri provvedano affinché tutti i documenti accessibili siano anche riutilizzabili, anche a fini commerciali**.

Direttiva 2013/37/UE

La [Direttiva 2013/37/UE](#) del Parlamento Europeo e del Consiglio, del 26 giugno 2013, apporta alcune modifiche alla Direttiva 2003/98/EC relativa al riutilizzo dell'informazione del settore pubblico.

Tra le principali novità, la Direttiva 2013/37/UE prescrive che:

- l'ambito di applicazione riguardi **tutti gli enti pubblici** di uno Stato membro, ivi compresi musei, archivi e biblioteche (anche universitarie);
- il principio generale di tariffazione, nel caso sia previsto un corrispettivo, sia quello del **costo marginale** sostenuto per la riproduzione, messa a disposizione e divulgazione dei dati. Per musei, archivi e biblioteche resta ancora in vigore la possibilità di generare un congruo utile sugli investimenti; simile deroga vale per gli enti pubblici che devono generare utili per coprire una parte sostanziale dei costi inerenti allo svolgimento dei propri compiti di servizio pubblico (e/o per specifici set di dati soggetti ad analoga previsione);
- qualora siano applicate **tariffe** per il riutilizzo, le condizioni applicabili, compresa la base di calcolo utilizzata, siano **fissate in anticipo e pubblicate**, ove possibile e opportuno, per via elettronica;
- gli Stati membri adottino **modalità pratiche per facilitare la ricerca** dei documenti disponibili per il riutilizzo: ad esempio attraverso elenchi dei documenti più importanti (insieme ai rispettivi metadati);
- ogni decisione sul riutilizzo contenga un riferimento ai **mezzi di ricorso** a disposizione del richiedente qualora questi intenda impugnarla. I mezzi di ricorso comprendono la **possibilità di revisione da parte di un organo imparziale** dotato delle opportune competenze, le cui decisioni sono vincolanti per l'ente pubblico interessato;

- il generale **divieto di accordi di esclusiva** per il riutilizzo dei dati (che comunque non opera nel caso in cui un diritto esclusivo si renda necessario per l'erogazione di un servizio d'interesse pubblico) non si applichi alla **digitalizzazione di risorse culturali**, qualora il periodo di esclusiva non ecceda, di norma, i **dieci anni**.

Al fine di fornire alcuni **principii guida unitari** nella gestione di specifici aspetti inerenti al riutilizzo dell'informazione del settore pubblico (impiego di licenze, tariffazione e dataset di rilevanza prioritaria), la Commissione Europea ha inoltre pubblicato - nel 2014 - una *Notice* composta da un set di [Guidelines on recommended standard licences, datasets and charging for the re-use of documents](#). I contenuti delle linee guida sono stati realizzati anche attraverso gli spunti raccolti mediante una propedeutica [consultazione pubblica](#) lanciata nel 2013.

*Le linee guida della
Commissione*

1.3.2 In Italia

L'**attuazione italiana della direttiva** comunitaria è avvenuta con il [Decreto legislativo 24 gennaio 2006, n. 36](#), pubblicato nella G.U. del 14 febbraio 2006, n. 37. Il Decreto Legislativo 36/2006 è stato successivamente modificato dalla [L. 96/2010](#) (art. 44) a seguito di una procedura di infrazione avviata dalla Commissione europea nei confronti dell'Italia per incompleto e scorretto recepimento della Direttiva. In particolare, la normativa italiana faceva salve vecchie norme sui dati catastali e ipotecari; non prevedeva un'informazione su mezzi di ricorso contro il mancato consenso alla riutilizzazione; ammetteva tariffe di riutilizzo determinate con criteri di dubbia compatibilità rispetto al dettato comunitario; non prevedeva un vero e proprio obbligo per le pubbliche amministrazioni di consentire il riutilizzo dei dati, né garantiva che questi venissero resi disponibili in formati aperti e realmente riutilizzabili.

Formati aperti

Secondo la legge italiana, si definisce aperto il formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi (art. 68 comma 3 lett. a) del Codice dell'Amministrazione Digitale).

Ulteriori informazioni, tra cui un repertorio dei formati aperti utilizzabili nelle pubbliche amministrazioni, sono disponibili sul sito Web dell'[AglD](#).

Alcuni esempi concreti sono forniti all'interno delle [Linee guida per i siti Web della PA](#), previste ai sensi della Direttiva 8/2009 del Ministro per la pubblica amministrazione e l'innovazione, in cui viene raccomandato l'uso dei seguenti formati aperti e standardizzati: HTML/XHTML per la pubblicazione di informazioni pubbliche su Internet; PDF con marcatura (secondo standard ISO/IEC 32000-1:2008); XML per la realizzazione di database di pubblico accesso ai dati; ODF e OOXML per documenti di testo; PNG per le immagini; OGG per file audio; Theora per file video; Epub per i libri.

Più recentemente, in Italia sono stati emanati **diversi atti normativi**, volti a riconoscere ed ufficializzare il ruolo dei dati aperti. Tra queste, la Legge di semplificazione (Legge 4 aprile 2012, n. 35) introduce l'Agenda Digitale Italiana (art. 47) e promuove il paradigma dei dati aperti quale modello di valorizzazione del patrimonio informativo pubblico (art. 47, comma 2-bis lett. b). Il Decreto Legge 18 ottobre 2012, n. 179 recante "Ulteriori misure urgenti per la crescita del Paese" (c.d. [Decreto Sviluppo bis](#)) introduce il concetto di "open by default" per tutte le informazioni pubblicate dalla PA (art. 9), promuove l'adozione di software libero e prescrive l'impiego di formati aperti nella pubblicazione dei dati da parte della pubblica amministrazione (art. 9-bis); il Decreto Legge porta queste misure nel [Codice per l'Amministrazione Digitale](#) (CAD, Decreto Legislativo, 7 marzo 2005, n. 82), modificandone gli artt. 52 e 68. Il Decreto Legislativo, 14 marzo 2013, n. 33 (c.d. [Decreto Trasparenza](#)), riorganizza la disciplina riguardante gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte

CAD e Decreto Trasparenza

delle pubbliche amministrazioni, introducendo per dette informazioni, soggette a pubblicazione obbligatoria, il diritto d'accesso civico (art. 5) e la conseguente messa a disposizione in formato aperto, senza "ulteriori restrizioni diverse dall'obbligo di citare la fonte e di rispettarne l'integrità" (art. 7). Il [Decreto Legislativo n. 90/2014](#) convertito con modificazioni in [Legge n. 114/2014](#), individua - all'art. 24-quater, comma 2 - l'obbligo per tutte le pubbliche amministrazioni e società partecipate in modo totalitario o prevalente dalle pubbliche amministrazioni "di comunicare all'Agenzia per l'Italia digitale, esclusivamente per via telematica, l'elenco delle basi di dati in loro gestione e degli applicativi che le utilizzano". Per agevolare questi adempimenti, l'Agenzia per l'Italia Digitale (AgID, ex AIPA, CNIPA e DigitPA) ha reso disponibile sul proprio sito una [procedura on line](#) per consentire ai soggetti interessati di trasmettere il catalogo delle basi dati e degli applicativi che le utilizzano ed una sezione di [FAQ](#) contenenti chiarimenti sugli adempimenti e sulla procedura d'invio.

Decreto Trasparenza

Nelle aspettative di molti, il "[Decreto Trasparenza](#)" non doveva essere altro che una trasposizione dell'americano FOIA (Freedom of Information Act). In tal senso spingevano e spingono tuttora varie iniziative della società civile, quali <http://www.foia.it/> o <http://www.foia4italy.it/>. La scelta del legislatore (con la legge 6 novembre 2012, n. 190) è andata in una direzione differente, ossia quella di predisporre nuove forme di pubblicità assieme ad un riordino della disciplina riguardante gli obblighi già esistenti in tal senso.

Da un lato, il decreto assicura, per i documenti oggetto di pubblicazione obbligatoria, un grado di efficienza maggiore del modello "a richiesta" del FOIA. D'altro canto, la normativa italiana in materia di richiesta di accesso a documenti non generalmente pubblici (l. 241/90) resta molto restrittiva rispetto ai paesi dotati di moderni FOIA.

Un primo indiscutibile punto di forza del Decreto Trasparenza è la **definizione del concetto di trasparenza**, intesa ora come "*accessibilità totale delle informazioni concernenti l'organizzazione e l'attività delle pubbliche amministrazioni, allo scopo di favorire forme diffuse di controllo sul perseguimento delle funzioni istituzionali e sull'utilizzo delle risorse pubbliche*" (art. 1). Rispetto al passato, quindi, l'accesso ai documenti amministrativi viene abilitato a quelle stesse "forme diffuse di controllo" che l'art. 24 comma 3 della l. 241/90 vietava. Su questa scia, l'art. 5 comma 1 del Decreto Trasparenza istituisce un nuovo tipo di accesso, denominato "**accesso civico**", il cui esercizio non è sottoposto "*ad alcuna limitazione quanto alla legittimazione soggettiva del richiedente*", né deve essere motivato ed è assolutamente gratuito (art. 5 comma 2).

Il secondo punto di forza del decreto è che quanto appena visto, e cioè una nuova forma di accesso ai documenti amministrativi, non è rimasta un'innovazione a sé stante, bensì affiancata dalla previsione che il **riutilizzo dei documenti**, che vengono pubblicati e cui si accede, costituisce un diritto vero e proprio. L'art. 3 del decreto sancisce infatti che "*tutti i documenti, le informazioni e i dati oggetto di pubblicazione obbligatoria ai sensi della normativa vigente sono pubblici e chiunque ha il diritto di conoscerli, di fruirne gratuitamente e di utilizzarli e riutilizzarli ai sensi dell'art. 7 del presente decreto*".

Tra le stonature del decreto si colloca, invece, la discordanza rispetto all'art. 18 del Decreto Sviluppo, poi abrogato proprio dal Decreto 33/2013, che prevedeva, in materia di amministrazione aperta, un regime di pubblicità per le "*sovvenzioni, i contributi, i sussidi ed ausili finanziari alle imprese e l'attribuzione dei corrispettivi e dei compensi a persone, professionisti, imprese ed enti privati*", comprendendovi comunque anche i "*vantaggi economici di qualunque genere*". Nel testo confluito negli artt. 26-27 del Decreto Trasparenza è stato ridotto il campo degli obblighi di pubblicità in merito all'uso delle risorse pubbliche. Per fare qualche esempio, basti pensare che non sono più oggetto di pubblicazione obbligatoria "*l'attribuzione dei corrispettivi e dei compensi*" e, ancor peggio, i contratti (e relativi capitolati) che le pubbliche amministrazioni stipulano con soggetti terzi. Inoltre, sparisce l'obbligo di pubblicità per le informazioni relative ad erogazioni di importo inferiore a mille euro.

Ai sensi del Decreto Trasparenza sono comunque oggetto di pubblicazione numerosi documenti, suddivisi per capi:

- Organizzazione ed Attività della Pubblica Amministrazione: i documenti e le informazioni di indirizzo politico come atto di nomina e curriculum, costo del personale, bandi di concorso, rendiconti dei gruppi consiliari regioni e provinciali, ecc.;

- Prestazioni offerte e Servizi erogati: tempi di pagamento della PA, procedimenti amministrativi e controlli sulle dichiarazioni sostitutive e l'acquisizione d'ufficio dei dati;

- Settori Speciali: contratti pubblici di lavori, servizi e forniture (procedure per l'affidamento e l'esecuzione); pianificazione, realizzazione e valutazione delle opere pubbliche (documenti di programmazione anche pluriennale delle opere pubbliche di competenza dell'amministrazione, le linee guida per la valutazione degli investimenti; le relazioni annuali; ogni altro documento predisposto nell'ambito della valutazione); pianificazione e governo del territorio (piani territoriali, piani di coordinamento, piani paesistici, strumenti urbanistici, generali e di attuazione, nonché le loro varianti);

- Informazioni ambientali;

- Servizio Sanitario Nazionale: procedure di conferimento degli incarichi di direttore generale, direttore sanitario e direttore amministrativo, nonché degli incarichi di responsabile di dipartimento e di strutture semplici e complesse;

- Interventi straordinari e di emergenza che comportano deroghe alla legislazione vigente: provvedimenti contingibili e d'urgenza adottati, i termini temporali eventualmente fissati per l'esercizio dei poteri di adozione dei provvedimenti straordinari, costo previsto degli interventi e il costo effettivo sostenuto.

Va infine ricordata la figura del “**responsabile della Trasparenza**”, un ruolo che l'art. 43 affida, all'interno di ogni amministrazione, al responsabile per la prevenzione della corruzione. Quest'ultimo ha il compito di aggiornare il Programma triennale per la trasparenza e l'integrità, controllare e assicurare la regolare attuazione dell'accesso civico sulla base di quanto stabilito dal decreto 33/2013 e segnalare “i casi di inadempimento o di adempimento parziale degli obblighi in materia di pubblicazione previsti dalla normativa vigente, all'ufficio di disciplina, ai fini dell'eventuale attivazione del procedimento disciplinare”. Il responsabile è infine competente anche a segnalare gli inadempimenti al vertice politico dell'amministrazione, all'OIV ai fini dell'attivazione delle altre forme di responsabilità.

Ai sensi dell'art. 47, l'**inadempimento degli obblighi** di pubblicazione o la mancata predisposizione del Programma triennale “costituiscono elemento di valutazione della responsabilità dirigenziale, eventuale causa di responsabilità per danno all'immagine dell'amministrazione e sono comunque valutati ai fini della corresponsione della retribuzione di risultato e del trattamento accessorio collegato alla performance individuale dei responsabili”. Sanzioni amministrative specifiche (da 500€ a 10.000€) sono invece previste, a carico del responsabile della mancata comunicazione o violazione, nel caso di omessa o parziale pubblicazione dei dati di cui agli artt. 14 e 22.

Si segnalano inoltre le misure per l'applicazione concreta dell'**Agenda Digitale Italiana** - individuate nel già citato Decreto Sviluppo Bis e trasfuse all'interno dell'art. 52, comma 6 e 7, del CAD - che contemplano

[L'Agenda nazionale e le Linee guida dell'AgID.](#)

[l'Agenda nazionale per la valorizzazione del patrimonio informativo pubblico](#) da aggiornarsi all'inizio di ogni anno, le complementari [Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico](#) (che estendono le precedenti [Linee guida per l'interoperabilità semantica attraverso i Linked Open Data](#) della Commissione di Coordinamento SPC - Sistema Pubblico di Connettività) ed il Rapporto sullo stato di avanzamento del processo di valorizzazione del patrimonio pubblico, predisposti dall'Agenzia per l'Italia Digitale.

1.3.3 A livello delle autonomie locali

Una forte sensibilità verso la valorizzazione del patrimonio di dati pubblici quale strumento per l'esercizio dei diritti all'informazione e di

[Leggi regionali open data.](#)

partecipazione democratica traspare dalle leggi promulgate dalla [Regione Emilia Romagna](#) e dalla [Regione Toscana](#) già nel 2004; prima ancora, cioè, che in Italia venisse recepita la direttiva PSI. Successivamente, la **Regione Piemonte è stata la prima ad aver predisposto strumenti normativi per agevolare il riutilizzo dell'informazione pubblica**. A seguire, numerose altre amministrazioni regionali si sono dotate di strumenti simili: così, ad esempio, la [Regione Friuli Venezia Giulia](#), con la Legge regionale 14 luglio 2011, n. 9. In altre Regioni sono state approvate o sono in corso di approvazione, sotto forma di disegno o proposta di legge, prescrizioni normative sul tema degli open data: è il caso delle amministrazioni regionali di [Lazio](#), [Basilicata](#), [Umbria](#), [Lombardia](#), [Sicilia](#), [Puglia](#), [Toscana](#), [Campania](#), e della [Provincia Autonoma di Trento](#). Un **elenco**

delle **principali leggi regionali** in materia di open data è contenuto nelle [Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico](#) (pp. 24 e ss.); inoltre, [Regolamenti e Direttive Locali](#) sono oggetto di uno specifico approfondimento contenuto nelle Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico, curato da Regioni-Cisis. Una sintesi interessante emerge anche dalle [infografiche](#) del sito Web dati.gov.it, in particolare quelle relative alla distribuzione geografica delle pubbliche amministrazioni che rilasciano dati aperti e al numero di dataset pubblicati.

Regione Piemonte

Nel 2009, Regione Piemonte si è dotata di **linee guida** per i processi di riutilizzo del proprio patrimonio informativo, associate alla definizione di licenze standard per tipologia di dati e categorie di utenza, che definiscono le discipline d'uso dei dati regionali ([Delibera di Giunta 31 - 11679 del 29 giugno 2009](#)). Tali linee guida sono state aggiornate nel 2010 ([Delibera di Giunta regionale 36 - 1109 del 30 novembre 2010](#)).

Sul finire del 2011 è entrata in vigore la [Legge Regionale 23 dicembre 2011, n. 24](#), "**Disposizioni in materia di pubblicazione tramite la rete Internet e di riutilizzo dei documenti e dei dati pubblici dell'amministrazione regionale**", mediante la quale l'amministrazione regionale garantisce la disponibilità, la gestione, l'accesso, la trasmissione, la conservazione e la fruibilità dei dati in modalità digitale. Sono pertanto riutilizzabili tutti i dati in possesso della Regione di tipo aggregato/anonimo o senza vincoli di privacy e che adottano come licenza standard la licenza Creative Commons CC0 (o, in alternativa, CC BY; vedi p. 21 in merito alle licenze Creative Commons).

Nel 2012 sono state adottate **linee guida regionali attuative** della Legge 23 dicembre 2011, n. 24 ([D.G.R. 8 Ottobre 2012, n. 22-4687](#)). Tali linee guida mettono a disposizione la piattaforma Open Data al fine di favorire i principi di efficienza ed economicità dell'azione amministrativa, favoriscono il riutilizzo (preferibilmente a titolo gratuito) dei dati pubblici e dei documenti contenenti dati pubblici e l'accesso ai relativi servizi, prodotti e banche dati regionali da parte degli interessati. Le Linee guida prevedono poi ulteriori indicazioni circa il perseguimento di condizioni di riutilizzo dei dati eque, adeguate e non discriminatorie, preferibilmente ricorrendo a licenze standard (come le licenze Creative Commons) e la diffusione dei dati grezzi e/o aggregati, organizzati razionalmente, comprensivi dei relativi metadati attraverso Internet e in formati elettronici standard ed aperti. Le linee guida prevedono possibilità di poter richiedere il riutilizzo di dati diversi o ulteriori rispetto a quelli presenti sulla piattaforma e di poter presentare reclamo contro la decisione sulla richiesta di riutilizzo, in caso di mancato o parziale accoglimento.

A livello comunale.

Dato il crescente interesse per le politiche open data, è difficile fornire un quadro esaustivo di tutte le iniziative a livello cittadino: le giunte e i consigli comunali che hanno adottato delibere per avviare il processo di apertura dei dati sono moltissime. Per ottenere un quadro delle amministrazioni comunali attualmente impegnate nel rilascio di dati aperti si può ricorrere, anche in questo caso, all'[infografica](#) "Dove sono i data store italiani?" del sito Web dati.gov.it, la quale, opportunamente settata sulla sola dimensione comunale, restituisce una mappa delle amministrazioni che pubblicano dati a livello cittadino.

Un approccio interessante, come quello piemontese, abilita i vari comuni (per altro stimolati da un progetto sostenuto dall'ANCI) ad aderire alle iniziative portate avanti dalla propria regione: in questo modo, le infrastrutture come i portali open data regionali vengono messe a disposizione di soggetti pubblici per i quali sarebbe stato inefficiente creare una propria autonoma piattaforma di pubblicazione dei dati.

1.4 La pratica dell'apertura dei dati: strumenti e processi

1.4.1 Fare open data nella pratica

L'esperienza maturata all'interno dei progetti sui dati aperti di maggiore successo, dagli USA alla Nuova Zelanda, passando per il Regno Unito e certamente anche in Italia, ha portato nel tempo ad individuare alcune buone pratiche e raccomandazioni.

Buone pratiche di apertura dei dati

Le principali buone pratiche per impostare un processo di apertura dei dati si possono riassumere nei principi seguenti. Come sempre, è possibile immaginare molte eccezioni, ma – specialmente qualora non si sia in possesso di particolare esperienza nel campo open data – è buona norma motivare attentamente le deviazioni dal principio generale.

Creare una squadra. Per poter partire, è importante che ci sia un leader, un politico o un funzionario che “ci metta la faccia”. Ma, soprattutto, serve una squadra, che abbia alcune competenze chiave (alcune delle quali possono essere possedute dalla stessa persona): serve un'idea di quali sono i dati a disposizione; e qualcuno che possa facilmente relazionarsi con i vari dirigenti responsabili di ogni dataset; serve qualcuno che abbia competenze legali, in particolare di proprietà intellettuale e di tutela dei dati personali (e magari di diritto amministrativo); è inoltre utile che qualcuno abbia un'idea di come i dati sono gestiti tecnicamente e che sappia dare qualche indicazione di base sui modi tecnici per mettere i dati a disposizione; possibilmente, qualcuno dovrebbe conoscere (o studiare) gli aspetti principali del movimento open data, per emulare le migliori pratiche.

Fare cose semplici. Soprattutto, partire da progetti piccoli e semplici, usando i primi risultati concreti per rafforzare il processo. Non è necessario aprire tutti i dati; non è necessario aver mappato tutti i dataset che si possiedono prima di partire; non è necessario “aprire” prima i dati più interessanti, se farlo solleva problemi particolarmente complessi. Si può partire dai dati nella sfera di competenza e controllo dei politici e dei dirigenti più aperti, innovativi o, semplicemente, ai quali si è riusciti a spiegare meglio a cosa serva aprire i dati. L'esperienza suggerisce che essere trasparenti chiarendo che quello che si sta facendo è solo l'inizio di un processo complesso ed il fatto stesso di aver cominciato, anche se a piccoli passi, sarà certamente apprezzato. Per partire da cose semplici, non è necessario abbandonare progetti ambiziosi: è opportuno però prediligere progetti modulari.

Dare aggiornamenti frequenti. Una delle ragioni per fare cose semplici è poterle fare abbastanza in fretta. Un'altra è apprendere dall'esperienza. Rilasciare spesso il materiale che si ha a disposizione permette due grandi vantaggi: mantenere alto l'interesse e rendere credibile la strategia modulare adottata (poiché ci saranno frequentemente nuovi piccoli dataset a disposizione); e imparare dai propri errori o dalle esperienze positive.

Coinvolgere le persone (fin da subito e spesso). Parlare con i riutilizzatori attuali e potenziali è sempre un approccio vincente. L'idea non è necessariamente quella di raccogliere tutti gli stakeholder potenzialmente interessati attorno ad un tavolo di lavoro. L'invito è quello di confrontarsi con le esigenze di chi i dati li potrà poi riutilizzare. Nel farlo, è centrare ricordarsi che alcuni riutilizzatori sono “indiretti”, ovvero degli intermediari dell'informazione: solo pochi cittadini e imprese saranno in grado di riutilizzare direttamente i dati pubblici; molti altri, potranno accedere a questi dati grazie al lavoro di questi intermediari capaci di riorganizzare, trasformare e presentare nuovamente i dati. L'intermediazione e a volte l'aggregazione e la scelta di punti di vista parziali sui dati stessi, dunque, restano necessari: il vantaggio dell'approccio open data è proprio la presenza (almeno potenziale) di una pluralità di intermediari e di punti di vista.

Non chiedersi “a chi può servire o interessare?”. Spesso ci si chiede come possano i cittadini usare in qualunque modo sensato dei database di enorme ricchezza e complessità. Non bisogna, invece, sottovalutare gli intermediari della Rete: può darsi che siano in grado di creare interfacce tramite le quali molti altri cittadini (o direttamente i loro smartphone e computer!) saranno in grado di fruire dei dati, eventualmente arricchendoli e correggendoli. Inoltre, come già sottolineato: se non ci sono buone ragioni per non condividere un dato e non si sa a cosa possa essere utile quel dato, allora c'è un'ottima ragione per “aprirlo”: sicuramente, se qualcuno lo riuserà, si tratterà di un riutilizzo a cui il detentore del dato difficilmente avrebbe potuto pensare!

Non spaventarsi troppo e, nel caso, confrontarsi. Capita spesso che qualcuno sollevi problemi apparentemente difficili o impossibili da risolvere. Quasi sempre, questi problemi hanno una soluzione, anche se capita che tale soluzione passi per una riduzione della qualità informativa dei dati (anonimizzazione, aggregazione – inclusa la pubblicazione di medie o dati statistici – o simili procedure che fanno allontanare dal dato grezzo). In caso di incertezza, è utile confrontarsi con chi ha esperienza di progetti relativi ai dati aperti, perché è probabile che i problemi di ciascuno siano già stati affrontati (e superati) da altri. Senza nulla togliere alla competenza degli esperti (ad esempio di problemi legali) che stanno dentro ciascuna istituzione, spesso è interessante e produttivo mettere queste persone in contatto con chi sta fuori dall'istituzione stessa e ha esperienza di open data, perché il punto di vista da cui affrontare le questioni relative ai dati aperti è molto diverso da quello tradizionalmente adottato per la gestione interna della PA.

Non essere integralisti (in nessun senso). I dati pubblici sono tanti e diversi, è dunque naturale che non tutti possano essere trattati allo stesso modo. Quando si introduce qualcuno a questo tema, si può certamente mirare ai principi generali relativi all'apertura e alla libertà massime. Tuttavia, ricordate e spiegate sempre che non ci sono questioni di principio, ma solo questioni di buon senso; che tutte le regole hanno eccezioni, ma che fare eccezioni non dovrebbe essere la regola. Nell'incertezza, è quasi sempre possibile e sensato adottare più di una soluzione: offrire il dato in più formati, applicare più licenze (ciascuna delle quali non esclusiva). L'unica cautela da adottare è evitare la confusione: un certo formato e licenza (possibilmente i più standard, aperti e liberi) dovrebbero essere quelli più chiaramente pubblicizzati, offrendo comunque le alternative per massimizzare la libertà dei riutilizzatori.

Rendere i dati disponibili gratuitamente. Di fatto, questo principio è una sorta di corollario dell'uso di licenze libere, poiché un dato liberamente riutilizzabile e redistribuibile non è facile da monetizzare (ossia da dare in licenza ad un prezzo che renda conveniente la transazione). Tuttavia, ci sono altre ragioni per distribuire gratuitamente i dati. La letteratura economica su questo tema, in effetti, ha confermato il valore della regola aurea "Prezzo = Costo Marginale" in questo campo. E, data la distribuzione online, il costo marginale è sostanzialmente zero. Infine, e nella pratica questa potrebbe essere l'argomentazione più persuasiva, non è raro che il costo di esigere una tariffa sia maggiore del ricavo generato dalla stessa! Si consideri, del resto, che la nuova Direttiva 2013/37/UE sulla PSI ha sposato in pieno questo orientamento (salvo che per ristrette categorie di enti e documenti), limitando al solo costo marginale il prezzo che gli enti pubblici possono richiedere per il riutilizzo dei dati.

Non discriminare tra umani e macchine. Non si dovrebbe mai discriminare l'accesso ai dati tra umani e macchine. A prima vista questo principio potrebbe far sorridere, ma la questione è rilevante ed è una questione di libertà (di impresa, ma anche di azione democratica). La necessità di registrarsi su un portale, per esempio, oppure la necessità di accettare esplicitamente una licenza e simili sono piccole noie per un essere umano, ma possono diventare ostacoli insormontabili per una macchina. E questi ostacoli possono rendere i propri dati difficili da trovare perché i motori di ricerca non li vedono, oppure difficili da scaricare e aggiornare in automatico, nell'ambito di applicazioni e servizi di terzi.

Le fasi del processo di apertura dei dati.

Come indica l'"[Open Data Handbook](#)", il manuale predisposto dalla Open Knowledge Foundation sul tema Open Data, autorevole risorsa informativa per il vasto pubblico interessato al tema, di cui sono disponibili traduzioni

in varie lingue tra cui l'italiano, il processo di apertura dei dati si articola in una serie di passi che consistono in:

1. **Individuare i dati da aprire.** Analizzare i passi successivi in astratto è difficile o impossibile, per questo motivo è utile poter discutere sempre di dati abbastanza ben determinati, di cui sia possibile mostrare un campione ai vari possibili interlocutori (giuristi, dirigenti, amministratori, potenziali riutilizzatori). La scelta non è necessariamente esaustiva e globale, in quanto l'apertura dei dati va considerata un processo continuo e perpetuo, in cui si può procedere con un iniziale set di dati, operando successive integrazioni e approfondimenti, anche in base ai feedback dei potenziali utilizzatori.
2. **Individuare quali diritti esistano sui dati.** Un passo legato alla fase di individuazione dei dati da aprire attiene alla considerazione del regime giuridico dei dati. Da dove vengono i dati? Chi li ha raccolti? Con quali risorse? Ci sono (stati) contratti di qualche tipo che regola(va)no l'attività di raccolta e/o l'uso dei dati stessi? Parte di questi dati sono personali? O addirittura sensibili? Ci possono essere diritti di proprietà (ad esempio informazioni commerciali segrete)? Se troppe di queste domande sembrano difficili o impossibili da rispondere, anche con l'aiuto dell'ufficio legale del proprio ente, si può prendere in considerazione l'idea di scegliere un altro dataset, in particolare se ci si trova all'inizio dell'attività di apertura dei dati.
3. **Applicare una licenza aperta.** Una volta individuati i dati da aprire, va determinato il livello di apertura giuridica da operare. Quali diritti si vogliono concedere ai riutilizzatori? A che condizioni? Chi è titolato a rilasciare la licenza in questione? In generale, è opportuno che esista un documento, condiviso e formalmente approvato, che legittimi la procedura di

licenza dei dati. Può essere uno dei passaggi istituzionali più utili nell'evoluzione da sporadici progetti pilota Open Data ad una vera politica dei dati aperti. Peraltro, si tenga presente che, laddove i dati venissero pubblicati senza adottare alcuna licenza, si applicherebbe il principio dell'“open by default”, disposto dall'art. 52, comma 3 del CAD, il quale prevede che “*I dati e i documenti che le amministrazioni titolari pubblicano, con qualsiasi modalità, senza l'espressa adozione di una licenza di cui all'articolo 2, comma 1, lettera h), del decreto legislativo 24 gennaio 2006, n. 36, si intendono rilasciati come dati di tipo aperto*”.

Creative Commons

Creative Commons (CC) è un'organizzazione senza fini di lucro, che ha pubblicato, a partire dal 2002, un insieme di licenze divenute uno standard a livello internazionale.

Come descritto sul sito <http://creativecommons.it/>, “*le licenze Creative Commons offrono sei diverse articolazioni dei diritti d'autore per [...] creatori che desiderino condividere in maniera ampia le proprie opere secondo il modello "alcuni diritti riservati". Il detentore dei diritti può non autorizzare a priori usi prevalentemente commerciali dell'opera (opzione Non commerciale, acronimo inglese: NC) o la creazione di opere derivate (Non opere derivate, acronimo: ND); e se sono possibili opere derivate, può imporre l'obbligo di rilasciarle con la stessa licenza dell'opera originaria (Condividi allo stesso modo, acronimo: SA, da "Share-Alike"). Le combinazioni di queste scelte generano le sei licenze CC*”.

- 4. Mettere i dati a disposizione online.** Oltre agli aspetti giuridici dell'apertura, vanno definiti gli aspetti tecnici, individuando sia i formati machine-readable con cui rendere disponibili i dati, sia le modalità tecniche per metterli a disposizione online. Per quanto riguarda i formati, entrano in gioco le opzioni definite nel [sistema di rating](#) di Tim Berners-Lee. Per quanto attiene invece alle modalità di distribuzione online, le alternative sono molteplici. La scelta ottimale sarebbe pubblicare sul Web i dati, istituendo un sito ad hoc oppure attraverso pagine del proprio sito pre-esistente.
- 5. Far trovare i dati.** Se nascosti, anche quei dati che hanno grandi potenzialità resteranno inutilizzati perché nessuno saprà che esistono e dove si trovano. Qualche energia va dunque spesa anche per rendere i propri dati facilmente individuabili sul Web. Volendo, si può creare un proprio sito/catalogo o portale apposito (a volte detto anche Information Asset Register). Oppure si può verificare se ne esiste uno, ad esempio, nella propria regione (<http://dati.piemonte.it> è un caso di portale regionale aperto a varie pubbliche amministrazioni e istituzioni piemontesi). In Italia, il 18 ottobre 2011 è stato inaugurato il portale <http://www.dat.gov.it>, un catalogo pubblico nazionale dedicato ai dati aperti.

1.4.2 Costi, profitti e tariffe

Come si è accennato, la migliore politica per massimizzare le opportunità del riutilizzo dei dati consiste nel **mettere a disposizione i dati gratuitamente**, richiedendo altrimenti il mero [costo marginale di messa a disposizione dei dati](#) stessi. Ai fini del presente lavoro, mutuando dalla

Recupero dei soli costi marginali. Online, sostanzialmente gratis!

definizione generale, con "costo marginale" si intende la variazione nei costi totali di messa a disposizione del dato che si verifica quando viene riprodotta e distribuita ("scaricata") un'unità in più del dato. Tale costo, in ambito digitale, è in sostanza pari a zero. Questo approccio, che permette la massima diffusione dei dati e, di conseguenza, degli effetti positivi generati, è stato precisamente recepito dall'Unione Europea: a seguito delle modifiche alla Direttiva PSI, infatti (e salvo che per alcune limitate tipologie di enti pubblici e documenti), l'eventuale tariffa non può più prevedere l'apposizione di un congruo utile sugli investimenti, impedendo così alle Amministrazioni di “fare cassa” scoraggiando al tempo stesso il riutilizzo. È dunque possibile considerare, ad esempio la (minima) quota di ulteriore banda necessaria, o analoghi costi di natura tecnica, ma certamente non i costi fissi di realizzazione del servizio. Senza dimenticare che ogni pagamento implica non solo introiti, ma nuovi costi per la PA (legati alla gestione delle transazioni), costi che in

alcuni casi possono rendere addirittura antieconomico il rilascio dei dati a titolo oneroso (come nel caso del Sistema Informativo Territoriale della Regione Piemonte).

Sistema Informativo Territoriale della Regione Piemonte

Il Sistema Informativo Territoriale della Regione Piemonte comprende una serie articolata di **repertori cartografici**, aggiornati nel tempo. Si tratta di un insieme di dati geografici, rappresentati nella forma di mappe che riportano informazioni geometriche, topologiche e descrittive del territorio regionale. Questi dati – che rappresentano un’istantanea del territorio regionale e degli oggetti, naturali e artificiali, che lo compongono – si prestano a svariate forme di riutilizzo di particolare importanza, dalla creazione di nuove mappe (con l’aggiunta di ulteriori informazioni) alla georeferenziazione dei servizi digitali. L’aggiornamento dei dati cartografici, pur facendo parte del mandato pubblico, comporta per l’amministrazione regionale un ingente costo. Per dare una misura concreta, la realizzazione della carta di un grande centro urbano (a scala 1:2000, con un aggiornamento tendenzialmente necessario ogni due anni) comporta una spesa di circa 50 euro all’ettaro.

Prima dell’approvazione delle Linee Guida relative al riutilizzo e all’interscambio del patrimonio informativo regionale del novembre 2010 (abrogate dalla Legge Regionale n° 24 del 23/12/2011, che ha riorganizzato le disposizioni circa la pubblicazione e il riutilizzo di documenti e dati dell’amministrazione regionale, ma che già prevedevano il rilascio libero e gratuito dei dati dell’amministrazione pubblica piemontese), parte dei repertori geografici veniva rilasciata a titolo oneroso, secondo uno specifico tariffario e, naturalmente, condizioni di rilascio compatibili con quanto prescritto dalla direttiva europea INSPIRE (relativa appunto ai dati geografici). I ricavi netti annui stimati erano compresi tra i 10mila e i 20mila euro. I costi cagionati dalla natura onerosa e manuale delle transazioni (prive di un canale di distribuzione digitale) – ossia le risorse umane dedicate, nonché le spese di fatturazione e altre spese accessorie – assommavano invece a circa 50mila euro all’anno, rendendo dunque, di fatto, **antieconomica l’offerta di dati a titolo oneroso** e incentivando un rilascio libero attraverso la rete.

La scarsità di risorse non consente peraltro di effettuare gli aggiornamenti dei database geografici con la frequenza che sarebbe desiderabile: per questo, meccanismi virtuosi di collaborazione con i riutilizzatori delle mappe potrebbero rappresentare un’opportunità molto più interessante della “vendita” del dato stesso.

Dati aperti e concorrenza.

Un aspetto economico relativo a riutilizzo merita un commento specifico: la possibilità che il riutilizzo favorisca un determinato soggetto

commerciale, apparentemente a spese dei contribuenti, come nel caso del servizio di traduzione GoogleTranslate. La risposta, qui, è dunque del tipo “**si al riutilizzo commerciale, purché aperto a tutti**”. Ciò, val la pena esplicitarlo, è garantito dal fatto che le licenze standard largamente consigliate per la messa a disposizione dei dati non sono mai esclusive; inoltre, le licenze esclusive sono esplicitamente scoraggiate (e normalmente vietate) dalla Direttiva PSI (e suo recepimento nazionale). Eventuali casi particolari, in cui malgrado l’assenza di licenze esclusive alcuni soggetti economici rafforzeranno la loro posizione dominante grazie alla PSI, saranno poi da gestirsi in via successiva, ad esempio tramite l’intervento dell’autorità antitrust.

Trattando di concorrenza, un altro caso merita una menzione. Immaginiamo che qualche riutilizzatore “consolidato” si lamenti di un’eventuale riduzione delle tariffe o in generale di una maggiore apertura (che faciliti l’ingresso di nuovi entranti, nonché suoi concorrenti), sostenendo magari che la sua possibilità di stare sul mercato potrebbe a quel punto essere minacciata. Se così accadesse, si avrebbe un netto indizio del fatto che quell’operatore basava la propria attività su un posizionamento (quasi) monopolistico (o comunque da rendita di posizione), più che sulla creazione di servizi realmente innovativi. Altrimenti non avrebbe di che preoccuparsi. Di conseguenza, perché una PA dovrebbe farsi scrupoli ad aprire i dati, malgrado queste obiezioni?

Google Translate

Nel Maggio 2010, la parlamentare europea Evelyn Regner ha indirizzato alla Commissione Europea un'interrogazione riguardante l'utilizzo di documenti tradotti dell'UE da parte di Google, per lo sviluppo del suo software di traduzione.

Come noto, Google ha sviluppato il servizio di traduzione GoogleTranslate, liberamente disponibile online. Evelyn Regner chiedeva dunque: “È la Commissione a conoscenza del fatto che l'azienda statunitense Google Inc. utilizza i documenti dell'UE tradotti in 23 lingue quale base per un servizio gratuito, perseguendo tuttavia un interesse economico? Un siffatto servizio gratuito induce un numero sempre maggiore di utenti Internet in tutto il mondo a consultare la homepage di Google. Qual è la posizione della Commissione rispetto a tale riguardo?” Continuando poi: “Come giustifica la Commissione ai contribuenti europei il fatto che per i servizi di traduzione siano stanziati milioni di euro e un'azienda privata statunitense ne tragga indirettamente beneficio?”

Il Commissario Kroes ha risposto a nome della Commissione, confermando una piena consapevolezza dei fatti descritti. Nella risposta, il Commissario ha fatto addirittura riferimento ad altri usi, commerciali e non-commerciali, di altri documenti pubblicati in più lingue dalla Commissione, dal Parlamento e da varie istituzioni europee.

Il riutilizzo descritto è stato posto nel contesto della Direttiva PSI (2003/98/EC), che ha precisamente lo scopo di incoraggiare questo tipo di attività e le loro ricadute (economiche e non), normalmente positive per la società. Rispetto al caso specifico, poi, la Commissione ha ricordato che Google, pur essendo un operatore basato negli Stati Uniti, ha affiliate in più di una dozzina di paesi europei (inclusa l'Italia). Inoltre – e soprattutto – la Commissaria ha fatto osservare che l'informazione in esame è disponibile a qualsiasi compagnia alle medesime condizioni e che i costi di traduzione, lungi dall'essere un aiuto a Google, sono spese inevitabili e connesse ai compiti delle istituzioni europee. In conclusione, la Commissione non ha preso in considerazione l'imposizione di alcun limite aggiuntivo all'accesso e al riutilizzo dei documenti in questione.

1.4.3 Le possibilità di riutilizzo e la tutela dei dati personali

1.4.3.1 Un bilanciamento di interessi

La disciplina del riutilizzo dell'informazione pubblica e la necessità di tutelare i dati personali pongono le amministrazioni pubbliche, e in generale i soggetti che intendano sviluppare e diffondere open data, di fronte a una sfida complessa e non priva di rischi: sebbene i dataset contenenti dati personali siano tra quelli con maggior valore, infatti, la decisione da parte di un soggetto pubblico di mettere a disposizione alcuni dataset non può mettere in discussione il diritto dei cittadini alla tutela dei loro dati personali e della loro riservatezza.

Con questa consapevolezza, la normativa sul riutilizzo dei dati fa completamente salva la normativa sui dati personali (art.4, c.1, lett. a), D. Lgs. 36/2006; art. 1 Direttiva 2003/98/CE).

Pertanto, in base all'attuale normativa europea ed italiana, il riutilizzo di dati personali non è escluso a priori, ma implica un **giudizio di bilanciamento** tra i diversi diritti e interessi coinvolti “caso per caso”, considerando che la tutela della privacy e dei dati personali prevale sugli altri interessi in gioco.

I **dati personali** sono una classe molto ampia: comprendono qualsiasi informazione concernente una persona fisica identificata o identificabile. Il nome e cognome, l'indirizzo di residenza, il codice fiscale, l'indirizzo mail, ma anche l'immagine fotografica di una persona e la registrazione della sua voce sono alcuni tipici esempi di dati personali. Inoltre, quando si parla di persona identificabile, lo si fa in senso abbastanza ampio e anche indiretto, ad esempio mediante riferimento ad un numero di identificazione o ad uno o più elementi specifici caratteristici della sua identità fisica, fisiologica, psichica, economica, culturale o sociale.

I **dati personali sensibili**, invece, sono quei dati personali che si riferiscono alla sfera più intima del soggetto, “idonei a rivelare l'origine razziale ed etnica, le convinzioni religiose, filosofiche o di

Dati personali?

Tutti quelli associati ad una persona identificata o identificabile.

altro genere, le opinioni politiche, l'adesione a partiti, sindacati, associazioni od organizzazioni a carattere religioso, filosofico, politico o sindacale, nonché i dati personali idonei a rivelare lo stato di salute e la vita sessuale". Questi godono di una **protezione maggiore**. Alcuni esempi comprendono l'elenco dei titoli relativi alle condizioni di salute nei concorsi pubblici (es. i dati che permettono di individuare soggetti diversamente abili); l'iscrizione ad un partito politico e ad un'associazione sindacale, l'elenco degli studenti che seguono o non seguono le ore di religione a scuola o i file di log di navigazione Internet su siti erotici, politici, sindacali. Si noti pure che in taluni casi un dato che può sembrare semplicemente personale — ad esempio l'indirizzo mail — è in realtà un dato sensibile, poiché rivela l'adesione ad una determinata associazione, sindacato o partito, p. es. alberto.rossi@cgil.it o mario.verdi@pd.it.

Esiste poi una categoria di **dati detti "semi-sensibili"**, il cui trattamento "presenta rischi specifici". In merito, il Codice della privacy (art. 17) stabilisce che "il trattamento dei dati diversi da quelli sensibili e giudiziari che presenta rischi specifici per i diritti e le libertà fondamentali, nonché per la dignità dell'interessato, in relazione alla natura dei dati o alle modalità del trattamento o agli effetti che può determinare, è ammesso nel rispetto di misure e accorgimenti a garanzia dell'interessato, ove prescritti. [Tali misure e accorgimenti] sono prescritti dal Garante [...], anche a seguito di un interpello del titolare". Un esempio di "trattamento che presenta rischi specifici" è la pubblicazione online dell'elenco di genitori che non pagano le rette della scuola o dell'asilo.

Va ricordato che la possibilità di diffusione di dati personali da parte di un soggetto pubblico può avvenire, a norma del Codice della Privacy solo in presenza apposita previsione di legge o di regolamento (art. 19, c. 3, D. Lgs. 196/2003).

1.4.3.2 L'ampiezza del riutilizzo di dati personali

I dati personali nelle norme sugli open data

A livello europeo, con le Direttive PSI del 2003 e del 2013, il legislatore ha consentito la possibilità di ammettere al riutilizzo anche dati personali. In quelle sedi, tuttavia, non si è assunto l'onere di individuare casi specifici di trattamenti legittimi di diffusione in formato aperto di dati personali, né

chiare modalità di selezione dei dati personali diffondibili come open data. La stessa scelta è stata portata avanti dal legislatore italiano nel 2006 (D. Lgs. 36/2006 di recepimento della direttiva europea sul riutilizzo) e nel 2012 (D. L. 179/2012, conv. con L. n. 221/2012, che modifica il CAD includendovi la disciplina sul riutilizzo dell'informazione pubblica). Nel 2013, invece, il legislatore ha avvertito la necessità di intervenire in modo più diretto: nell'ambito del riordino della normativa sulla trasparenza amministrativa è stato infatti ridefinito il bilanciamento tra riutilizzo dell'informazione e tutela dei dati personali fissandone nuovi confini. Così, sono stati introdotti specifici limiti normativi alla diffusione in vista del riutilizzo di dataset contenenti dati personali (cfr. D. lgs. 33/2013, art. 4).

Il risultato è un **ridursi degli spazi legittimi per il riutilizzo di dati personali**.

Sebbene infatti sia stato introdotto l'obbligo di pubblicazione adottando formati di tipo aperto per alcune categorie di dati e documenti inerenti la trasparenza delle amministrazioni pubbliche, si è stabilito che laddove non sia previsto un obbligo di pubblicazione sia possibile diffondere - per il loro riutilizzo - dataset contenenti dati personali solo in seguito all'adozione di tecniche di anonimizzazione. Così, per i casi di dati personali a pubblicazione obbligatoria, l'art. 4 comma 1 stabilisce che debba esser possibile una loro diffusione in formato di tipo aperto attraverso siti istituzionali, che il loro trattamento segua modalità tali da consentire l'indicizzazione e rintracciabilità tramite motori di ricerca Web e che il loro riutilizzo avvenga nel rispetto dei principi sul trattamento dei dati personali. In aggiunta, il comma 3 dello stesso articolo, prevede la possibilità di rendere disponibili e riutilizzabili anche tutti quei dati, informazioni e documenti che le pubbliche amministrazioni non hanno l'obbligo di pubblicare ai sensi del decreto trasparenza o di altre specifiche previsioni di legge o regolamenti, una volta che si sia proceduto ad anonimizzare i dati personali eventualmente presenti (art. 4, c.3).

1.4.3.3 Le linee guida del Garante Privacy

Il Garante privacy si è espresso di recente sulla materia in esame nelle [Linee Guida in materia di trattamento dei dati personali, contenuti anche in atti e documenti amministrativi, effettuato per finalità di pubblicità e trasparenza sul Web da soggetti pubblici e da altri enti obbligati](#) (Garante Privacy, 12 giugno 2014).

L'interpretazione del decreto Trasparenza fornita dal Garante risulta restrittiva. Nelle Linee Guida si distingue nettamente tra dati a pubblicazione obbligatoria per finalità di trasparenza e dati a pubblicazione obbligatoria per qualunque altra finalità (es. pubblicità degli atti, etc.). Si noti che il Garante ha così interpretato la distinzione nell'area di applicazione del D. Lgs. 33/2013 ritenendo che l'obbligo di pubblicazione in formati aperti per il riutilizzo (art. 7, D. Lgs. 33/2013) riguardi solo i dati a pubblicazione obbligatoria per finalità di trasparenza relativi all'organizzazione e l'attività amministrativa, e non gli altri casi. Scrive il Garante che "Di conseguenza, tutte le ipotesi di pubblicità non riconducibili a finalità di trasparenza (...), qualora comportino una diffusione di dati personali, sono escluse dall'oggetto del d. lgs. n. 33/2013 e dall'ambito di applicazione delle relative previsioni fra cui, in particolare, quelle relative all'accesso civico (art. 5), all'indicizzazione (art. 4 e 9), al riutilizzo (art. 7), alla durata dell'obbligo di pubblicazione (art. 8) e alla trasposizione dei dati in archivio (art. 9)."

Il Decreto Trasparenza secondo il Garante.

Inoltre, distinguendo ulteriormente tra documenti a pubblicazione obbligatoria e dati personali in essi contenuti, ha chiarito che ove si sia un obbligo di pubblicazione per finalità di trasparenza dei documenti in formati aperti, non perciò quest'obbligo si estenda senza ulteriore valutazione ai dati personali. Per questi ultimi varranno i normali giudizi sui principi di finalità, necessità, pertinenza e non eccedenza, previsti dal Codice della Privacy, da compiersi avendo riguardo alle finalità di raccolta e trattamento e alle possibili finalità di trattamento successivo alla pubblicazione.

In merito alla presenza di eventuali dati sensibili, il Garante precisa che "È, invece, sempre vietata la diffusione di dati idonei a rivelare lo "stato di salute" (art. 22, comma 8, del Codice) e "la vita sessuale" (art. 4, comma 6, del d. lgs. n. 33/2013)" e che "anche alla luce di un'interpretazione sistematica delle disposizioni del decreto sulla trasparenza, i dati personali sensibili e giudiziari sono espressamente esclusi dal riutilizzo (art. 4, comma 1, e art. 7 del d. lgs. n. 33/2013). Inoltre, operando un'interpretazione forzatamente restrittiva del disposto legislativo, il Garante sostiene che l'obbligo di pubblicazione in formati di tipo aperto per consentire il riutilizzo ai sensi della normativa vigente (CAD, D. Lg. 36/2006) di documenti, dati e informazioni a pubblicazione obbligatoria (art. 7, D. Lgs. 33/2013), non implichi la pubblicazione di dati personali "di tipo aperto", che al contrario ritiene vietata: "Da ciò consegue che i dati personali oggetto di pubblicazione obbligatoria non sono liberamente riutilizzabili da chiunque per qualsiasi ulteriore finalità".

Ne consegue che, una volta che i dati personali a pubblicazione obbligatoria per finalità di trasparenza siano stati pubblicati in formati di tipo aperto sui siti istituzionali "Amministrazione trasparente", l'amministrazione dovrà **valutare caso per caso l'eventuale riutilizzabilità** di quei dati, procedendo a una valutazione rigorosa del rischio e del rispetto del principio di finalità (art. 11 Codice della Privacy), con l'obiettivo di ridurre il rischio di perdere il controllo sulle medesime informazioni o di dover far fronte a richieste di risarcimento del danno da parte degli interessati; a tale scopo ne ha indicato il contenuto e le modalità. Tale valutazione d'impatto potrà avvenire ex ante e per tipologia di dati, oppure su istanza qualora l'amministrazione ricevesse richiesta di riutilizzo ai sensi del D. Lgs. 36/2006.

Per questi casi, il Garante ritiene che sarà necessario predisporre una **Valutazione di Impatto** documentata e rigorosa al fine di ridurre il rischio di perdere il controllo sulle medesime informazioni o di dover far fronte a richieste di risarcimento del danno da parte degli interessati. Tale valutazione dovrà:

Valutazione di Impatto su privacy e dati personali.

1. stabilire se è lecito, alla luce dell'esistenza di un presupposto normativo idoneo, che i dati

- personali pubblicamente accessibili sui siti Web istituzionali possano essere riutilizzati da terzi e per scopi ulteriori (art. 11, comma 1, lett. a e b del Codice);
2. in caso di valutazione positiva, occorre poi verificare se l'utilizzo ulteriore di questi dati possa essere consentito:
 - limitatamente ai dati rielaborati in forma anonima e aggregata, individuando il livello appropriato di aggregazione e la specifica tecnica di anonimizzazione da utilizzare sulla base di una ponderata valutazione del rischio di re-identificazione degli interessati oppure rispetto a tutti o soltanto ad alcuni dei dati personali resi pubblici (cfr. artt. 3 e 11, lett. d, del Codice);
 - per qualsiasi scopo ulteriore o solo per taluni scopi determinati (art. 11, comma 1, lett. b, del Codice);
 - secondo modalità di messa a disposizione online conformi ai principi di necessità, proporzionalità e pertinenza (artt. 3 e 11 del Codice);
 - a condizione che gli utilizzatori adottino modalità tecniche e rispettino specifici vincoli giuridici definiti in apposite licenze predisposte al fine di individuare idonee cautele per tutelare i diritti degli interessati nei successivi trattamenti di dati a fini di riutilizzo.

Impiego di licenze e note legali

Il Garante Privacy, inoltre, raccomanda l'uso di licenze standard qualora l'amministrazione decida di diffondere dati personali per il riutilizzo. Si ricorda come già il Working Party Art. 29 nella sua [Opinion 06/2013 on open data and public sector information \('PSI'\) reuse](#) suggerisse di adottare

una licenza che restringesse le possibilità di riutilizzo di dati personali a ciò che sia strettamente compatibile con gli scopi con cui i dati sono stati raccolti. Il Garante Privacy italiano ha seguito questa raccomandazione specificandola sia con riguardo al riuso di dati personali, sia con riguardo ai dati anonimi derivati da dati personali. Per quest'ultima ipotesi le condizioni della licenza "dovrebbero vietare ai titolari delle licenze di re-identificare gli interessati e di assumere qualsiasi decisione o provvedimento che possa riguardarli individualmente sulla base dei dati personali così ottenuti, nonché prevedere in capo ai medesimi titolari l'obbligo di informare l'organismo pubblico nel caso in cui venisse rilevato che gli individui interessati possano essere o siano stati re-identificati".

Nel caso di riutilizzo di dati personali, invece, le condizioni di licenza dovrebbero indicare chiaramente "le finalità e le modalità degli ulteriori trattamenti consentiti". Quest'ultima raccomandazione del Garante pare tuttavia di dubbia praticabilità. Si imporrebbe in capo all'amministrazione una complessa definizione ex ante delle possibili ipotesi di trattamenti compatibili con la finalità di raccolta originaria. Ad una simile scelta seguirebbero inevitabili problemi di incompatibilità tra licenze simili. Inoltre, sembra opportuno non gravare una licenza di clausole non direttamente correlate con la gestione di diritti d'autore e connessi. Uno strumento negoziale come quello delle licenze non risulta infatti il mezzo più idoneo per definire aspetti di tutela dei dati personali, potendo sopperire in modo più consono alle esigenze di tutela l'indicazione delle finalità di raccolta originaria dei dati personali nella scheda di metadattazione dei dati pubblicati, accompagnata da una nota che ricordi l'osservanza della normativa sulla privacy, inserita tanto **nella scheda di metadattazione dei dati, quanto nelle note legali del sito** o portale istituzionale che esponga i dati in oggetto (così, ad esempio, il portale open data del Riksdag svedese, che nei [Terms of Use](#) spiega come il riutilizzatore di dati parlamentari contenenti dati personali debba assicurarsi che detto riutilizzo risulti conforme alle disposizioni del Personal Data Act, 1998: 204).

Quest'ultima possibilità è stata peraltro adottata anche dal Garante individuando la necessità che le PA provvedano ad inserire un messaggio di allerta specifico per i dati personali a pubblicazione obbligatoria in modalità open data presenti sulla pagina "Amministrazione trasparente" nei propri

siti Web. L'allerta deve informare il pubblico che i dati personali sono riutilizzabili in termini compatibili con gli scopi per i quali sono raccolti, e nel rispetto delle norme sulla protezione dei dati personali.

1.4.3.4 L'anonimizzazione

Come si è già evidenziato, i dati personali presentano delle peculiarità che rendono la loro gestione piuttosto delicata. Fortunatamente, è sempre possibile adottare una **soluzione tecnica** che consente la pubblicazione di un dato semplicemente eliminandone la natura personale. Infatti, un dato è personale finché resta riconducibile ad una persona fisica e questa riconducibilità può venir meno grazie a varie tecniche di **anonimizzazione** (ai sensi dell'art. 4 del Codice della privacy, il dato "anonimo" è il "il dato che in origine, o a seguito di trattamento, non può essere associato ad un interessato identificato o identificabile"). Tra queste tecniche rientra anche l'**aggregazione** di più dati personali, ad esempio calcolando una media per quartiere, città, fascia d'età, ecc.

Anonimizzazione dei dati e reidentificazione.

A scanso di equivoci, l'anonimizzazione è una soluzione efficace, ma radicale, poiché comporta la perdita di molte informazioni potenzialmente utili. In ogni caso, è inevitabile che l'aggregazione e l'anonimizzazione portino ad allontanarsi dal dato grezzo, poiché rimuovendo l'informazione sull'identità dell'individuo si elimina il punto di contatto, la chiave comune tra i dati menzionati poco sopra; ovvero, la riconducibilità al singolo individuo che ha una certa convinzione religiosa, un certo reddito, una certa età e vive in un certo luogo. Va del resto ricordato che l'anonimizzazione non può essere considerata come una soluzione sempre sicura e idonea a proteggere il titolare dei dati anonimizzati. Esiste, infatti, la possibilità che i **dati anonimizzati siano reidentificabili** attraverso la loro combinazione con altri dati; in conseguenza di un tale procedimento, i dati anonimizzati non sono più anonimi e, pertanto, torneranno ad essere attratti sotto il regime della normativa sui dati personali.

Quest'eventualità deve esser fronteggiata predisponendo una clausola o in alternativa (come anche suggerito dal Working Party Art. 29 nella già citata Opinion 06/2013) un chiaro **messaggio di avvertenza** ai riutilizzatori sul portale dei dati aperti che avvisi che non possono essere attuate riaggregazioni di dati anonimizzati al fine di aggirare il divieto di reidentificazione. Ed, infatti, il Garante ha specificato che, nel caso di clausole che riguardano il riutilizzo di dati anonimi derivati da dati personali occorrerebbe "vietare ai titolari delle licenze di re-identificare gli interessati e di assumere qualsiasi decisione o provvedimento che possa riguardarli individualmente sulla base dei dati personali così ottenuti, nonché prevedere in capo ai medesimi titolari l'obbligo di informare l'organismo pubblico nel caso in cui venisse rilevato che gli individui interessati possano essere o siano stati re-identificati".

1.4.4 Formati e formalismi

Per valutare e migliorare i progetti di apertura dei dati, [Tim Berners-Lee](#), co-inventore del World Wide Web e tra i principali promotori del movimento Open Data, propone di classificare i progetti Open Data secondo il **livello tecnico di condivisione dei dati**. Per quantificare la progressione, ha proposto il "[5 star rating scheme](#)", un modello a 5 livelli, corrispondenti a "stelle", definiti come segue:

La scala di riferimento: le cinque stelle di Berners-Lee.

- ★ Mettere a disposizione i dati sul Web (in qualsiasi formato, "così come sono") con licenza aperta (per esempio, in formato immagine o PDF).
- ★★ Metterli a disposizione come dati strutturati in un formato "machine readable" (per esempio, in formato Excel, anziché come immagini scannerizzate di una tabella) con licenza aperta.
- ★★★ Utilizzare formati "machine readable" non-proprietari/aperti (per esempio, CSV,

anziché Excel), sempre con licenza aperta.

★★★★ Oltre alle condizioni dei livelli precedenti, usare gli standard del W3C (RDF e SPARQL) e URI (stabili) per identificare ogni cosa a cui i dati si riferiscono, così che altri possano creare puntatori ai tuoi dati (ad esempio, http://it.dbpedia.org/resource/Uniform_Resource_Identifier).

★★★★★ Oltre alle condizioni dei livelli precedenti, collegare i tuoi dati ad altri dati per fornire un contesto ([Linked Data](#)).

Da questo modello si può desumere facilmente che, oltre che esserlo dal punto di vista giuridico, i dati aperti devono essere riutilizzabili anche tecnicamente.

In particolare, il secondo livello di apertura si concentra sull'importanza di distribuire i dati secondo un formato strutturato (es. una tabella di Excel) e leggibile dalle macchine. Concettualmente il passaggio dal primo livello al secondo livello corrisponde al passaggio dal testo, in cui i dati non sono riconoscibili, ai dati strutturati. La struttura del dato è lo schema con il quale i dati possono essere immagazzinati ed organizzati, in modo tale da renderli **processabili automaticamente**.

Dati processabili automaticamente

I dati processabili automaticamente sono [dati](#) (o [metadati](#)) esposti in un formato che li renda interpretabili da un [computer](#).

Vi sono, da un lato, dati interpretabili dagli esseri umani che sono anche elaborabili da computer, in particolare mediante [marcatori](#) o espressioni codificate, come ad esempio i [microformat](#) o le estensioni [RDFa](#). Dall'altro, esistono formati di file di dati espressamente intesi per essere elaborati da computer, come ad esempio [RDF](#), [XML](#), [JSON](#).

Processabile automaticamente non è sinonimo di *digitalmente accessibile*. Un documento può essere digitalmente accessibile online, il che facilita l'accesso allo stesso da parte di umani via computer; tuttavia, se i dati contenuti nel documento non sono processabili automaticamente, sarà molto più difficile usare un computer per estrarli, trasformarli ed elaborarli. Per chiarire tale distinzione si pensi, ad esempio, a un documento [PDF](#), il quale è sì digitalmente accessibile, tuttavia l'informazione in esso contenuta è difficilmente processabile da una macchina.

Per esemplificare, si può immaginare di strutturare un dato che originariamente non lo era:

- **Dato non strutturato:** *La Primavera è un dipinto a tempera su tavola (203x314 cm) di Sandro Botticelli, databile al 1482 circa. Realizzata per la villa medicea di Castello, l'opera d'arte è attualmente conservata nella Galleria degli Uffizi a Firenze.*
- **Dato strutturato:**

Attributo	Valore
Titolo dell'opera	Primavera
Autore	Sandro Botticelli
Tipo	Dipinto a tempera su tavola
Dimensione	203x314 cm
Data di realizzazione	1482 circa
Destinazione originaria	Villa medicea di Castello
Città di conservazione	Firenze

Luogo di conservazione	Galleria degli Uffici
------------------------	-----------------------

Non sempre l'informazione può essere così strutturata, ma organizzare i dati in questo modo permette un riuso più semplice e comparazioni più utili tra dati della stessa tipologia.

Lo scenario migliore consiste nel predisporre il download – auspicabilmente gratuito - dei dati e che gli stessi possano essere **scaricati in blocco**, possibilmente come un unico archivio compresso. Inoltre, a scampo di equivoci, rendere i dati disponibili solo tramite servizi Web, ad esempio in pagine con menù a discesa che restituiscano alcuni dati sulla base delle scelte dell'utente, non costituisce una modalità adeguata per il riutilizzo (sebbene il servizio in quanto tale possa essere utilissimo, se affiancato al download in blocco).

Il terzo livello di apertura si riferisce alla riduzione delle possibili barriere tecnologiche date da **formati non aperti** che richiedano la disponibilità di specifico software non aperto per essere letti (quindi si consiglia l'uso di un file .CVS piuttosto che un file .XLS).

Formati aperti e machine-readable

Come previsto dall'art. 68 del CAD e più volte ricordato dall'Agenzia per l'Italia Digitale, il formato dei dati si definisce "aperto" quando ne viene resa pubblica, mediante esaustiva documentazione, la sintassi, la semantica, il contesto operativo e le modalità di utilizzo. Tali informazioni, unitamente ad una guida all'uso del formato, orientata alla lettura da parte dell'utilizzatore, devono essere presenti in uno o più documenti rilasciati dall'ente proponente lo standard.

Tuttavia, i formati aperti sono una condizione necessaria (e comunque assai raccomandabile) ma non sempre sufficiente: il livello minimo è l'adozione di un formato aperto e **leggibile in modo automatico dalle macchine (machine-readable)**, dal momento che la possibilità di processare in automatico i file, a volte, facilita il riutilizzo anche più di un formato aperto. Esistono grandi quantità di dati in formati standard "de facto", ovvero formati non aperti ma ugualmente leggibili e visualizzabili con software proprietari e open source. Per esempio, una tabella di Microsoft Excel (formato XLS) non è aperta in senso tecnico, ma è machine-readable e, a fini pratici, risulta infinitamente più utile di un PDF, che è un formato aperto da un certo punto di vista (nel senso che chiunque può creare applicazioni che leggono e scrivono file PDF senza pagare i diritti alla Adobe Systems), ma dal quale può essere molto difficile estrarre dati ben organizzati. Alcuni esempi di ottimi formati machine-readable (ed aperti) sono l'XML e il CSV (Comma Separated Value).

Inoltre, esperienze concrete suggeriscono che qualora la distribuzione dei dati su Internet avvenga attraverso protocolli di rete, anche questi siano aperti: esempi possono essere HTTP, FTP, WEBDAV e Torrent (si vedano in questo senso le [Linee guida per il riutilizzo e la diffusione dei dati pubblici](#) adottate dalla Provincia Autonoma di Trento).

In estrema sintesi: **l'apertura di un dato in senso tecnico riguarda il formato e il protocollo** con cui questo è diffuso; a ciò si deve aggiungere il profilo giuridico che coinvolge la licenza con cui è condiviso.

Quando un dataset raggiunge il livello delle tre stelle, ci si trova davanti alla minima buona pratica di Open Data in senso tecnico.

1.4.4.1 I Linked Data

Grazie all'impetuoso e inarrestabile accumulo di immagini, testi e contenuti multimediali, il World Wide Web che utilizziamo oggi è uno strumento dai benefici inconfutabili, atto a veicolare informazioni eterogenee e definibile come il più vasto archivio della conoscenza umana.

Sino ad ora, tuttavia, il suo potenziale non è ancora stato sfruttato appieno: l'informazione condivisa al suo interno, infatti, è costituita da documenti complessi intrecciati tra loro mediante collegamenti ipertestuali, che tuttavia non esplicitano la relazione tra le diverse pagine e concetti.

*Il Web of Data o
Web semantico.*

Chiamando di nuovo in causa Tim Berners-Lee, “per un computer, il Web è un mondo noioso, piatto, privo di significato” (citazione originale: “*To a computer, then, the Web is a flat, boring world devoid of meaning*”, [WWW](#)

[Plenary Meeting, Ginevra, 1994](#)). Un file CSV condiviso nel Web può infatti essere molto utile, ma non è altro che una monade isolata e priva di semantica, soprattutto se processato automaticamente da una macchina.

La nuova sfida si chiama “Web of Data”, o più esattamente “Web semantico”, evoluzione del “Web of documents”, e consiste nel processo volto a strutturare i contenuti messi a disposizione sul Web, arricchendoli della loro connotazione semantica e collegandoli con altre informazioni, al fine di creare nuova conoscenza. Nel “Web of data” Berners-Lee aspira che “i computer diventino in grado di analizzare tutti i dati sul Web - il contenuto, i link, e le transazioni tra le persone e computer” (da “Weaving the Web” di Berners-Lee e Fischetti, HarperSanFrancisco, 1999, ISBN 9780062515872). Per ottenere ciò non basta che i formati sul Web siano “machine-readable”: occorre che siano “machine-comprehensible”.

Secondo Berners-Lee quindi, ma anche secondo la maggior parte dei commentatori ed esperti del settore, i dati aperti generano il loro vero valore quando sono accompagnati da semantica e sono collegati tra loro, ovvero quando sono anche “linked data”: la loro utilità aumenta sensibilmente quando diversi insiemi di dati, prodotti e pubblicati in modo indipendente da diversi soggetti, possono essere incrociati liberamente anche da terze parti. Il Web of Data pertanto rappresenta una estensione e un miglioramento della rete esistente.

La filosofia dei Linked Data dimostra come la conoscenza sia un fattore incrementale e condiviso, e che derivi da quella “intelligenza collettiva” che il Web fin da principio si prefiggeva di abilitare. Questa è la direzione verso cui l'intero Web si sta dirigendo: da uno spazio globale di informazione costruito da documenti granitici ad uno scenario in cui documenti e dati sono collegati tra loro in una dimensione strutturale e partecipata.

1.4.4.2 La quinta stella

I dati aperti, connotati del significato semantico e connessi tra loro, Linked Open Data, sono certamente più impegnativi da mettere a disposizione rispetto ai dati “così come sono”, ma rispettare alcune semplici norme di comportamento potrebbe permettere un enorme salto di qualità per i riutilizzatori.

*URI stabili: un link per ogni
entità o concetto.*

Come già introdotto sopra, il riutilizzo è facilitato dal collegamento dei dati tra loro, stabilendo un link diretto se e quando i dati (eventualmente provenienti da diverse sorgenti) si riferiscono a oggetti identici o

comunque relazionati tra loro. Usare a questo scopo gli **standard riconosciuti, accettati e condivisi** del W3C, l'ente di certificazione del Web, permette di abilitare il modello a quattro stelle: **URI, RDF e SPARQL**.

Questo approccio è teso a spostare l'attenzione dal dato in quanto tale all'entità a cui il dato si riferisce, cioè all'oggetto fisico, astratto o digitale che il dato rappresenta.

*I Linked Data: dati strutturati,
collegati da link con una
semantica esplicita.*

La quinta stella si raggiunge quando dataset rispondenti al modello a 4 stelle, che richiede l'adozione di URI, RDF e SPARQL, sono **collegati a dataset di altre organizzazioni**, che adottano la medesima tecnologia.

Questo permette di raggiungere un miglioramento a due dimensioni nella conoscenza resa disponibile sul Web: orizzontalmente, poiché il dato è meglio contestualizzato, ad esempio perché connesso ad altri dati che hanno qualche affinità con l'entità descritta; verticalmente, perché si aumenta la quantità di dati che descrivono la medesima entità.

Per chiarire meglio le due dimensioni, supponiamo di avere, da una parte, amministrazioni locali che pubblicano dati aperti relativi ai monumenti storici e agli hotel che si trovano nelle vicinanze di quei monumenti; dall'altra, Sovrintendenze ai beni culturali che pubblicano dati dettagliati sui

monumenti, gli artisti e i periodi storici, e sui quadri esposti nei musei o nei palazzi.

Considerando, invece, la dimensione verticale del collegamento dei dati possiamo immaginare che più dataset relativi alla medesima opera d'arte siano rilasciati da diverse strutture pubbliche e private. Le Sovrintendenze, ad esempio, condividono alcuni dati relativi allo stato di conservazione dell'opera, mentre la storia, le informazioni sull'autore e l'anno di produzione sono collezionate da Wikipedia (o dal suo corrispondente già in formato “linked”, [DBPedia](#)).

Tutti questi dati, una volta connessi tra loro e aggregati intorno alla medesima entità, permettono di **aumentare esponenzialmente la quantità di informazioni** che un utente può navigare e riutilizzare, ma soprattutto la disponibilità delle ontologie che accompagnano i dati abilitandone la loro portata semantica, ne consente il riuso anche da parte di applicazioni e sistemi. Combinare tutti questi dataset potrebbe essere di grande utilità per l'imprenditoria locale che, ad esempio, potrebbe costruire applicazioni per offrire un servizio personalizzato sugli itinerari turistici in base agli interessi culturali specifici dell'utente.

Al fine del “collegamento” dei dati va sottolineato l'aspetto assolutamente centrale dell'utilizzo di rappresentazioni del dominio (**ontologie**) che esplicitino, in modo formalizzato, la **semantica dei dati pubblicati**. Solo in questo modo si potranno utilizzare tali informazioni per collegare dati che si riferiscono, in contesti diversi, alla medesima entità.

In conclusione, possiamo definire il paradigma de Linked Open Data come un insieme di principi, tecnologie e buone pratiche per connettere tra loro dati strutturati nel Web, resi machine-readable, in cui i collegamenti e le relazioni tra le entità descritte nei dataset sono resi espliciti e le informazioni relative alla medesima entità sono collezionate in modo collaborativo, e resi inoltre **machine-comprehensible** attraverso l'esposizione di ontologie che ne esplicitano il livello semantico, consentendo l'arricchimento della conoscenza attraverso meccanismi inferenziali.

Il paradigma Linked Open Data mira, dunque, alla creazione di una “nuvola” definita “**linked open data cloud**” che sostituisce la miriade di isole di repository e dataset esistenti con un singolo gigante aggregato di conoscenza distribuita in continuo movimento ed evoluzione. Nel lungo periodo, questo approccio cooperativo permetterà di trasformare il Web in uno spazio distribuito di entità che ogni persona, amministrazione o azienda potrà contribuire a migliorare, con l'obiettivo di aumentarne l'informatività e la qualità.

Seguire l'approccio fin qui delineato offre sicuri benefici per utenti e sviluppatori. Gli utenti potranno navigare tra entità specifiche in modo più semplice rispetto ad inviare complesse interrogazioni a grandi database. Allo stesso modo, gli sviluppatori potranno investire meno tempo nella pulizia, nella bonifica e nel collegamento dei dati e potranno concentrarsi sul miglioramento dell'idea e del servizio offerto.

1.4.4.3 Le tecnologie in gioco

Per realizzare il Web semantico e raggiungere il livello “5 stars” dello [schema di rating](#) degli Open Data, è necessario seguire le regole stabilite dal W3C:

[URI e HTTP](#)

- usare URI per identificare le risorse pubblicate,
- usare il protocollo HTTP per “raggiungere” quelle risorse sul Web,
- usare gli standard RDF e SPARQL per descrivere le risorse ed interrogarle,
- includere collegamenti ad altre URI per accrescere il patrimonio di conoscenza iniziale.

Gli **URI** (Uniform Resource Identifiers) – o, seguendo le specifiche più recenti ed attente al multilinguismo, gli **IRI** ([Internationalized Resources Identifiers](#)) – sono identificatori univoci per gli “oggetti” presenti sul Web ed hanno lo scopo di disambiguare il significato delle risorse. In

particolare, la locuzione Uniform Resource Identifier si riferisce a una stringa che identifica univocamente una risorsa generica, come ad esempio un indirizzo Web, un documento, un'immagine, un file, un servizio, un indirizzo di posta elettronica. La **disambiguazione** (in inglese **Word Sense Disambiguation** o, abbreviato, **WSD**) è l'operazione con la quale si precisa il significato di una parola o di un insieme di parole (frase), che denotano significati diversi a seconda dei contesti e che quindi sono ambigue.

Un tipo particolare di URI, molto utilizzato, sono gli URL (Uniform Resource Locator), familiari a chiunque usi il Web, che sono gli indirizzi che compaiono nella barra di navigazione del browser e cominciano con "http://". Il protocollo **HTTP** (HyperText Transfer Protocol) permette di raggiungere le risorse tramite la rete Internet in modo semplice.

RDF, il Resource Description Framework, è il paradigma per esporre informazioni sul Web dei dati.

Gli indirizzi URI e il protocollo HTTP sono funzionali ad una terza tecnologia fondamentale per il Web of Data: **RDF** (Resource Description Framework). Questo strumento è il principale standard di riferimento per la modellazione, la codifica, lo scambio e il riutilizzo di dati collegati in modo da essere massimamente interoperabili. Non si tratta propriamente

di un formato di dati, bensì di un formalismo per rappresentare dati – un “data model” – basato sul concetto di **tripla**. Le triple RDF sono asserzioni rappresentabili anche in forma di grafo, **composte da soggetto, predicato e oggetto**. Attraverso le triple RDF si rappresentano le risorse, le loro classi e proprietà e i rispettivi valori. Ad esempio, una tripla RDF può definire che il quadro "Primavera" (soggetto) è dipinto (predicato) dal pittore "Botticelli" (oggetto) (livello estensionale). Soggetto ed oggetto della tripla sono entrambi URIs che identificano due risorse. Il predicato descrive come soggetto e oggetto sono connessi ed anch'esso un URI. Al tempo stesso una tripla RDF può asserire che i “pittori” “dipingono” i “quadri” (livello intenzionale). Parimenti, con una tripla RDF si asserisce che “Botticelli” è un “pittore”.

L'esempio appena portato è solo uno degli scenari possibili in cui i Linked Open Data possono favorire l'interoperabilità e l'aggregazione tra dataset. Le possibilità sono infinite se pensiamo alla vasta quantità di Linked Open Data già presenti sul Web.

[DBPedia.org](http://dbpedia.org), per esempio, espone una grande porzione di dati di Wikipedia.org come Linked Open Data. Esiste anche it.dbpedia.org, che raccoglie e struttura dati provenienti dalle pagine Wikipedia in italiano. [Geonames.org](http://geonames.org) offre descrizioni RDF di entità geografiche. [Linkeddata.org](http://linkeddata.org) fornisce un quadro dello stato corrente della “Linked data cloud”, e mostra un ecosistema di database interconnessi in rapida crescita.

Ciascuno di questi database fornisce dati che possono essere linkati e che possono aggiungere valore ad altri e nuovi dati.

Proviamo a trasformare un dato strutturato tabellare in triple RDF, riferendoci alla versione italiana di DBPedia.org:

- Dato strutturato riguardante la Primavera di Botticelli:

<i>Attributo</i>	<i>Valore</i>
<i>Autore</i>	<i>Sandro Botticelli</i>
<i>Luogo di conservazione</i>	<i>Galleria degli Uffizi</i>

- Triple RDF:
 - Prima tripla:
 - Soggetto: [http://it.dbpedia.org/resource/Primavera_\(Botticelli\)](http://it.dbpedia.org/resource/Primavera_(Botticelli))
 - Predicato: <http://it.dbpedia.org/ontology/author>
 - Oggetto: http://it.dbpedia.org/resource/Sandro_Botticelli
 - Seconda tripla:

- Soggetto: [http://it.dbpedia.org/resource/Primavera_\(Botticelli\)](http://it.dbpedia.org/resource/Primavera_(Botticelli))
- Predicato: <http://it.dbpedia.org/ontology/location>
- Oggetto: http://it.dbpedia.org/resource/Galleria_degli_Uffizi

I Linked Open Data presuppongono formati e modelli standard, interoperabili, machine-readable e ontologie di riferimento, al fine di rendere possibile la massima integrazione tra fonti diverse di conoscenza. Per fare ciò servono standard, non necessariamente imposti centralmente, ma comunque condivisi o realizzati in modo tale da minimizzare gli sforzi per raggiungere l'interoperabilità anche nella definizione delle ontologie e dei vocabolari di riferimento.

SPARQL, acronimo di SPARQL Protocol and RDF Query Language, è un linguaggio di interrogazione e un protocollo per i dati rappresentati con RDF. Le query SPARQL, molto simili a quelle SQL, sono inviate da un client ad un server (l'end-point SPARQL) usando il protocollo http. L'interazione tra client ed end-point avviene attraverso una pagina Web che contiene un form dove poter scrivere le query e lanciarle. Il risultato di una query può avere diversi output, quali ad esempio HTML, CSV, JSON, RDF/XML, TURTLE.

SPARQL, il linguaggio per interrogare il Web dei dati

Gli strumenti necessari sono chiamati “**vocabolari**”: come nel mondo analogico, si tratta di collezioni di termini e delle relative definizioni pensati per risolvere l'ambiguità tipica del linguaggio naturale. Definire la terminologia con cui si individuano classi di entità e le proprietà delle stesse aiuta ad aggregare i dati in modo preciso. Condividere, utilizzare e continuamente migliorare vocabolari comuni permette di rendere interoperabili tra loro i datasets: si tratta di un passaggio sempre più necessario per riuscire ad aggregare - e dunque estrarre valore - dai datasets condivisi nel mondo aperto del Web.

Vocabolari

Gli strumenti tecnici che permettono di creare questi vocabolari sono RDFs (Resource Description Framework Schema) e OWL (Web Ontology Language).

L'**ontologia** è una rappresentazione condivisa ed esplicita di un dominio di interesse. A differenza di vocabolari, le ontologie consentono di esprimere con maggior dettaglio le relazioni presenti tra le entità che costituiscono il dominio e i vincoli che queste devono soddisfare. Le ontologie meglio si prestano a descrivere i domini applicativi dai quali sono tratti i dati. Il linguaggio utilizzato per descrivere le ontologie è principalmente OWL (Web Ontology Language) poiché offre una espressività migliore rispetto a RDF Schema.

Ontologie e OWL

La costruzione delle ontologie richiede una buona capacità di modellazione, cioè di rappresentazione logica, da parte dell'esperto di dominio. Tramite essa si attribuisce un significato univoco alle informazioni descritte. Poiché tali informazioni potranno poi essere riprese e collegate da altre organizzazioni è di tutta evidenza quanto la stabilità, la coerenza e la credibilità di tali schemi costituisca un valore che può determinare il successo o l'insuccesso del dataset all'interno della nuvola. Di nuovo: l'utilizzo di schemi concettuali condivisi e affermati facilita l'assolvimento di tali requisiti rendendo la fase di concettualizzazione del dominio un sicuro punto cruciale nella costruzione della catena del valore legata al dato. [Dublin Core](#), [Friend Of A Friend](#) e [Schema.org](#) sono alcuni degli standard più conosciuti ed utilizzati.

Infine, val la pena mettere in guardia contro la chimera rappresentata dal “progetto perfetto”. Quando è possibile collegare i propri dati a semplici dizionari standard di riferimento come quelli menzionati sopra, è sicuramente opportuno farlo. Lo stesso vale per l'uso di metadati che rappresentino lo standard in un certo settore scientifico o disciplinare (ad es. lo standard [SDMX](#) nel campo dei dati statistici). Ma sarebbe alquanto pericoloso e controproducente ritardare la messa a disposizione di alcuni dati, attendendo di aver individuato gli standard di riferimento ideali per “linkarli” in modo opportuno.

Allo stesso modo, l'esperienza di molti informatici che hanno avuto modo di collaborare con le pubbliche amministrazioni pare concorde nel mettere in guardia rispetto a progetti che puntino a

realizzare “ontologie” della PA capaci di rappresentare in modo standard ed esaustivo la pletora di complesse attività, ruoli ed entità che le PA trattano o con cui entrano in contatto.

Un'ottima analisi delle tecnologie e dei progetti che hanno iniziato a sperimentarne i benefici è stata realizzata dall'Agenzia per l'Italia Digitale nel contesto delle già citate [Linee guida per l'interoperabilità semantica attraverso i linked open data](#).

Triplestore

Il processo di trasformazione dei propri dati in triple RDF definisce uno o più grafi di relazioni tra entità, che vengono poi inseriti in un **triplestore** ovvero un repository specifico per la memorizzazione e la ricerca di dati RDF. Per interrogare i dati inseriti nel triplestore, viene fornito un **end-point SPARQL**, ovvero un indirizzo Web dove poter effettuare le query SPARQL. Questo strumento permette anche di interrogare contemporaneamente diversi triplestore non formalmente collegati tra loro, estraendo così informazioni da diversi archivi condivisi nel Web. Anche qui è necessario un avvertimento: l'accesso tramite end-point SPARQL è necessario, ma non dovrebbe essere l'unico metodo di fruizione e consumo dei dati in quanto – per funzionare correttamente - richiede un certo livello di conoscenze tecniche. È quindi sempre opportuno affiancare a questo strumento tecnico, una forma di visualizzazione grafica delle triple RDF pubblicate. Due buoni esempi sono [LodLive](#), prodotto “made in Italy”, e [RelFinder](#). Altri strumenti imprescindibili sono i frontend per SPARQL end-point, navigabili sia da parte di esseri umani che da agenti automatici, come il noto [Pubby](#), standard del settore, o l'ottimo e più ricco [LodView](#), dello stesso team di LodLive e finalista del prestigioso Challenge LODLAM al [Summit2015](#).

1.4.5 Piattaforme Open Data

1.4.5.1 L'obiettivo dell'“Open by design”

Open by Design: un ciclo di vita dei dati pensato con l'apertura in mente!

In relazione ai processi di apertura di dati delle PA, numerosi addetti ai lavori, ma anche semplici osservatori, menzionano quale principio ordinatore da perseguire il cosiddetto “**Open by design**”, ossia l'insieme di meccanismi e protocolli da applicare a un sistema informativo volti ad

assicurare la disponibilità - in automatico e senza la necessità di interventi ad hoc - dei dati pubblici per il loro libero riutilizzo. Tali meccanismi riguardano, di fatto, tutti gli aspetti della catena di generazione, archiviazione, aggiornamento e rilascio dei dati che una PA si trova a dover gestire nell'ambito del suo mandato; così da raggiungere l'obiettivo di avere dati “nativamente” aperti, ossia entro un modello di governo che preveda che l'informazione gestita da una PA sia sempre accessibile - eventualmente con diversi livelli di privilegio a seconda del tipo di soggetto - e non secondo rilasci discrezionali e a frequenza variabile (e/o non dichiarata). In quest'ottica, da un lato, è necessario un ripensamento degli stessi sistemi informativi interni delle PA, nell'ottica di concepire **l'apertura come un aspetto cardine del più ampio ciclo di vita del dato**. Ad esempio, si pensi al fatto che gli stessi flussi informativi delle PA (e, in generale, delle grandi organizzazioni) sono spesso compartimentati, e il più delle volte sono “verticali” (ossia, tra i diversi livelli gerarchici di un medesimo dipartimento / ufficio) piuttosto che “orizzontali” (ossia, tra due differenti divisioni della medesima PA). Dall'altro lato, considerando il rapporto tra le PA e riutilizzatori esterni, il concetto sopra espresso pare in linea con l'adozione, tra l'altro, di una politica tecnologica evoluta per la messa a disposizione in via automatizzata degli insiemi informativi, secondo un approccio del tipo “[Everything should be an API](#)”.

1.4.5.2 Criteri e funzionalità

Piattaforme software per gli open data

Ridisegnare in ottica “aperta” i flussi informativi di una PA rappresenta, per la sua complessità, un obiettivo ragionevole a medio/lungo termine. Nel breve termine, è invece alla portata di ogni amministrazione volenterosa

dotarsi di efficaci strumenti per l'esposizione di dati aperti. Nella pratica, si tratta di strumenti

software (adattabili a seconda delle necessità) che consentono di interagire con i sistemi informativi interni, estrarre informazione (ad esempio, interrogando un database) secondo regole di ingaggio e livelli di automazione stabiliti, e pubblicare i relativi dati, tipicamente mediante un front-end dedicato.

In particolare, possibili **insiemi di funzionalità** (o criteri funzionali) utili per valutare le caratteristiche di una piattaforma software per la pubblicazione di dati aperti sono i seguenti:

- il **processo** che consente la pubblicazione dei dati, ad esempio:
 - l'esistenza o meno di un collegamento automatizzato con i database di applicazioni in uso interno alla PA;
 - l'uso di [Application Programming Interfaces \(API\)](#) per la trasformazione -ad es., in termini di formato- dei dati, e il loro riutilizzo;
 - l'impiego di metadati standard;
 - l'accoppiamento con un front-end / portale;
- le **tipologie standard di dati** che la piattaforma consente di pubblicare, ad esempio:
 - dati in tempo reale, o comunque ad alta frequenza di aggiornamento;
 - dati geospaziali, e dati geo-riferibili;
 - linked data;
- l'architettura della piattaforma, e altre **caratteristiche generali** come:
 - il tipo di licenza associata, free/open-source, o proprietaria;
 - (collegato al punto precedente) il grado di adattabilità / personalizzazione;
 - l'eventuale disponibilità in [cloud](#);
- **altri aspetti**, ad esempio
 - la possibilità, per gli utenti, di generare e condividere versioni aggiornate (ad es., corrette, arricchite, ecc.) dei dataset;
 - la disponibilità di un servizio di assistenza per gli utenti.

1.4.5.3 Alcune soluzioni esistenti

Tra le soluzioni esistenti, è possibile citare:

- piattaforme **open-source**, come ad esempio:
 - il [Comprehensive Knowledge Archive Network \(CKAN\)](#), sviluppata dalla Open Knowledge Foundation, che viene usato, tra l'altro, come back-end per il portale di dati aperti del Regno Unito <http://data.gov.uk/>. Il **Comprehensive Knowledge Archive Network (CKAN)** è una sistema open source e basato sul Web per l'immagazzinamento, la catalogazione e la distribuzione di dati, quali ad esempio fogli di calcolo o contenuti di database.;
 - [NuData](#), lanciata a giugno 2014, disponibile in cloud modalità [SaaS \(Software as a service\)](#);
- piattaforme **proprietarie**, quali [Socrata](#), impiegata in Italia, ad esempio, per il [portale di dati aperti della Regione Lombardia](#). (Socrata dispone anche di una [versione "community"](#), rilasciata in open-source).

1.4.5.4 La scelta di una piattaforma (o di singoli moduli)

Una prima distinzione possibile tra le piattaforme riguarda, da un lato, la disponibilità o meno del **codice sorgente**. Questo aspetto può rendere più o meno facile la personalizzazione del software, e l'integrazione con altri moduli al fine di conseguire finalità specifiche. La scelta dipende anche dalla valutazione del [Total Cost of Ownership](#) legato all'adozione di una piattaforma, che è a sua volta funzione di

*Disponibilità del sorgente e
altre caratteristiche funzionali*

numerosi parametri, non solo relativi al costo diretto (monetario) di adozione. In quest'ottica va comunque ricordato che l'art. 68 del Codice dell'Amministrazione Digitale prevede che un'amministrazione effettui un'analisi comparativa delle soluzioni secondo i principi di economicità e di efficienza, tutela degli investimenti, riuso e neutralità tecnologica, *a parità di condizioni economiche, funzionali e relative alle garanzie del fornitore*, riutilizzando soluzioni già disponibili all'interno della PA, raccolte dall'Agenzia per l'Italia Digitale in un [catalogo nazionale](#).

Altri importanti fattori sono le **caratteristiche funzionali** - i cui "pesi" relativi dipendono in larga parte dalle necessità della singola Pubblica Amministrazione - quali ad esempio:

- l'erogazione del servizio in cloud, e/o in locale (con pro e contro ampiamente dibattuti);
- le modalità di estrazione / copia dei dati dai propri sistemi interni, ai fini di una loro esposizione;
- la possibilità di federazione tra diversi cataloghi, come ad esempio nel caso di CKAN;
- le varie possibilità di manipolazione dei dati per l'utente, come la navigazione al livello del singolo dato, o l'abilitazione alla creazione di elaborazioni personalizzate;
- l'integrazione con un front-end: nel caso di Socrata, questa avviene nativamente (il che ha pro -facilità d'uso- e contro -minore capacità di personalizzazione-, a seconda degli obiettivi).

In linea generale, è opportuno segnalare che l'adozione di piattaforme che integrano più moduli e funzionalità non è, di per sé, l'unica possibilità. Di fatto, è possibile -e probabilmente consigliabile- valutare caso per caso anche l'impiego di **singoli tool**, in particolare open-source. Naturalmente, molto dipende non solo dalle necessità della singola amministrazione, ma anche dalla sua capacità di gestire l'interoperabilità con i propri sistemi interni (e dagli eventuali vincoli che questi pongono). Moduli specifici vanno, ad esempio, dai connettori per l'estrazione di dati dai database legacy (ad esempio, usando [Teiid](#)) ai software per la creazione e gestione API (come quello fornito da [WSO2](#)), fino a strumenti per la verifica e il miglioramento della qualità dei dati (come [OpenRefine](#)). Nel caso dell'esposizione di dati linked, alcune delle piattaforme summenzionate abilitano tale processo, ma è anche possibile dotarsi, "semplicemente", di strumenti ad hoc, ossia, come configurazione di base, un [triple store RDF](#), un [end-point SPARQL](#), con l'accortezza di adottare opportuni strumenti a corredo, ad esempio per la riconciliazione tra ontologie.

Piattaforme: alcuni approfondimenti utili

Un'ampia discussione in merito alle possibili funzionalità di piattaforme per l'esposizione di dati aperti è presente in [Zuiderwijk, A., Janssen, M., & Jeffery, K. \(2013, May\). *Towards an e-infrastructure to support the provision and use of open data*. In Conference for E-Democracy and Open Government \(p. 259\). Di interesse, nel medesimo filone, \[Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Sheikh_Alibaks, R. \\(2012\\). *Socio-technical impediments of open data*. Electronic Journal of e-Government, 10\\(2\\), 156-172.\]\(#\)](#)

Un primo tentativo di confronto tra alcune delle piattaforme per la messa a disposizione di dati pubblici (CKAN, Socrata ed altre) è presente in un articolo presentato da [R. Iemma, F. Morando e M. Osella](#) alla International Conference for E-Democracy and Open Government 2014 (e vincitore del "Best paper" award), disponibile all'indirizzo <http://nexa.polito.it/iemma2014open>. Un tentativo di maggiore formalizzazione di tale confronto (estraendo requisiti da documenti di policy) è disponibile nella wiki semantica, in via di popolamento, disponibile all'indirizzo <https://opendataplatforms.nexacenter.org/>.

1.4.5.5 "Raw data now"?

"Raw Data Now!" e basi dati complesse.

La breve analisi qui presentata si collega a una questione più ampia, di natura quasi filosofica, relativa al confine del mandato pubblico, e in particolare in ordine alla distinzione tra i compiti di chi espone di dati e quelli di soggetti esterni interessati a riutilizzarli. Da un punto di vista generale, pare auspicabile che le PA si dedichino esclusivamente a mettere a disposizione i propri dati in modo tale da garantire

che accesso ed elaborazione siano il più possibile agevoli, evitando manipolazioni ex-ante che comportino il rischio di ridurre il potenziale informativo dei dati, nonché la loro collegabilità ad altre fonti. Si tratta, in termini più diretti, del concetto di “[Raw data now](#)” più volte dibattuto nell'ambito della comunità Open Data. Nel caso di basi di dati complesse (si pensi, come semplici esempi, ai dati di spesa, ai bilanci, o anche agli stessi atti parlamentari), non si può ignorare esista un “divario cognitivo” tra i detentori originali dei dati, in questo caso una PA, e potenziali riutilizzatori. Senza una adeguata contestualizzazione, questi ultimi rischiano di non avere sufficienti strumenti interpretativi per cimentarsi in elaborazioni fruttuose. Esempi di questo tipo sono riscontrabili anche in ambito “linked data”, nei quali la navigazione stessa dei dati (del loro grafo) rappresenta uno strumento per comprenderne la struttura e individuare opportunità di interrogazione e collegamento. Non a caso, le stesse PA che espongono dati linked tipicamente descrivono query standard che possano fungere da esempio. Nell'ottica di ridurre l'asimmetria informativa tra PA e riutilizzatori, può rivelarsi strategico l'intervento di intermediari appartenenti alla società civile dotati di specifiche conoscenze di dominio (nonché tecnologiche), con il vantaggio di agevolare il riutilizzo dei dati senza tuttavia circoscrivere alla fonte i potenziali scopi ed esiti del riutilizzo.

1.4.6 Licenziare i dati

1.4.6.1 I diritti di proprietà intellettuale sui dati

Nella maggior parte dei casi è ragionevole presumere che su una banca di dati possano insistere alcuni diritti di proprietà intellettuale (in primis, il [diritto d'autore](#) e il [diritto sui generis](#)) e che sia necessario ottenere l'autorizzazione del titolare dei diritti per compiere la maggior parte delle operazioni di riutilizzo (estrazione di dati, riproduzione, aggiornamento, creazione di nuove banche dati che ne combinino di preesistenti, etc.). Di regola, il fatto che tali atti siano compiuti per fini non commerciali non legittima l'utilizzazione (il che vale in generale per tutti i materiali protetti diritto d'autore e dunque anche nel caso delle banche dati). In altre parole, il modello in cui ricadono automaticamente le banche di dati è quello “**Tutti i diritti riservati**”.

Pertanto, in considerazione dell'esclusività di tali diritti - che li rende non esercitabili da soggetti terzi in assenza di un'autorizzazione da parte del relativo titolare - diventa necessario porsi due domande: 1) può l'ente pubblico disporre legalmente dei dati in questione, usandoli liberamente ed eventualmente anche ri-licenziandoli come ritiene opportuno? 2) Se sì, com'è opportuno che l'ente si comporti nel gestire i propri diritti?

Aspetti di proprietà intellettuale.

La prima domanda riguarda la cosiddetta **rights clearance** e rientra tra quelle questioni tecnico-giuridiche che rendono assai opportuna la presenza di un giurista nella squadra operativa del proprio progetto open data. Per analizzare lo status giuridico di un database, sarà necessario rispondere a quesiti del tipo: da dove vengono i dati? Chi li ha raccolti? Con quali risorse? Ci sono (stati) contratti di qualche tipo che regola(va)no l'attività di raccolta e/o l'uso dei dati stessi? Inoltre, sarà opportuno chiedersi: parte di questi dati sono personali? O addirittura sensibili? Ci possono essere segreti aziendali od altri diritti di proprietà intellettuale di terzi? In prima approssimazione, si può ritenere che l'ente pubblico disponga sempre dei diritti di proprietà intellettuale necessari nel caso in cui la banca dati sia stata creata da propri dipendenti, con fondi dell'ente stesso.

Circa la seconda domanda; come abbiamo accennato, lo scenario base individuato dalla legge italiana sulla proprietà intellettuale e quello del “Tutti i diritti riservati”. In altre parole, in assenza di un'esplicita autorizzazione non sarebbe possibile impiegare i dati; al massimo, è permesso consultarli laddove reperibili online. Ovviamente, qualcuno potrebbe decidere di usare comunque i dati (anche perché potrebbe non essere a conoscenza del quadro giuridico che si è descritto); pertanto, se l'obiettivo di un ente pubblico è quello di aumentare in modo significativo l'uso delle risorse informative che possiede e **dato che l'incertezza sul regime giuridico adottato sui**

dati rischia facilmente di scoraggiarne il riutilizzo, è sempre opportuno associare ai dati un'esplicita licenza (o almeno una nota o liberatoria, che chiarisca la facoltà per chiunque di riutilizzare i dati liberamente).

1.4.6.2 Adottare una licenza aperta

Applicare una licenza.

Una licenza è un documento che descrive quali, tra le attività normalmente riservate al detentore dei diritti (il licenziante), possano essere compiute da chi riceve i dati sotto i termini della licenza stessa (il licenziatario).

Tipicamente, la licenza descrive alcune condizioni per godere dei diritti concessi. Coerentemente con la già discussa definizione di dati aperti (su cui vedi p. 6), una licenza aperta non può essere associata a condizioni troppo vincolanti. La scelta è limitata a tre tipologie di licenze:

1. quelle che non impongono **alcuna condizione**,
2. quelle che impongono **condizioni necessarie a documentare la provenienza del dataset**, o, eventualmente,
3. quelle che impongono **condizioni necessarie a mantenere il dataset aperto**.

Nella prima categoria rientrano varie forme di liberatorie e dediche al pubblico dominio. Tali strumenti sono pacificamente considerati idonei a rendere i dati "licenziati" aperti, a condizione che l'ordinamento giuridico consenta ai titolari una così ampia rinuncia ai loro diritti. Similmente, le condizioni legate alla documentazione della provenienza sono considerate accettabili dalla maggior parte della comunità dei dati aperti: in questo contesto, il titolare si limita a condizionare la licenza al riconoscimento di una specifica **attribuzione/origine/fonte dei dati** (ovviamente, con modalità che non traggano in inganno l'utilizzatore – ad esempio, un dataset modificato dovrà sempre essere identificato come tale e mai spacciato per l'originale). Al contrario, le condizioni necessarie a preservare l'apertura del dataset, in altre parole il requisito di condivisione allo stesso modo, presenta alcuni aspetti più complessi e delicati. Questa clausola (detta anche share-alike, [copyleft](#) o virale) consiste nell'obbligare il riutilizzatore a licenziare qualsiasi lavoro realizzato a partire dall'originale con la medesima licenza adottata per l'originale stesso (sicché chiunque altro resta a sua volta libero di apportare ulteriori modifiche e la catena di condivisione si prolunga, da cui il nome clausola virale). Per quel che riguarda i dati delle pubbliche amministrazioni, non tutta la comunità open data si trova d'accordo rispetto all'opportunità di assoggettare a questa clausola dati la cui creazione è stata già finanziata con denaro del contribuente.

Condivisione allo stesso modo

Spesso, i dati generati dalle community online sono rilasciati sotto licenze del tipo condividi-allo-stesso-modo (o, in inglese, share-alike). Questa è una **scelta di auto-difesa** della comunità, contro soggetti esterni che tentino di appropriarsi del lavoro della comunità stessa, mediante strategie che passino per la creazione di una versione modificata (e migliorata) del lavoro della community.

Per una comunità di utenti, che si unisce intorno all'idea di un progetto comune che rappresenta anche l'obiettivo attorno al quale si aggrega la comunità stessa, questa scelta può essere appropriata.

È più dubbio se lo stesso valga per un'amministrazione pubblica, che si ritrova a possedere dati generati come "effetto collaterale" di un suo compito istituzionale. In particolare, mentre lo scopo principale di una community online è quello di far progredire il progetto attorno al quale si è costituita, i dati pubblici sono messi a disposizione per rendere possibili riusi inattesi, spesso in combinazione con altri insiemi di dati. Per questo motivo, l'inserimento della clausola di condivisione-nello-stesso-modo è un'idea criticabile. Ogni licenza share-alike, infatti, è compatibile con se stessa e tendenzialmente incompatibile con altre licenze share-alike (ad esempio, una licenza [Creative Commons BY-SA](#) e una licenza [Open Data Commons Open Database License](#) - ODC ODbL - sono mutualmente incompatibili). Il risultato rischia di essere una sorta di "**Babele delle licenze**", in cui numerosi insiemi di dati non possono essere uniti e mescolati.

Se comunque si decidesse di adottare una licenza di tipo "condividi-allo-stesso-modo", una licenza [CC BY-SA](#) presenterebbe il vantaggio di essere lo standard dominante (ad esempio, perché utilizzata da Wikipedia e progetti collegati, come DBpedia). Tra le altre principali licenze contenenti una clausola condividi-allo-

stesso-modo, la [ODC ODbL](#) presenterebbe il vantaggio di essere più robusta in termini di possibilità di far valere con successo la clausola condividi-allo-stesso-modo in alcune situazioni, dal momento che è stata scritta appositamente per le banche di dati mentre le licenze CC sono più generiche. La [Italian Open Data License](#) (IODL) versione 2.0 presenterebbe invece il vantaggio di contenere un'esplicita clausola di compatibilità, che permette di mescolare dati ottenuti sotto la IODL con dati licenziati con una licenza compatibile (CC BY-SA o ODC ODbL), ri-licenziandoli con la seconda licenza. Forse, se proprio si vuole usare una licenza “virale”, la sua clausola di compatibilità rende la [IODL 2.0](#) la scelta preferibile, almeno in Italia.

Per applicare una licenza ad una base di dati distribuita online, la prassi è che basti **indicare la licenza che si intende adottare** nella pagina dalla quale si effettua il download; si noti che le licenze relative a diritti d'autore devono essere provate per iscritto, tuttavia, l'assenza di licenza equivale allo status “tutti i diritti riservati”, per cui un licenziatario che violi i termini della licenza indicata sul sito difficilmente potrà giovare del fatto che la licenza vada provata per iscritto. Questo aspetto, dunque, può essere ignorato, almeno in prima approssimazione. Esistono poi opportuni accorgimenti tecnici, per far sì che l'associazione tra un file e la licenza sia chiara anche alle macchine (quali, ad esempio, i motori di ricerca). Una soluzione particolarmente *user friendly*, ad esempio, è offerta da Creative Commons (CC): il percorso guidato del [license chooser](#) (selettore di licenze) di Creative Commons fornisce automaticamente un brano di codice da inserire nelle proprie pagine Web. Creative Commons ha inoltre sviluppato una specifica tecnica, [CC REL](#), per descrivere le informazioni della licenza in formato RDF.

Cap. 2. L'offerta di dati aperti parlamentari

2.1 Cosa si intende per Open Data parlamentari

Per poter dare una definizione di Open Data parlamentari occorre necessariamente adottare una convenzione, in quanto la conoscenza gestita in un parlamento è di per sé difficilmente circoscrivibile entro limiti predeterminati. L'ambito delle materie oggetto di attività parlamentare – sia che si tratti di funzione legislativa, sia che si tratti di funzione di indirizzo e controllo – è infatti estremamente ampio e potenzialmente illimitato, come emerso anche dalla considerazione delle attività dei Servizi Studi della Camera e del Senato, cui competono le attività di studio e di indagine preliminare sulle materie oggetto dell'attività parlamentare, con funzione di supporto decisionale ai parlamentari.

Ciò premesso, convenzionalmente si circoscrive l'ambito degli Open Data parlamentari entro i confini del dominio informativo costituito dalle **informazioni accumulate e prodotte in ambito parlamentare, che sono sicuramente attinenti, da un lato, alla struttura dell'istituzione** (organi e persone di cui si compone) **e dall'altro alla sua attività** (articolazione dei lavori parlamentari e relativi prodotti); tale dominio si estende però anche ad altri ambiti informativi, e in particolare a quelli a supporto dell'attività parlamentare e a quelli relativi alla gestione della struttura amministrativa. È questo effettivamente l'ambito informativo in cui attualmente non solo i due rami del parlamento italiano, ma anche i parlamenti esteri, stanno concentrando la produzione di Open Data.

Informazioni accumulate e prodotte in ambito parlamentare.

Sebbene già di per sé abbia fatto sì che il contributo nella produzione di Open Data da parte del Parlamento italiano sia ad oggi considerevole (oltre 500 milioni di triple pubblicate dalla Camera dei deputati e oltre 30 milioni di triple pubblicate dal Senato), questa convenzione deve essere auspicabilmente estesa affinché la collettività possa giovare di tutta la conoscenza gestita all'interno delle istituzioni parlamentari. Tutti i parlamenti, infatti, producono quotidianamente e trattano enormi volumi di informazioni che, se messi a disposizione in Open Data, possono contribuire ad un maggior coinvolgimento dei cittadini e ad una più profonda conoscenza dell'attività parlamentare.

2.2 Lo scenario italiano

Nella produzione di Open Data parlamentari è un fatto generalmente riconosciuto, anche a livello internazionale, che l'Italia si trovi in una situazione di avanguardia, paragonabile ad esempio a quella di Canada o Nuova Zelanda, pionieri in tale ambito, e con una posizione ragguardevole anche rispetto a USA e Regno Unito, tra i primi a stabilire strategie open data a livello nazionale.

L'**ambito informativo** coperto dagli Open Data parlamentari italiani è oggi estremamente ricco ed ampio. La massa di dati attualmente disponibile in rete, che conta oltre 530 milioni di triple RDF, si focalizza sulle informazioni relative alla struttura delle istituzioni e alla loro attività, ma si estende anche ad altri ambiti informativi contigui. Essa abbraccia un **arco temporale** che, a partire dalla legislatura corrente, si estende indietro nel tempo fino al 1848, in modo da rispondere non solo agli obiettivi di trasparenza e controllo democratico dell'operato delle istituzioni, ma anche a quelli di diffusione della conoscenza storica, attraverso la pubblicazione in rete dell'immenso patrimonio informativo racchiuso negli archivi parlamentari.

Inoltre, la posizione di spicco non è determinata solo dalla **quantità di dati** che oggi sono prodotti, ma anche dall'impiego di sofisticate tecnologie, quali le **ontologie formali** (per descrivere il dominio informativo), e dall'adozione di principi e modelli Linked Open Data, verso cui è stata orientata la scelta della

Quantità dei dati e qualità, grazie alle ontologie formali.

soluzione tecnologica. Difatti, con la scelta di utilizzare gli standard aperti del W3C (RDF e SPARQL) e con la contestualizzazione dei dati attraverso link verso altri dati in rete, le istituzioni parlamentari italiane hanno recepito appieno le indicazioni proposte da Berners-Lee e si sono candidate per le cinque stelle corrispondenti al massimo livello (p. 27).

I portali dati.camera.it e dati.senato.it.

Queste tecnologie sono alla base dei **due portali della Camera e del Senato**, dati.camera.it e dati.senato.it, operativi rispettivamente dal 2011 e dal 2013. Si tratta di **piattaforme di pubblicazione e condivisione di**

Linked Open Data che, sebbene con alcune differenze d'impostazione, mettono entrambe a disposizione, tra le altre cose:

- accesso agli end-point SPARQL per la ricerca sui dati;
- esempi di query SPARQL predefinite;
- sezioni dedicate al salvataggio dei dataset in vari formati;
- materiale di supporto alla comprensione delle ontologie [OCD](#) e [OSR](#).

Inoltre, dati.camera.it ha scelto di utilizzare [Lodview](#) come dereferenziatore delle IRI e [Lodlive](#) per navigare l'ontologia ed per esplorare i dati in modo libero; infine, fornisce un contatore dinamico alle triple che a mano a mano vengono pubblicate sull'end-point, presenta una sezione dedicata alle applicazioni realizzate sui Linked Open Data della Camera e uno spazio dedicato all'Hackathon Montecitorio, svolto nel 2014 per promuovere il riutilizzo dei dati da parte della comunità degli sviluppatori. [Dati.senato.it](http://dati.senato.it) offre invece un'interfaccia per definire filtri da applicare ai fini del salvataggio selettivo dei dataset.

Gli Open Data parlamentari sono disponibili con **licenza aperta CC-BY 3.0**, che consente di condividere (riprodurre e distribuire i dati con qualsiasi mezzo e formato) e adattare (modificare, trasformare il materiale e basarsi su di esso per trarne altre opere) i dati per qualsiasi fine, anche commerciale, con l'unica condizione di attribuire la paternità alla fonte e indicare se sono state effettuate modifiche. Costituiscono l'unica eccezione a questa licenza alcuni dataset, quali quelli bibliografici pubblicati dalla Camera dei deputati, che adottano la licenza CC-BY-SA 3.0, richiedendo di [condividerli allo stesso modo](#).

La situazione che si registra oggi non è certamente statica, bensì destinata ad evolvere rapidamente, e si prevede che l'offerta di Open Data parlamentari proceda nelle direzioni di un accrescimento e di una diversificazione. Il recupero dei dati e dei documenti storici è infatti un processo ancora in atto ed è quindi destinato a riversare in rete ulteriori masse di triple RDF che descrivano il parlamento italiano e la sua attività nel passato; inoltre, le scelte che guidano l'innovazione tecnologica delle istituzioni parlamentari italiane vanno sempre più nella direzione dell'adozione del paradigma Linked Open Data, visto non come un processo aggiuntivo rispetto alle tradizionali procedure di gestione dei dati, ma come il principio a cui ispirare i **processi di innovazione interna**. Tutto ciò amplifica il potenziale di esposizione di Open Data parlamentari, in quanto è sufficiente far transitare i dati da un end-point SPARQL interno a un end-point SPARQL esterno, per metterli a disposizione della collettività in Internet, senza bisogno di realizzare processi che trasformino i dati delle sorgenti interne in Open Data. Un esempio di questa innovativa impostazione tecnologica è dato dal [sistema PAD](#) (Piattaforma Aperta per la Documentazione), utilizzato dai Servizi di documentazione della Camera dei deputati e da diversi altri uffici.

2.3 La roadmap verso i Linked Open Data parlamentari

Per comprendere la costruzione, l'evoluzione nel tempo e le prospettive degli Open Data parlamentari è utile ripercorrere i passi dall'origine fino ad oggi, rappresentati nella *roadmap* illustrata nell'immagine qui di seguito riportata.

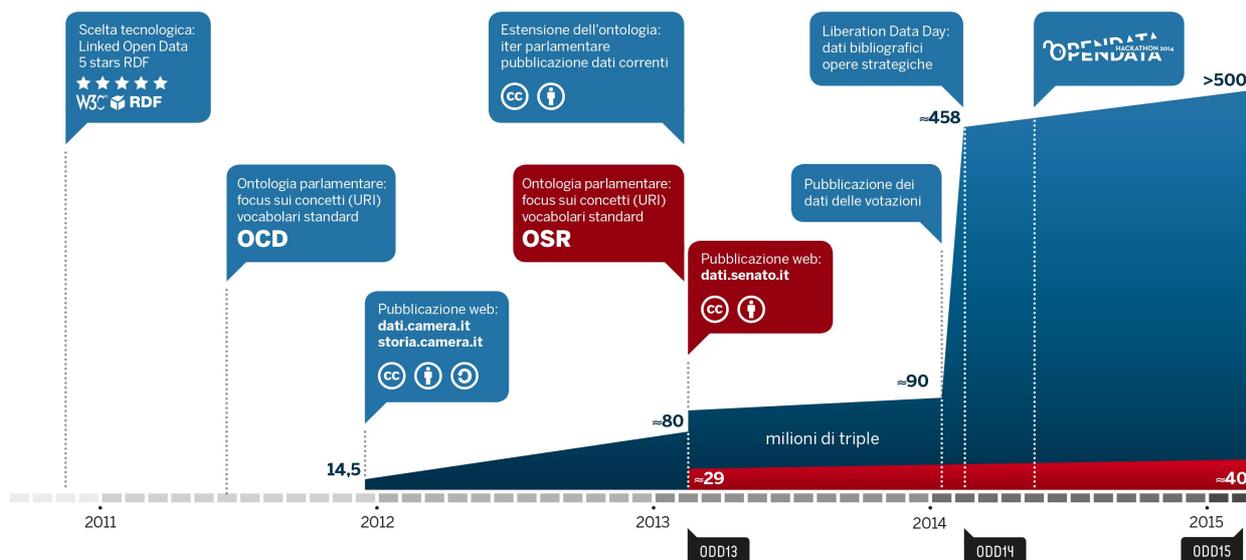


Immagine 1: Roadmap verso i Linked open Data Parlamentari.

2.3.1 La scelta tecnologica, l'ontologia OCD e dati.camera.it

I primi passi verso la produzione di Open Data parlamentari in Italia risalgono al **2010**, quando - in previsione delle celebrazioni dei 150 anni dell'Unità d'Italia - **la Camera dei deputati ha rivisto complessivamente la gestione delle risorse informative storiche** e ha avviato un intervento globale per unificare entro una struttura comune dati, informazioni, documenti digitali provenienti da applicazioni e sistemi differenti, struttura costituita da un comune repository di Linked Open Data (quindi con il massimo livello di apertura dei dati), senza ricorrere a mediazioni tecnologiche, conformemente alle *best practices* introdotte dal W3C.

A tal fine si è proceduto innanzitutto alla **formalizzazione di un'ontologia generale** in grado di descrivere il dominio informativo parlamentare, i diversi concetti e oggetti che lo compongono, le relazioni tra le risorse informative, in modo da rappresentare l'articolazione e il funzionamento dell'istituzione anche con le sue particolarità e le sue evoluzioni nel tempo: dalle schede anagrafiche dei parlamentari fino alle composizioni di commissioni e gruppi parlamentari, dalla gestione degli slittamenti nei mandati parlamentari alle evoluzioni degli organi (cambi denominazioni, accorpamenti, scissioni, etc.), dai progetti di legge alle interrogazioni parlamentari, dalle legislature alle elezioni, descrivendo via via oltre 150 anni di storia parlamentare.

La descrizione di questo insieme di risorse è rappresentata in un unico data model, identificato da uno specifico namespace, **OCD (Ontologia Camera dei Deputati)** – una ontologia [OWL](http://www.w3.org/2002/rdf-syntax-ns#OWL) (Ontology Web Language), [OCD, l'Ontologia Camera dei Deputati](http://dati.camera.it).

il linguaggio per pubblicare e condividere ontologie promosso dal W3C, espresso in triple RDF. L'ontologia OCD (su cui vedi p. 47), fortemente connotata in molti dei suoi elementi dal concetto di legislatura, che scandisce i tempi dell'attività parlamentare e contrassegna le strutture parlamentari, è stata rilasciata come vocabolario per la descrizione dei dataset della Camera dei deputati alla fine del 2011, in concomitanza con la pubblicazione del sito dati.camera.it (interamente dedicato agli Open Data) e del Portale Storico della Camera dei deputati (storia.camera.it), basato sui Linked Open Data delle 48 (oggi 49) legislature precedenti a quella corrente. In questa prima versione gli Open Data parlamentari sono stati rilasciati con la licenza CC-BY-SA 3.0, richiedendo quindi di condividerli allo stesso modo, e lanciando sul sito dati.camera.it un invito al riutilizzo e alla realizzazione di applicazioni basate su di essi.

L'ontologia OCD è nata come uno strumento flessibile, pensato per descrivere le sfumature

presenti in organizzazioni complesse come quella parlamentare, ma soprattutto è stata impostata come uno **strumento estensibile**, sia in senso verticale, per approfondire il dettaglio della rappresentazione di concetti anche solo accennati, sia in senso orizzontale, per ampliare la rappresentazione con concetti nuovi e nuove relazioni. Ed è proprio questa estensibilità che ha consentito le evoluzioni successive degli Open Data parlamentari e ha connotato le fasi successive della roadmap.

2.3.2 I Linked Open Data in ambito Intranet con PAD e il progetto Senato

Il **2012**, sebbene non abbia visto esposte sul Web sostanziali novità in tema di Open Data parlamentari, è stato in realtà segnato da un intenso lavoro interno sul tema, tanto da parte della Camera quanto da parte del Senato.

PAD: una piattaforma aperta e "linked" per la documentazione interna.

La **Camera dei deputati** si è infatti concentrata nella sperimentazione dell'adozione del **paradigma Linked Open Data in ambito Intranet**, raccogliendone indubbi vantaggi nell'incremento dell'efficienza interna, nella riduzione di costi e nel miglioramento dell'integrazione nella condivisione

delle informazioni entro strutture complesse, quali quelle del settore della documentazione. Ne è nato in questo ambito e in questa fase il **sistema PAD** (su cui vedi p. 87). Contemporaneamente ha proseguito sulla medesima direttrice analizzando l'estensione verticale dei concetti dell'ontologia attinenti all'iter legislativo (in particolare applicati alle attività correnti) al fine di rendere la pubblicazione dei dati parlamentari più tempestiva.

OSR, l'Ontologia Senato della Repubblica.

In parallelo, il **Senato della Repubblica** ha avviato l'analisi dei concetti e degli oggetti che compongono il proprio dominio informativo, progettando quella che poi sarebbe stata annunciata e pubblicata nel successivo 2013:

l'Ontologia Senato della Repubblica (OSR), una ontologia OWL affiancata e coordinata con l'ontologia OCD della Camera dei deputati. Lo studio dell'ontologia si inquadra nell'avvio di un ampio progetto, che aveva come obiettivo quello di pubblicare, secondo i suddetti formati aperti e le migliori pratiche internazionali, sia i dati, correnti e storici, già disponibili sul proprio sito Web, sia i testi inerenti l'attività legislativa.

2.3.3 L'estensione della ontologia OCD, l'ontologia OSR e dati.senato.it

Il lavoro interno svolto da entrambi i rami del parlamento ha prodotto nel 2013 diverse rilevanti novità nel campo degli Open Data parlamentari, che sono state annunciate in occasione dell'evento hub ODD13 (Open Data Day 2013), tenuto a Roma nel febbraio, e che sono state rese effettivamente disponibili in rete con la nuova legislatura – la XVII – a partire da marzo:

- la **nuova versione dell'ontologia OCD**, arricchita dei concetti descrittivi del dettaglio dell'iter legislativo;
- la pubblicazione in rete dei **LOD della Camera dei deputati** riferiti non più solo alle legislature storiche ma anche a quella corrente e **aggiornati quotidianamente**;
- il nuovo sito dati.senato.it, con i **Linked Open Data del Senato riferiti alle legislature dalla XIII fino a quella attuale**, secondo la formalizzazione della ontologia OSR (Ontologia del Senato della Repubblica) e **aggiornati quotidianamente**;
- la disponibilità, sul sito www.senato.it, dei testi dei disegni di legge presentati e discussi in Senato a partire dalla XVII legislatura, nel formato standard internazionale Akoma Ntoso.

2.3.4 Gli Open Data su votazioni e dati bibliografici/documentali

La produzione di **Open Data** parlamentari è proseguita nel 2014 – con la contestuale estensione dell'ontologia – con la pubblicazione dei dati relativi alle **votazioni in Assemblea di Montecitorio** riferite a tre legislature oltre quella corrente; ciò ha determinato una verticale

crescita del quantitativo di triple disponibili al riuso, che sono passate da meno di 100 milioni a oltre 400 milioni. Analogamente, il Senato della Repubblica ha messo a disposizione sul proprio portale Open data, i dati riferiti alle **votazioni svolte in Assemblea con sistema elettronico a partire dalla XIII legislatura**. I dati sulle votazioni in Aula pubblicati dalla Camera dei deputati e dal Senato della Repubblica costituiscono una massa informativa ingente, intorno a cui si concentra l'attenzione pubblica, e che organizzazioni che si occupano di monitoraggio dell'attività parlamentare – quali ad esempio Open Polis – avevano precedentemente acquisito attraverso operazioni manuali di estrazione dei dati dalle pagine dei siti Web. Questo metodo si è rivelato molto costoso e poco efficiente: una semplice modifica nella struttura delle pagine Web richiedeva infatti la rielaborazione degli strumenti di estrazione delle informazioni. La disponibilità di questi dati in formato aperto ha rappresentato quindi un passo importante nella direzione della trasparenza, aprendo la strada a nuovi modi, tecnologicamente più avanzati, di esercitare il controllo democratico sull'operato dei parlamentari.

I dati "parlamentari" per eccellenza: quelli sulle votazioni in Aula.

Nello stesso anno altre significative novità hanno interessato la produzione degli Open Data parlamentari: la **messa a disposizione di dati aperti in ambito bibliografico e documentale**. Si tratta di un esempio tipico di estensione dei dati aperti prodotti dal parlamento ad ambiti informativi continui al dominio informativo "parlamentare" in senso stretto (come definito a p. 41). Un precedente analogo in realtà si era già verificato quando, nella costruzione del repository Linked Open Data su cui si è basata la pubblicazione del [Portale Storico della Camera dei deputati](#), i dati relativi alla struttura amministrativa della Camera – con articolazione in Servizi e Uffici e relativi dirigenti che si sono succeduti nel tempo – e i dati relativi alla cronologia dei principali avvenimenti storici in ambito politico-istituzionale, socio-culturale e internazionale sono stati modellati e rappresentati come Linked Open Data, sebbene non direttamente esposti sul sito [dati.camera.it](#). Si trattava quindi di dati non propriamente "parlamentari" in senso stretto.

Dati non strettamente "parlamentari": l'esempio dei dati bibliografici.

La portata innovativa dell'adozione del paradigma Linked Open Data a informazioni di tipo bibliografico è stata oggetto di attenzione da parte della comunità dei bibliotecari. Ciò ha condotto, in particolare, all'adozione di un **modello RDF basato sull'adozione di ontologie standard di riferimento** per la rappresentazione dei dati, quali [BIBO](#) (Bibliographic Ontology), ISBD (International Standard Bibliographic Description) Ontology, DC (Dublin Core), e lo sperimentale utilizzo di interlinking con il catalogo internazionale delle voci controllate di autori [VIAF](#) (Virtual International Authority File) disponibile sul Web in Linked Open Data e collegato a sua volta con numerosi dataset bibliografici in tutto il mondo.

In ambito documentale, sebbene l'end-point di Linked Open Data interno sia ricco di dati RDF prodotti e utilizzati dai Servizi interni, sono stati resi disponibili dalla Camera dei deputati i dati – nel solo formato CSV – relativi a due temi di attualità dell'attività parlamentare:

- i dati di simulazione elettorale, derivanti dalla applicazione della nuova formula elettorale proposta dalla Commissione Affari costituzionali, approvata in Aula alla Camera il 12 marzo 2014, alle elezioni della Camera dei deputati del 2013, 2008 e 2006;
- i dati dell'attività di monitoraggio sull'attuazione delle infrastrutture strategiche comprese nella "legge obiettivo" (n.443/2001).

2.3.5 Il riutilizzo degli Open Data parlamentari: l'Hackathon Montecitorio 2014

Una volta raggiunto un meccanismo di produzione a regime di una massa informativa di Open Data parlamentari ingente come quella pubblicata dal parlamento italiano, l'attenzione si è spostata sul valore aggiunto che questo sforzo intrapreso dalle istituzioni parlamentari italiane negli ultimi anni poteva produrre per la collettività; ciò tenendo conto del fatto che uno dei principali obiettivi che il progetto Open Data si prefiggeva era quello di favorire il riutilizzo dei dati, la loro integrazione

con altre risorse informative e la promozione di applicazioni Web che potessero approfondire specifici aspetti a partire dagli Open Data parlamentari disponibili.

Sicuramente l'impegno profuso dalla Camera dei deputati e dal Senato nel campo degli Open Data ha creato valore in termini di trasparenza e ha contribuito a dare spicco alle istituzioni parlamentari italiane sia nel contesto parlamentare internazionale che nell'ambito del Web dei dati e della [Linked Open Data cloud](#).

*Code4Italy@Montecitorio
2014.*

Tuttavia, la mancata risposta all'appello – presente sia su [dati.camera.it](#) che su [dati.senato.it](#) – a contribuire con lo sviluppo di applicazioni ha mosso le istituzioni parlamentari a ricercare **nuovi mezzi per promuovere**

l'utilizzo dei Open Data parlamentari disponibili in rete. E' così che ha avuto origine l'iniziativa – assolutamente priva di precedenti nel panorama parlamentare – di un [Hackathon sugli Open Data parlamentari](#), organizzato presso la Camera dei deputati e svolto nel maggio del 2014 con la partecipazione del Senato della Repubblica. L'evento, aperto alla libera partecipazione previa registrazione on-line, si è articolato in una iniziale **sessione di formazione**, dedicata a chiunque fosse interessato a conoscere e lavorare sui dataset offerti dalla Camera dei deputati e dal Senato, con uno spazio per la **presentazione di pitch (proposte)** per la realizzazione di servizi ideati per la società civile al fine di comprendere meglio il mondo parlamentare, e in una successiva **sessione operativa** di 30 ore consecutive, svolta in forma competitiva e collaborativa, in cui i partecipanti hanno sviluppato **applicazioni basate sugli Open Data parlamentari**. È seguita una fase di valutazione delle numerose applicazioni realizzate, da parte di una Commissione giudicatrice appositamente costituita dal Comitato per la Comunicazione della Camera dei deputati, che ha individuato i primi tre progetti classificati, premiati nel corso di un evento nel successivo mese di luglio.

L'iniziativa è stata una esperienza particolarmente utile sia per le istituzioni che per i partecipanti, in quanto è stata un'occasione di riscontro e di diffusione sull'attività di produzione degli Open Data parlamentari, che ha dimostrato come la “apertura” dei dati parlamentari non sia soltanto una questione di licenze d'uso e di formati di dati, ma anche una questione di disponibilità all'incontro, al confronto e alla **condivisione tra istituzioni parlamentari e collettività**.

2.4 La Camera dei deputati

2.4.1 Introduzione

L'apertura e la trasparenza sono principi profondamente radicati nella cultura e nella storia della Camera dei deputati. Quando, l'8 maggio del 1848, due mesi dopo la promulgazione dello Statuto albertino, la Camera dei deputati si riunì per la prima volta a Torino e si diede immediatamente un [regolamento interno](#), fissò subito, in poche norme, le funzioni primarie della struttura amministrativa di supporto: la predisposizione del processo verbale e dei resoconti, la stampa degli atti parlamentari, la gestione della biblioteca e degli archivi di documentazione. Queste essenziali **finalità di apertura, trasparenza e comunicazione dell'attività parlamentare** sono state nel corso dei decenni declinate in vario modo, seguendo l'evoluzione delle tecnologie, fino all'odierno [dati.camera.it](#).

I Linked Open Data oggi sono una realtà tecnologica ormai nota e la Camera dei deputati nella adozione di questi standard e tecnologie è generalmente riconosciuta come una istituzione che ha percorso i tempi non solo in ambito italiano ma anche a livello internazionale.

L'appello lanciato da Tim Berners-Lee dalla tribuna di TED nel 2009 (http://www.ted.com/talks/tim_berniers_lee_on_the_next_Web) - “Raw Data Now!” – è stato raccolto sollecitamente dalla Camera dei deputati, che ha deciso di adottare le tecnologie Linked Data, avviando già l'anno successivo i lavori per progettare l'ontologia dei dati parlamentari – da cui è nata OCD (Ontologia Camera Deputati) – e realizzando il portale dedicato ai propri Open Data

(dati.camera.it) contestualmente al Portale storico della Camera (storia.camera.it), una vera e propria "Web app" unitaria, alimentata dal repository RDF attraverso il proprio Endpoint SPARQL, pubblicato alla fine del 2011, in occasione delle celebrazioni per i 150 anni dell'Unità d'Italia.

2.4.2 L'ontologia OCD Ontologia Camera deputati

Per la definizione dell'ontologia del dominio Camera è stata necessaria [L'ontologia OCD](#).
nonché proficua una stretta **collaborazione tra gli esperti di dominio parlamentare e i tecnici esperti di tecnologie LOD**. Ciò ha consentito di arrivare alla definizione di un modello astratto e allo stesso tempo efficace a rappresentare tutte le situazioni diverse che, come è possibile immaginare, si sono stratificate nel tempo. Se da una parte l'ontologia doveva essere in grado di rappresentare gli eventi "storici" delle prime legislature del Regno, essa doveva anche sostenere tutte le peculiarità degli eventi recenti, con tutta la granularità utile per produrre dataset consistenti, comprensibili da uomini e macchine, e per favorirne il riuso esterno.

Possiamo constatare oggi, a quasi quattro anni dalla nascita di OCD, che le tecnologie LOD hanno rappresentato un sistema molto efficace per la rappresentazione del modello di lavoro della Camera che, grazie alla naturale "elasticità" insita in RDF e OWL, è sempre molto **attualizzabile e rimodulabile**. Ad esempio, la prima versione dell'ontologia OCD non prevedeva la descrizione analitica dell'iter legislativo, mentre nella seconda versione pubblicata nel 2013 è stata aggiunta tutta la parte che modella l'iter parlamentare (dal punto di vista della Camera dei deputati) con tutti i riferimenti utili per analisi quantitative e qualitative dei processi.

Nella definizione delle classi e delle proprietà dell'ontologia si è cercato di riprodurre le caratteristiche tipiche di alcune situazioni e sono stati utilizzati **nomi il più possibile parlanti**, affinché l'utente finale potesse essere in grado di comprenderne la maggior parte e un utente esperto del dominio potesse esser in grado di comprendere i dati proposti nella loro interezza senza documentazione aggiuntiva. Per favorire la leggibilità, il riuso dei dati e l'apertura verso altre ontologie, si è provveduto ad assegnare a ciascuna classe e proprietà una etichetta, una descrizione e, laddove possibile, una definizione ufficiale tratta dal sito della Camera. In tutti i dataset le informazioni sono suddivise in una parte di **metadati descrittivi delle risorse** (rdfs:label, dc:title, dc:description, dc:date ad esempio) e in un'altra di **riferimenti a risorse interne al dominio o a risorse Web**. La classe "chiave" dell'ontologia è ocd:legislatura, che è legata a quasi tutte le risorse rappresentate.

La seguente figura illustra i principali concetti della ontologia OCD.

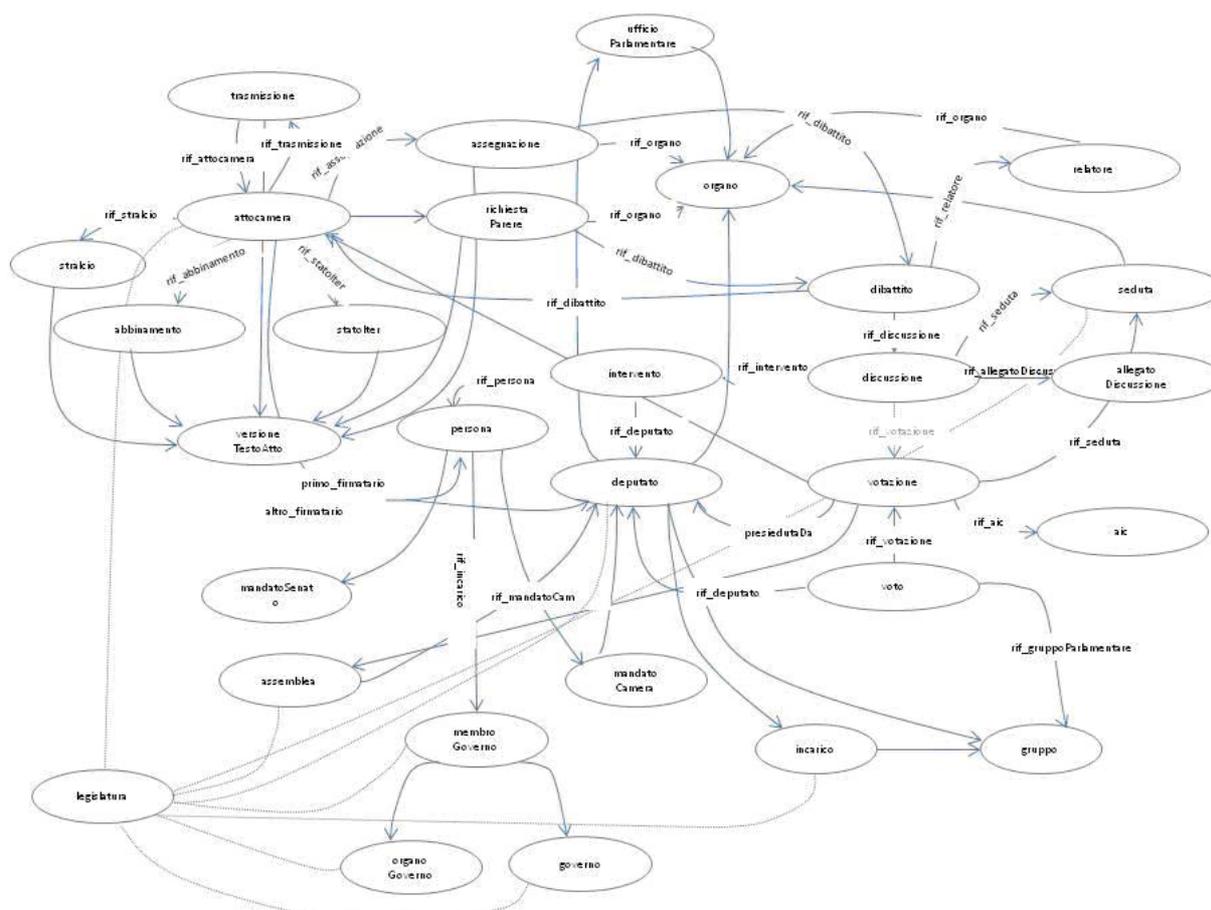


Immagine 2: Ontologia Camera Deputati.

Le **ontologie importate** nel modello OCD sono:

- [Dublin Core](#) e [Dublincore Terms](#) per la descrizione dei metadati più comuni come il titolo, la descrizione, le date, i riferimenti bibliografici;
- [Organization Ontology](#) per la descrizione generale dell'Istituzione Camera, recentissima raccomandazione del W3C;
- [Bio Ontology](#) per la descrizione degli eventi biografici dei deputati;
- [SKOS](#) per gli schemi di classificazione;
- [FOAF](#) per i metadati descrittivi delle persone;
- [BIBO Ontology](#) per la descrizione dei metadati bibliografici;
- [ISBD ontology](#) per la descrizione dei metadati bibliografici.

Un ulteriore lavoro preliminare alla predisposizione dell'ambiente è stata l'individuazione di griglie di classificazione e tesauri utilizzati all'interno del dominio (ad esempio titolari di classificazioni, soggetti, indici, rubriche) e la loro conversione in formato SKOS.

Thesaurus

Il **thesaurus** (o tesoro) è un elenco strutturato in cui le parole sono raggruppate per somiglianza semantica (contenente [sinonimi](#) e qualche volta [antonimi](#)); si differenzia dal [dizionario](#) che contiene definizioni e pronuncia. Secondo la definizione [ISO](#) il thesaurus è "un vocabolario di un linguaggio di indicizzazione controllato in maniera formalizzata in modo che le relazioni a priori tra i concetti sono rese

esplicite” (ISO 2788-1986). Le relazioni tesaurali sono utilizzate per facilitare la navigazione nel thesauro agli utenti. Il thesauro inoltre si differenzia dall'ontologia, in quanto quest'ultima può contenere relazioni più complesse e una logica inferenziale inerente al [modello](#).

In [informatica giuridica](#), il thesaurus può essere considerato una sorta di vocabolario dei termini presenti in una banca dati. Il suo scopo è quello di facilitare l'utente nella ricerca dei termini. Spesso chi compie una ricerca può andare incontro a [polisemie](#), sinonimie ecc. Il thesaurus, partendo da una categoria generale, guida il ricercatore fino all'uso della parola esatta, contenuta nella banca dati, che esprime esattamente il concetto che si voleva cercare. Le relazioni all'interno dei termini del thesaurus sono di gerarchia, preferenza o affinità.

SKOS (Simple Knowledge Organization System) è un sotto-linguaggio di OWL e propone una serie di classi e proprietà di RDF e RDFS per rappresentare il contenuto e la struttura di base dei cosiddetti Concept-Scheme, cioè tesauri, schemi di classificazione, tassonomie, terminologie, glossari e altri tipi di vocabolari.

La Skossification delle griglie di classificazione ha consentito di collegare l'ambiente LOD in modo trasversale così da favorire la ricerca delle informazioni all'interno dell'intera struttura e favorendo l'interoperabilità con altri sistemi di classificazione (ad esempio EUROVOC).

2.4.3 Il Portale Open Data [dati.camera.it](#)

Il Portale [dati.camera.it](#), pubblicato nel 2011, è dedicato agli Open Data della Camera dei deputati. Risponde all'obiettivo di dare libero e completo **accesso ai dataset Open Data della Camera dei deputati** che mano

Modalità di pubblicazione dei dati.

a mano vengono resi disponibili, attraverso processi di produzione continui, oltre a fornire informazioni, link e documenti di supporto. Vi si ritrovano le informazioni riguardanti l'attività e la struttura parlamentare sin dalla I Legislatura del Parlamento sabauda fino a quella attuale, con una consistenza in triple presenti nell'end-point SPARQL la cui costante crescita è monitorabile da un **contatore** visibile in homepage che ne visualizza in tempo reale il conteggio.

Il sito, completamente conforme agli standard del W3C per il semantic Web, consente il **download** di alcuni dataset predefiniti, ma attraverso l'**end-point SPARQL** è possibile estrarre le informazioni e aggregarle secondo specifiche necessità.

Il sito, in continua evoluzione, oggi si articola nelle seguenti **sezioni**:

- Dati
- Apps
- Linked Data
- Risorse
- Progetto

Sempre disponibile, in homepage e nelle sezioni, è l'accesso diretto alle funzioni principali del Portale: lo scarico dei dataset (**download**) e l'accesso all'**end-point SPARQL** per l'interrogazione e l'utilizzo dei dati.

Del portale è inoltre disponibile una **versione in inglese**, in forma ridotta e semplificata.

2.4.3.1 Dati

Nella sezione Dati si elencano e descrivono i dataset disponibili per il download, fornendo informazioni relative alle entità rappresentate, con rappresentazioni grafiche dell'ontologia. L'insieme dei dati è periodicamente aggiornato ed esteso, come risulta sia dalle notizie fornite in homepage che dal conteggio in tempo reale delle triple presenti nell'end-point SPARQL.

Download dei dataset.

I dati attualmente disponibili sono qui di seguito elencati.

- **L'anagrafica delle persone:** nel dataset è presente la descrizione delle persone censite all'interno delle basi dati legislative della Camera, sia deputati e senatori, sia membri non parlamentari di governo. Il dataset Persona riporta le informazioni anagrafiche ed è collegato con i mandati aperti alla Camera o al Senato e con gli eventuali incarichi di governo ricoperti. La persona è stata descritta utilizzando lo standard FOAF.

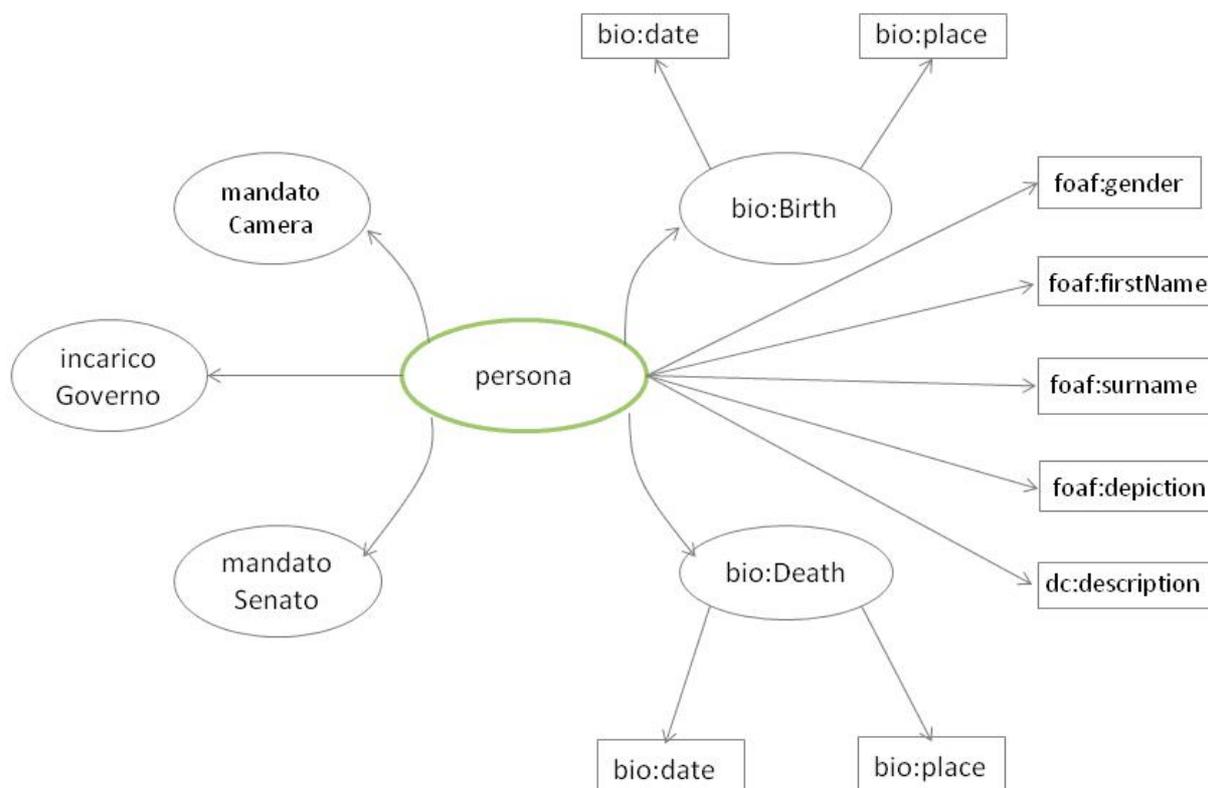


Immagine 3: Ontologia Anagrafica persone.

- **I deputati, i mandati ed i gruppi parlamentari:** il dataset comprende tutte le persone che hanno un mandato aperto alla Camera. Il deputato è strettamente connesso al concetto di mandato Camera e a quello di persona perché rappresenta una persona che riceve un mandato in una determinata legislatura; le informazioni anagrafiche del deputato si trovano nel dataset persona, mentre in questo dataset è possibile individuare tutti i ruoli svolti da una persona nel corso della sua attività parlamentare, l'appartenenza agli Organi della Camera e ai Gruppi, gli incarichi negli Organi e gli Uffici parlamentari ricoperti.

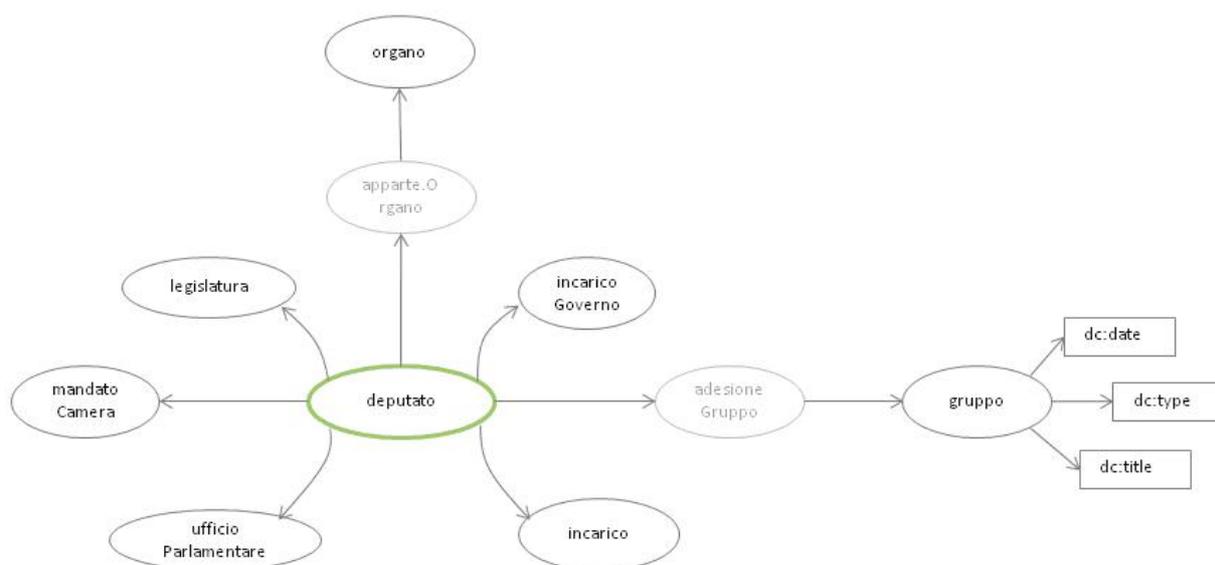


Immagine 4: Ontologia Deputati Mandati Gruppi.

- **Gli organi parlamentari e l'Assemblea:** il dataset Organo comprende la descrizione di Organi della Camera dei deputati con le informazioni relative a commissioni permanenti, commissioni bicamerali e d'inchiesta, giunte e altri organi interni.

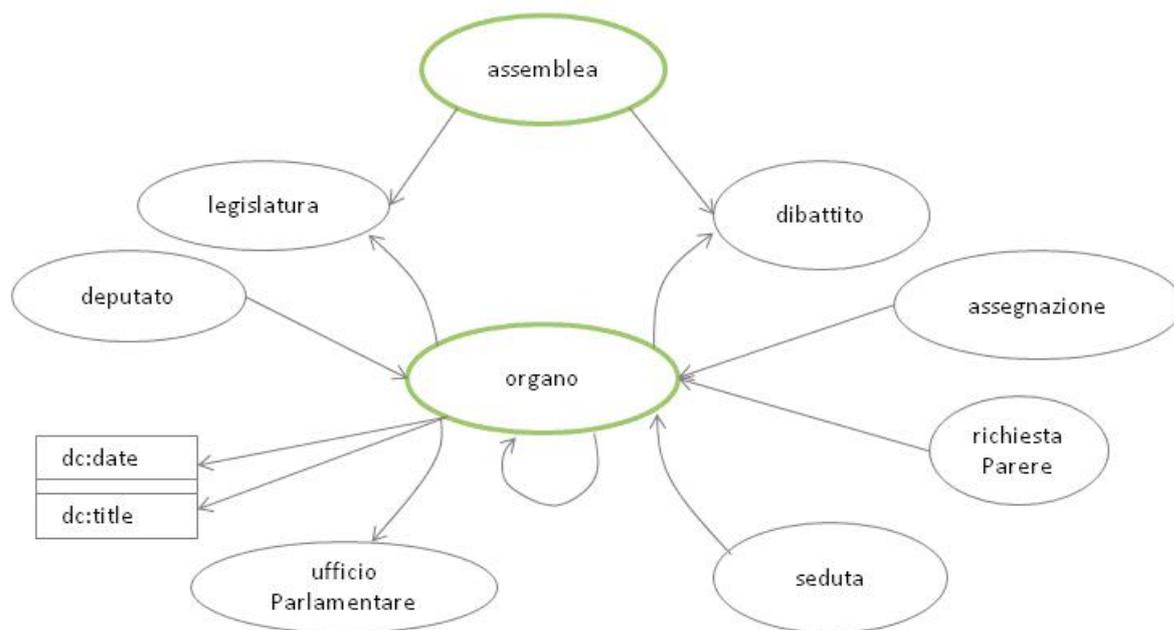


Immagine 5: Ontologia Organi Assemblea.

- **Gli incarichi e gli uffici parlamentari:** il dataset Incarico comprende tutti i dati relativi agli incarichi nei Gruppi parlamentari ricoperti dai deputati durante la loro attività; il dataset "ufficio Parlamentare" allo stesso modo conserva i dati riguardanti gli incarichi ricoperti dai deputati negli Organi parlamentari.

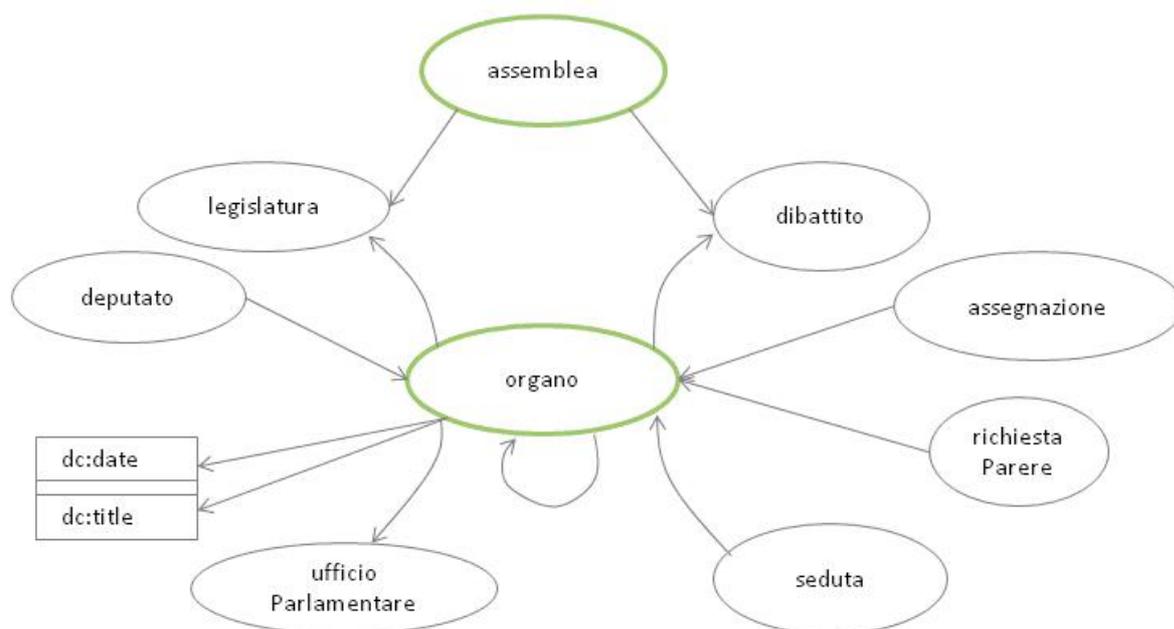


Immagine 6: Ontologia Organi Assemblée.

- **I Presidenti della Camera e i Presidenti della Repubblica:** il dataset Presidenti della Camera e Presidenti della Repubblica presenta la descrizione di tutti i Presidenti che si sono susseguiti nel corso degli anni, fino alla legislatura corrente.
- **I progetti di legge, l'iter legislativo e le leggi:** il dataset contiene la descrizione di tutto l'iter legis degli atti Camera della XVI legislatura fino alla loro conversione in Legge; per definire questo dataset l'ontologia OCD è stata integrata con tutte le classi e le proprietà necessarie ad una totale esplicitazione del dominio. Partendo dalla classe *atto* sono state aggiunte le seguenti classi:
 - la classe *assegnazione* serve a descrivere la fase di discussione in commissione in sede referente, legislativa e redigente. Dall'assegnazione parte il collegamento con i dibattiti e poi con le discussioni nelle sedute. Ogni discussione verrà descritta dettagliatamente fino al singolo intervento del deputato. Alla discussione è collegata la classe *votazione* che si articola in una votazione complessiva e una votazione analitica per singolo deputato;
 - la classe *richiesta Parere* descrive la fase di discussione in sede consultiva;
 - la *discussione* in Assemblée viene descritta utilizzando una proprietà diretta dall'atto Camera al *dibattito* (che nel caso specifico avrà un riferimento all'Assemblée);
 - le classi *abbinamento* e *stralcio* riportano le indicazioni della data dell'abbinamento o dello stralcio e il riferimento all'atto Camera abbinato o stralciato;
 - la classe *statolter* descrive lo stato dell'iter in un determinato momento importante riportando la data e lo stato;
 - la classe *trasmissione* serve a gestire le trasmissioni da e per il Senato. L'atto Camera avrà più trasmissioni (qualificate in alcuni casi da un *dc:type* navette se si tratta di un ritorno dal Senato con la navetta) con l'indicazione della data di trasmissione, la provenienza ed il collegamento con l'atto senato o con il nuovo atto Camera;
 - la classe *votazione*, collegata all'atto Camera e, in modo inverso, a tutti i singoli voti dei deputati è utilizzata per descrivere tutti i dati che rappresentano la

partecipazione al voto, l'esito e l'eventuale approvazione;

- la classe *voto* serve a descrivere il voto del singolo deputato.

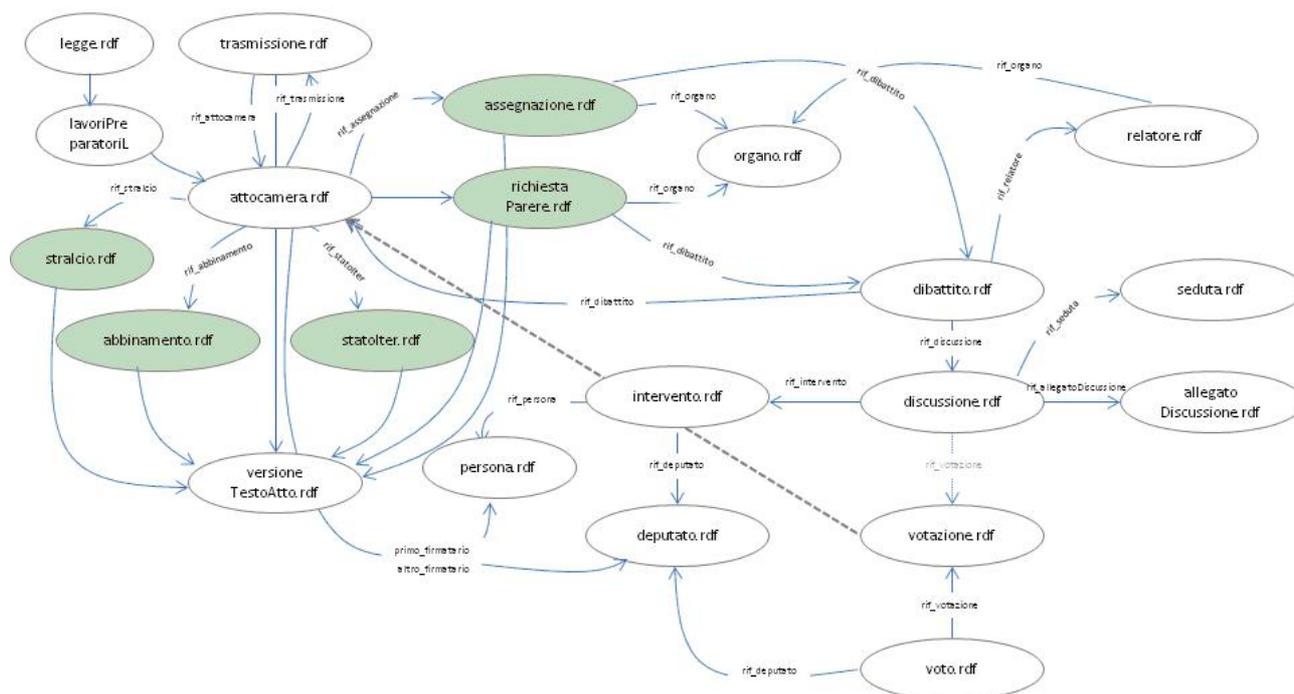


Immagine 7: Iter

- **Le votazioni ed i voti:** il dataset contiene la descrizione delle votazione elettroniche eseguite nelle sedute in Assemblea nel loro complesso e dei voti dei singoli deputati.
- **Le sedute, i dibattiti ed i bollettini delle giunte e delle commissioni:** nel dataset è presente la descrizione delle sedute, dei dibattiti e dei bollettini delle Giunte e delle Commissioni.

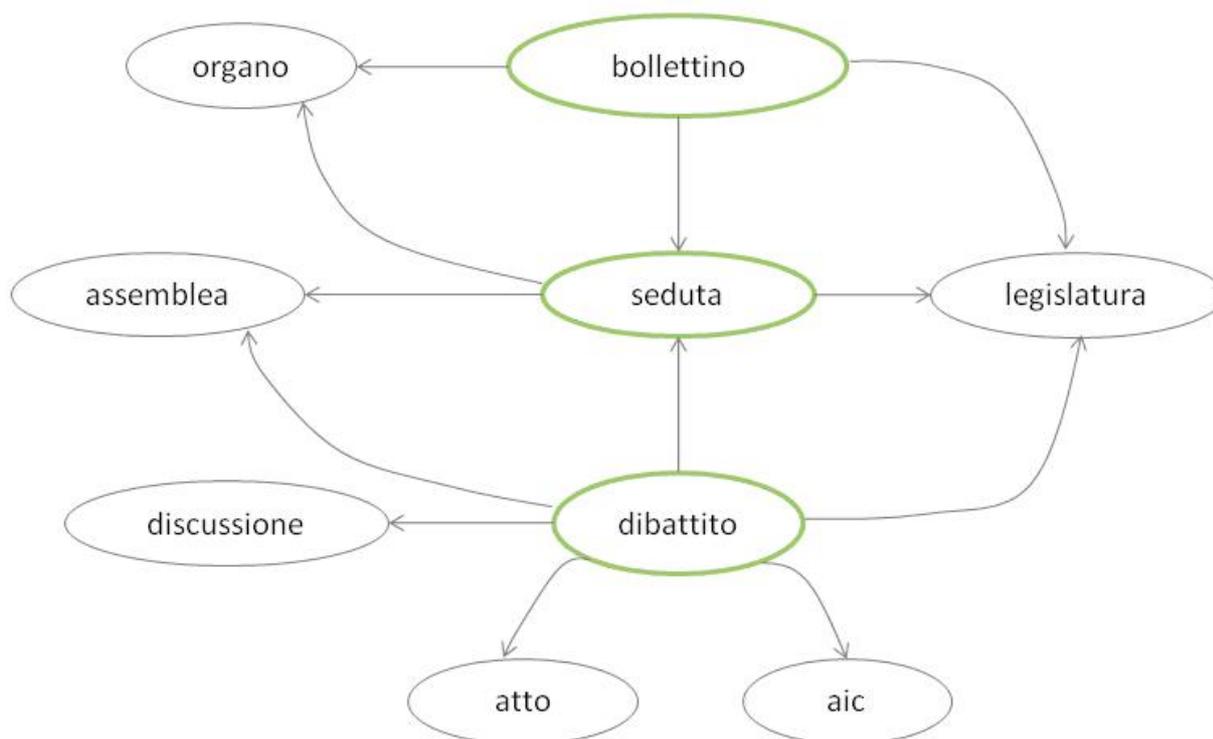


Immagine 8: Ontologia Sedute Dibattiti Bollettini

- **I documenti parlamentari:** nel dataset è presente la descrizione dei documenti parlamentari stampati dalla Camera dei deputati, i cosiddetti "DOC" in numero romano: documenti e relazioni, trasmessi alla Camera dal Governo o da altri soggetti istituzionali ovvero predisposti da organi della Camera stessa.
- **Interrogazioni, interpellanze e mozioni:** il dataset contiene gli Atti di indirizzo e controllo, quali interrogazioni, interpellanze, mozioni, ordini del giorno, risoluzioni, dalla VII alla XVI Legislatura della Repubblica.
- **Le legislature, i sistemi elettorali e le elezioni:** il dataset comprende la descrizione delle Legislature, dei Sistemi elettorali e delle Elezioni.
- **I Governi, gli organi governativi, gli incarichi ed i Presidenti del Consiglio:** il dataset comprende la descrizione dei Governi, dei Ministeri, degli incarichi ricoperti e dei Presidenti del Consiglio.
- **BPR - Bibliografia del Parlamento italiano e degli studi elettorali:** il dataset della BPR (Bibliografia del Parlamento italiano e degli studi elettorali), curata dalla Biblioteca della Camera dei deputati, censisce le pubblicazioni riguardanti il Parlamento italiano, la Consulta e l'Assemblea Costituente e le elezioni politiche a partire dal 1848. Una parte dei riferimenti bibliografici è arricchita dai testi integrali. Ad ogni citazione sono assegnati uno o più codici di classificazione sulla base di uno schema di oltre 100 classi organizzate in sette grandi aree. Per la modellazione RDF dei dati è stato deciso di riutilizzare ontologie per la descrizione bibliografica già diffuse a livello internazionale come la Bibliographic Ontology (**BIBO**), l'International Standard Bibliographic Description (**ISBD Ontology**) e il **Dublin Core Vocabulary**. Lo schema di classificazione è stato convertito in formato **SKOS**.
- **Dati dei Servizi di documentazione:** i dataset contengono dati provenienti o utilizzati dai Servizi di documentazione della Camera dei deputati. Sono gli unici dati in formato per

ora diverso da Linked Open Data, resi disponibili come **CSV**. Si tratta di dati attinenti a un tema della attività parlamentare – la riforma elettorale – e a una attività di monitoraggio sulla attuazione della cosiddetta “legge obiettivo”.

In particolare, ai fini di una simulazione della **riforma elettorale**, ai voti espressi alle elezioni della Camera dei deputati del 2013, 2008 e 2006 è stata applicata la formula di trasformazione dei voti in seggi, prevista nella proposta della Commissione Affari costituzionali, approvata in Aula il 12 marzo 2014. Metodologia e risultati in sintesi sono esposti nel Dossier del Servizio studi n. 98 sul calcolo dei risultati elettorali. I dataset CSV resi disponibili costituiscono i voti a cui è stata applicata la formula (voti validi riaggregati nelle circoscrizioni elettorali e nei collegi plurinominali previsti nelle tabelle A e B della proposta di legge) e il dettaglio del risultato finale, vale a dire i seggi attribuiti alle liste nei collegi plurinominali.

L'altro insieme di dati rappresenta l'esito della attività di **monitoraggio sull'attuazione delle infrastrutture strategiche** comprese nella “legge obiettivo” (legge 443/2001), svolta dal Servizio Studi della Camera in collaborazione con l'Autorità per la vigilanza sui contratti pubblici di lavori, servizi e forniture (AVCP); tale esito, presentato alla Commissione ambiente nella seduta del 4 febbraio 2014, contiene dati aggiornati al 31 ottobre 2013. I dati sono stati recentemente integrati con gli esiti dell'ultimo monitoraggio, presentato alla Commissione ambiente nella seduta dell'11 marzo 2015, aggiornato al 31 dicembre 2014. Le opere pubbliche sono suddivise in lotti e in ulteriori interventi che rappresentano un segmento dei lavori da svolgere. In considerazione del fatto che il monitoraggio coinvolge in prevalenza grandi infrastrutture il livello di segmentazione è più articolato e dettagliato. Il monitoraggio prende in considerazione 1.420 lotti, tra opere, interventi, sottointerventi e ulteriori dettagli, relativi a 419 infrastrutture. Per ciascun intervento sono riportati la denominazione, il soggetto competente, la localizzazione regionale, lo stato di attuazione e la data presunta di ultimazione dei lavori. Tutti gli interventi sono classificati a seconda se siano o meno inclusi nell'ultimo documento programmatico trasmesso dal Governo ovvero se siano stati o meno oggetto di delibera del CIPE (Comitato interministeriale per la programmazione economica). La delibera del CIPE segna il passaggio nell'iter procedurale rispetto all'approvazione dei progetti e all'assegnazione delle risorse. I dati di carattere finanziario riguardano il costo delle opere riportato nei documenti programmatici trasmessi dal Governo, il costo presunto aggiornato alle date di monitoraggio (31 ottobre 2013 e 31 dicembre 2014), le risorse disponibili per la realizzazione dei lavori e quelle mancanti.

- Le **relazioni con altri dataset online** e gli **allineamenti semantici**: tassello essenziale per la realizzazione di una interoperabilità semantica dei dati è l'interlinking verso altri dataset già esposti nella linked data cloud. Questa interconnessione è, come già detto, ciò che permette l'accrescimento del patrimonio informativo di partenza grazie ai dati e alle informazioni provenienti dal Web e allo stesso tempo consente la disambiguazione delle entità e la riconduzione a concetti univoci.

Cercando quindi di perseguire l'obiettivo di pubblicazione “5 stars”, i dati pubblicati nel dominio Camera sono stati interconnessi con altri dataset LOD utilizzando procedure automatiche basate sulla sovrapposizione puntuale di nome e cognome, data e luogo di nascita – per i nomi di persona – e di luogo con i relativi dati geografici gerarchicamente superiori – per i nomi di luogo presenti nei documenti. L'interlinking, finora, è stato realizzato con i dataset di DBpedia.org, DBpedia Italia, Wikidata, Freebase per i nomi dei deputati e delle persone presenti nei dati; di VIAF (Virtual International Authority File) per gli autori pubblicati dalla Biblioteca; di Geonames e LinkedGeoData per i nomi di luogo. È proprio grazie a questi allineamenti semantici che è stato possibile collocare i dataset della Camera nella versione della Linked Data Cloud pubblicata ad agosto 2014.

2.4.3.2 Apps

La produzione di App come dimostratori delle potenzialità dei Linked Data parlamentari.

La produzione di Linked Open Data della Camera dei deputati non è stata concepita come iniziativa fine a se stessa, ma sin dall'inizio è stato dato particolare risalto all'obiettivo di favorire il riuso dei dati prodotti e la loro integrazione con altre raccolte informative, promuovendo la realizzazione

di applicazioni Web al fine di approfondire aspetti diversi e specifici della vita parlamentare presente e passata.

Le Apps sono concepite come un nuovo, semplice, intuitivo e creativo modo di navigare attraverso i dati conservati negli archivi della Camera. Dal Portale dati.camera.it si richiamano alcune di queste Apps. Prima di tutto ci si riferisce allo stesso Portale storico, che costituisce il principale esempio di riuso dei Linked Open Data pubblicati, arricchito di cinque percorsi di approfondimento – Legislature, Presidenti, Regolamenti, Montecitorio e Amministrazione – e un'ampia sezione multimediale, alimentata dall'archivio fotografico, dalla collezione di opere d'arte, dalle immagini dei palazzi della Camera, essa offre oltre alla cronologia degli eventi politico-istituzionali, nazionali e internazionali, anche molteplici chiavi di lettura dei dataset, che possono essere selezionati e consultati attraverso un'intuitiva e innovativa navigazione basata su filtri a “faccette”, tipica del Web semantico. Sono poi state realizzate specifiche pagine Web interattive che, utilizzando tecnologie innovative, consentono all'utente di riaggregare liberamente i dati in funzione dei propri scopi di approfondimento, sfruttando il patrimonio di relazioni semantiche tra essi e la più innovativa tecnologia Web 2.0: una Mappa delle elezioni del Regno d'Italia e un Atlante Istituzionale.

L'iniziativa dell'Hackathon stesso, intrapresa nel 2014, è stata essenzialmente volta a promuovere la realizzazione di Apps basate sui Linked Open Data parlamentari e i primi tre progetti premiati sono accessibili dalla pagina dedicata all'evento <http://hackathon.camera.it>.

2.4.3.3 Linked Data

Molteplici strumenti per accedere ai dati in tempo reale o analizzarli off-line.

I Linked Data contenuti all'interno del repository possono essere utilizzati direttamente all'interno di applicazioni esterne con un accesso online.

Per un utilizzo off-line dei dati, si rende possibile effettuare il download delle serializzazioni dei dati in formato RDF/XML o CSV.

Inoltre, è possibile interrogare direttamente l'[end-point SPARQL](#) del repository: ad esempio può essere utilizzato da applicazioni mash-up per creare collegamenti con risorse di altri dataset. Per agevolare l'espressione di query SPARQL, viene proposta una collezione di esempi semplici che mostrano alcune delle tipologie di query che è possibile effettuare sull'*end-point*. Questa collezione viene ricorrentemente integrata sulla base delle specifiche richieste che provengono dal pubblico interessato ai dati.

Infine, per una semplice e intuitiva visualizzazione e consultazione dei dati viene suggerito l'utilizzo di navigatori RDF pubblici e accessibili via Web, che offrono un quadro esaustivo delle risorse in esame esplicitando tutti i collegamenti con le altre informazioni (fotografie, documenti pdf, ecc.). A tal fine si richiama una specifica implementazione di LodLive, configurata per permettere l'esplorazione delle classi e delle proprietà attualmente utilizzate nei dati pubblicati nell'*end-point* SPARQL.

2.4.3.4 Risorse

Nel portale dati.camera.it una specifica sezione è dedicata ad autorevoli risorse informative dei principi dei Linked Open Data: tutorial e specifiche relative a SPARQL, tutorial, guide e specifiche relative ai Linked Data, guide e articoli sugli Open Data, guide e realizzazioni di Open Government Data. Si tratta di materiale che, senza alcuna pretesa di esaustività, costituisce una base informativa per chiunque sia interessato a tali temi, periodicamente aggiornata con nuovi riferimenti.

2.4.3.5 L'end-point SPARQL

Da qualsiasi pagina del portale dati.camera.it si può accedere all'end-point SPARQL per l'interrogazione e l'utilizzo dei dati, che è gestito dal software open source **Virtuoso**, gestore di triplestore con motore SPARQL.

La pagina di interrogazione dei dati su end-point richiede la formulazione di query nel linguaggio SPARQL. E' stata resa più *user friendly* attraverso l'inserimento di una serie di **query predefinite** e spiegate.

[End-point SPARQL e query predefinite di esempio.](#)

Una volta selezionata la query, questa viene visualizzata nel box e si può lanciare. I risultati si possono ottenere in diversi formati, tra cui il CSV, l'HTML o il JSON.

2.4.3.7 Download dei dataset

Dal portale dati.camera.it è possibile scaricare sia i dataset relativi alla legislatura corrente sia quelli delle legislature precedenti, a partire dal 1848. I dataset storici disponibili sono organizzati per tipologia e fanno riferimento a tutte le legislature passate. Quelli relativi alla legislatura corrente sono scaricabili separatamente.

[Download di dati storici e correnti e interrogazioni tramite SPARQL.](#)

In tutti i dataset le informazioni sono suddivise in una parte di metadati descrittivi delle risorse – *rdfs:label*, *dc:title*, *dc:description*, *dc:date* – e una parte di riferimenti ad altre risorse interne al dominio o a risorse Web.

Ogni dataset è stato costruito mirando alla massima leggibilità e chiarezza, l'utente finale dovrebbe essere in grado di comprenderne la maggior parte mentre un utente esperto del *dominio* dovrebbe essere in grado di comprendere i dati proposti nella loro interezza senza documentazione aggiuntiva. Particolare attenzione è stata posta nella denominazione delle *classi* e delle *proprietà* ed all'utilizzo di standard internazionali ove questi si intersecavano con precisione con le informazioni presenti nei *dataset*.

La Camera dei deputati ha prodotto e pubblicato i dataset dei *Senatori* con i dati delle relative *Persone*, dei *mandati al Senato* e dei *Presidenti della Repubblica* con un livello di dettaglio inferiore rispetto agli altri dataset al solo intento di garantire l'integrità delle relazioni.

L'end-point SPARQL è aggiornato ogni giorno e analogo livello di aggiornamento si offre per i dati disponibili al download.

2.4.4 Le licenze d'uso

Originariamente, quando nel 2011 si è pubblicato il repository Linked Open Data della Camera dei deputati, la licenza d'uso con cui i dati sono stati resi disponibili è stata la [CC-BY-SA 3.0](http://creativecommons.org/licenses/by-sa/3.0/).

[Creative Commons: da Condividi-allo-stesso-modo a semplice Attribuzione.](http://creativecommons.org/licenses/by-sa/3.0/)

A partire dal 2013, con l'avvio della legislatura XVII, si è passati alla licenza [CC-BY 3.0](http://creativecommons.org/licenses/by/3.0/) con l'obiettivo di promuovere un più diffuso riuso dei dati resi disponibili. Unica eccezione è rappresentata dai dati bibliografici relativi alla BPR che hanno mantenuto la licenza [CC-BY-SA 3.0](http://creativecommons.org/licenses/by-sa/3.0/).

2.5 Il Senato della Repubblica

2.5.1 Introduzione

Iniziative pregresse.

Il Senato italiano, sin dai tempi del Senato del Regno, ha mostrato grande attenzione riguardo l'archiviazione e la pubblicazione di atti, documenti, leggi e altri contenuti a carattere giuridico. A partire **dai primi anni '70** del XX secolo queste attività hanno ricevuto un notevole impulso, soprattutto grazie all'installazione di un elaboratore centrale e alla realizzazione di un sistema informatico all'interno del quale sono state via via archiviate le informazioni relative all'iter dei disegni di legge, alla composizione degli organi del Senato e all'attività non legislativa. **Negli anni '90**, l'avvento di Internet ha comportato la nascita del sito ufficiale www.senato.it (1996), attraverso il quale gli atti e i dati, precedentemente messi a disposizione di un ristretto pubblico di specialisti, sono stati aperti alla consultazione da parte di tutti i cittadini. Il Senato della Repubblica è stata una delle prime istituzioni italiane ad essere presenti sulla rete Internet e, sin dall'inizio, ha avuto come obiettivo quello di essere, piuttosto che un sito "vetrina", un vero e proprio sito di servizio per i cittadini, che consentisse di far conoscere quanto accadeva all'interno dei "palazzi" ad un pubblico più vasto possibile. Questo obiettivo ha determinato tutte le successive evoluzioni del sistema informativo che, nel 2000, è passato ad un'architettura più moderna, di tipo client/server, basata su database relazionale; le successive versioni del sito Internet, realizzate sulla nuova architettura, hanno reso l'accesso ai dati ancora più semplice, anche grazie all'introduzione di numerosi motori di ricerca (uno generale e una ventina specialistici) per le diverse banche dati. Altre iniziative, quali la pubblicazione di una versione del sito per dispositivi mobili (mobile.senato.it) e di un sito divulgativo destinato ad un pubblico di ragazzi (www.senatoperiragazzi.it), hanno continuato il cammino precedentemente intrapreso.

A metà degli anni 2000 anche la legislazione nazionale ha posto particolare accento sulla necessità, da parte delle istituzioni, di rendere pubblici i dati in proprio possesso nella modalità più semplice possibile. In particolare, il Codice dell'Amministrazione Digitale (CAD, Decreto Legislativo, 7 marzo 2005, n. 8 e s.m.i.) ha introdotto il principio di "disponibilità dei dati pubblici", cioè la possibilità, per soggetti pubblici e privati, "di accedere ai dati senza restrizioni non riconducibili a esplicite norme di legge"; il Codice sollecita le pubbliche amministrazioni a diffondere i dati in proprio possesso utilizzando formati aperti che ne consentano il riutilizzo. Il libero accesso all'informazione pubblica consente di conseguire due principali obiettivi:

- **Trasparenza:** consentire a tutti i cittadini di controllare l'operato delle amministrazioni pubbliche;
- **Apertura:** permettere a chiunque di riutilizzare i dati pubblici, sviluppando servizi e applicazioni di utilità generale.

La grande novità introdotta dal Codice consiste proprio nell'invito ad utilizzare formati aperti e a promuovere il riutilizzo dei dati da parte di terzi, anche per fini commerciali. Fino ad allora, tutte le istituzioni, Senato della Repubblica compreso, si erano adoperate per pubblicare i dati in formati adatti alla lettura da parte di utenti umani (pagine HTML, file in formati PDF, ecc.); chi avesse voluto riutilizzare i dati, ad esempio per condurre studi o ricerche, o per presentare quegli stessi dati in altre forme, a fini divulgativi o di analisi politica, poteva soltanto estrarre quei dati dal codice HTML delle pagine Web (cosiddetto Web scraping), con il problema di dover adattare la procedura di estrazione ad ogni modifica della struttura delle pagine Web.

Attività odierna.

Il Senato della Repubblica ha quindi avviato un **processo di "apertura" dei propri dati, a partire dalla fine del 2012**; tale processo ha portato alla realizzazione di un portale ad hoc (dati.senato.it) pubblicato in forma sperimentale all'inizio del 2013 e in forma definitiva nel 2014. Tale portale consente sia di **scaricare insieme predefiniti di dati** (dataset), sia di **eseguire interrogazioni sull'insieme di tutti i dati** disponibili, andando così incontro alle esigenze di differenti tipologie di utenti: dagli sviluppatori, che costruiscono

interrogazioni specifiche per realizzare delle applicazioni, ai giornalisti e ricercatori, che possono scaricare i dataset per finalità di analisi e approfondimento.

2.5.2 I dati pubblicati

Si è deciso di mettere a disposizione in formato aperto la maggior parte delle informazioni già disponibili attraverso il sito Web istituzionale del Senato. L'apertura di questi dati è stata semplificata dalla struttura del sistema informativo e dall'architettura di alimentazione delle pagine del sito Web istituzionale. Le pagine Web sono infatti costruite in modo automatico a partire dai dati presenti nel database, tramite delle procedure giornaliere di aggiornamento, basate su interrogazioni. È stato perciò possibile riutilizzare le medesime interrogazioni per ottenere un formato di output alternativo all'HTML; a tale scopo si è deciso di utilizzare il **formalismo RDF**, lo standard in ambito open data (e, in particolare, linked open data), definito dal W3C, anche al fine di semplificare il collegamento con altre sorgenti di dati (quali, ad esempio, i dati aperti della Camera dei deputati). Tutto ciò seguendo i **principi del Web semantico**, ovvero mirando ad affiancare al Web di documenti, tra loro collegati da link "muti", un Web dei dati che renda possibile la comprensione semantica e l'elaborazione automatica degli stessi.

Modalità di pubblicazione dei dati.

Per quanto riguarda l'arco temporale dei dati pubblicati, essi spaziano dalla XIII legislatura (1996) ad oggi, per complessivi 35 milioni di triple RDF attualmente disponibili.

I dati pubblicati sono relativi a:

- **Composizione:** la composizione del Senato, dei Gruppi parlamentari, delle Commissioni e del Consiglio di Presidenza, con tutte le variazioni intervenute nel corso delle legislature;
- **Atti:**
 - i disegni di legge presentati alla Camera e al Senato, gli iter legislativi, le iniziative, le assegnazioni alle commissioni, i relatori e gli interventi in Assemblea, la classificazione TESEO (vedi p. 71);
 - i Documenti non legislativi e la loro iniziativa;
 - gli atti di Sindacato Ispettivo e la loro iniziativa.
- **Votazioni:** le votazioni elettroniche svolte in Assemblea, sia su disegni di legge, sia su documenti non legislativi, con il dettaglio dei voti espressi da ciascun senatore.

Per consentire una più immediata comprensione dei dati pubblicati, questi sono affiancati da un'ontologia, che descrive in modo formale i concetti e le relazioni tra di essi. Tale ontologia, denominata **Ontologia del Senato della Repubblica (OSR)** è pubblicata nel formato standard RDF/XML, oltre che in Turtle, è costituita da 20 classi, ed estende l'ontologia della Camera dei deputati (OCD), oltre a riutilizzare numerosi concetti di altre ontologie standard, quali [FOAF](#) e [BIO](#) rispettivamente per quanto riguarda le proprietà delle persone e degli eventi di interesse biografico.

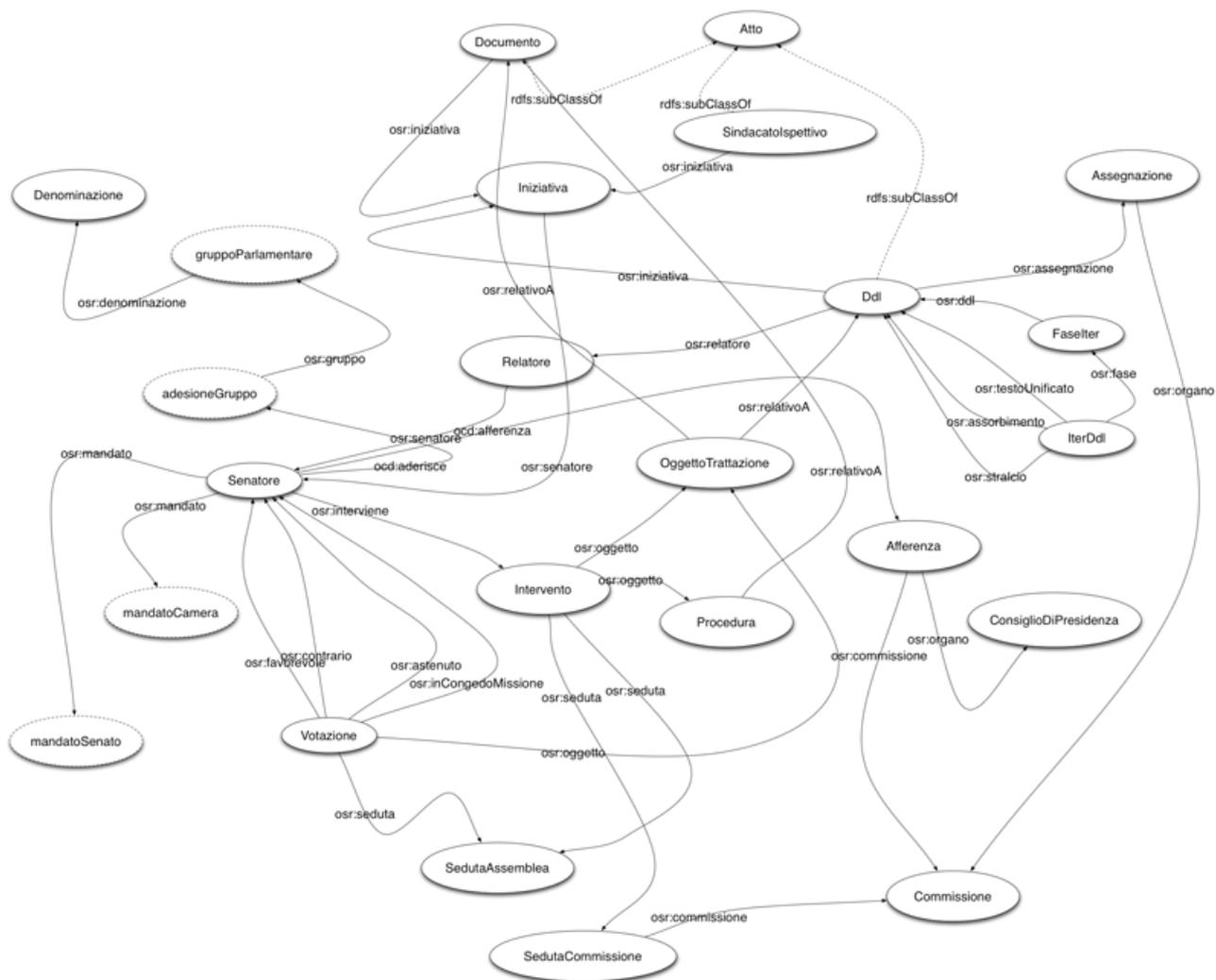


Immagine 9: Diagramma dell'ontologia del Senato della Repubblica.

2.5.3 La licenza d'uso

Date le esigenze di massima “apertura” dei dati, la licenza di pubblicazione dei dati è di tipo aperto, tale cioè da garantire all'utente una serie di diritti, in tema di utilizzo, ripubblicazione e modifica dei dati.

Le licenze più diffuse per i dati aperti della pubblica amministrazione sono quelle basate su Creative Commons (CC).

La licenza scelta:
Creative Commons
Attribuzione.

Da un'analisi dei dataset pubblicati sul portale dati.gov.it si osserva che le licenze maggiormente applicate dalle amministrazioni italiane sono la CC-BY e la [IODL 2.0](https://www.iodl.it/). Esse sono sostanzialmente equivalenti, ma le [linee guida per l'introduzione degli open data nelle pubbliche amministrazioni](#)

raccomandano l'utilizzo di licenze che siano leggibili e riconoscibili non solo a utenti nazionali ma anche a utenti internazionali "in un'ottica di interoperabilità transfrontaliera". Da ciò deriva, per i dati messi a disposizione dal Senato della Repubblica, la scelta della **licenza CC-BY 3.0 Unported**. Essa permette al soggetto utilizzatore di riprodurre, distribuire, comunicare, esporre, rappresentare, nonché di modificare e usare i dati, anche a fini commerciali, con il solo obbligo di attribuire la paternità dell'opera.

Similmente, anche la Commissione Europea ha di recente raccomandato, all'interno delle Linee guida già menzionate (p. 13) l'applicazione di licenze standard (quali le Creative Commons) ai dati aperti, in luogo di licenze proprietarie.



Immagine 10: Il portale dati.senato.it.

2.5.4 Il portale

Il portale dati.senato.it, pubblicato in versione sperimentale all'inizio del 2013, ed in versione definitiva nel 2014, presenta un menu di navigazione principale attraverso il quale gli utenti possono accedere alle **sezioni principali del sito**, che sono:

- Il progetto;
- Scarica i dati;
- Interroga i dati;
- Feed RSS.

2.5.4.1 Il progetto

In questa sezione sono riportate le **informazioni generali** relative al progetto di pubblicazione di dati aperti da parte del Senato della Repubblica. Sono poi descritti nel dettaglio:

- L'Ontologia del Senato della Repubblica (OSR), tramite un grafo riassuntivo e tramite l'elenco di tutte le classi e le proprietà che ne fanno parte, con descrizione di ciascuna di esse;
- La licenza d'uso, che, come descritto sopra, è la CC-BY, con breve spiegazione della licenza stessa e un link che rimanda al testo ufficiale;
- Le specifiche del risolutore di URN del Senato della Repubblica.

Sempre all'interno di questa sezione sono poi presenti le informazioni di contatto ed un modulo tramite il quale chiunque abbia sviluppato un'applicazione basata sui dati aperti del Senato può segnalare tale applicazione in modo che possa essere resa pubblica.

2.5.4.2 Scarica i dati

È la sezione del sito destinata ad **utenti non tecnici**, quindi principalmente ricercatori, giornalisti, ma anche cittadini che siano interessati ad ottenere in modo semplice i dati sull'attività parlamentare per effettuare analisi, statistiche, studi di vario tipo, e ad integrare tali dati con quelli provenienti da altre fonti informative. Per rendere semplice l'accesso ai dati, questo è realizzato mediante dei **moduli dinamici** nei quali gli utenti possono specificare dei parametri di ricerca quali la legislatura d'interesse e una data (o un intervallo di date), oltre al formato di output che può essere uno a scelta tra XML, JSON e CSV (Comma Separated Values, valori separati da virgola).

[Download dei dataset.](#)

I dataset disponibili sono suddivisi in tre sotto-sezioni: Composizione, Atti e Votazioni.

The screenshot shows the 'dati.senato.it' website interface. At the top, there is a navigation bar with links for 'Home', 'Il progetto', 'Scarica i dati', 'Interroga i dati', and 'Feed RSS'. The main content area is titled 'Composizione' and contains text explaining the data available, including a list of senators and their roles. Below this, there are three data download forms, each with a date selector and a 'download' button. A sidebar on the right provides a 'Searica i dati' menu with expandable options for 'Introduzione', 'Composizione', 'Atti', and 'Votazioni'. It also includes 'Riferimenti' (Il formato CSV, Il formato JSON, Il formato XML) and 'Link' (dati.camera.it, dati.gov.it) sections.

Immagine 11: Esempio di form per scaricare i dati dal portale open data del Senato della Repubblica.

2.5.4.3 Composizione

La sezione Composizione mette a disposizione i dataset relativi alla **composizione dell'Assemblea del Senato e dei suoi organi**, oltre ad informazioni relative all'attività di ciascun senatore, quali gli incarichi all'interno delle commissioni e dei gruppi parlamentari, gli interventi effettuati nel corso delle sedute di Assemblea e di commissione, e l'iniziativa sia di tipo legislativo che non legislativo.

2.5.4.4 Atti

La sezione Atti mette a disposizione i dataset relativi agli **atti legislativi e non legislativi**. In particolare, è possibile ottenere gli elenchi dei disegni di legge, degli atti di Sindacato Ispettivo e dei documenti non legislativi presentati in un certo intervallo di date, nonché, per ciascun disegno di legge, il dettaglio del suo iter (successione delle letture parlamentari) e l'elenco dei presentatori.

2.5.4.5 Votazioni

La sezione Votazioni mette a disposizione i dataset relativi alle **votazioni elettroniche svoltesi in Assemblea**. Sono pubblicati i dati relativi a tutte le votazioni elettroniche svolte nel corso della legislatura, con l'indicazione del voto espresso da ciascun senatore.

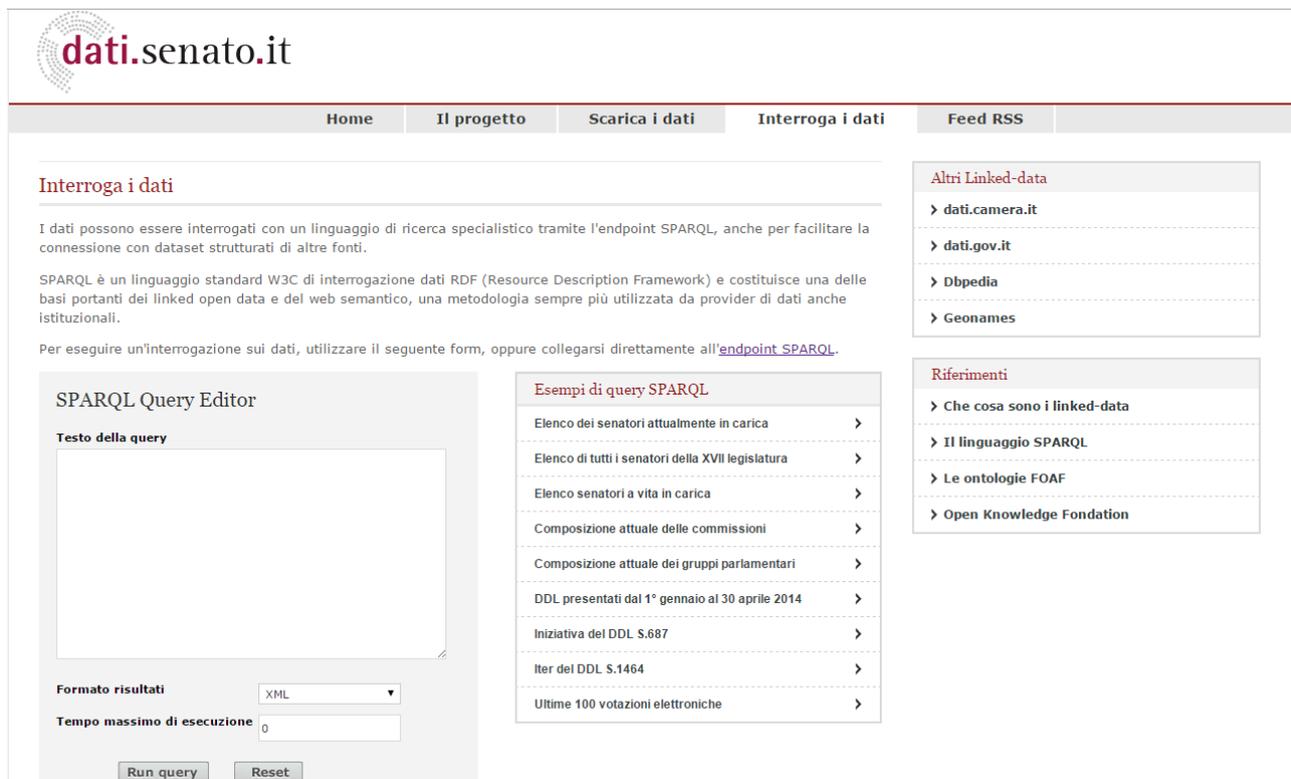


Immagine 12: La sezione "Interroga i dati" del portale open data del Senato della Repubblica.

2.5.4.6 Interroga i dati

È la sezione del sito in cui gli **utenti tecnici** possono ottenere dei dataset specifici eseguendo delle **interrogazioni in linguaggio SPARQL** sull'intero insieme di triple RDF. Il sito presenta un'area di testo all'interno della quale è possibile scrivere un'interrogazione ed eseguirla, specificando il formato di output tra quelli proposti (XML, CSV, JSON, HTML, Javascript, ecc.); si ottiene in tal modo un file contenente il dataset, nel formato specificato. Sono inoltre proposte alcune interrogazioni di esempio, che l'utente può esaminare, copiare in modo automatico nell'area di testo, ed eseguire; ciò allo scopo di semplificare la scrittura di nuove interrogazioni, che possono essere ottenute modificando quelle proposte.

[Interrogazioni in SPARQL](#)

2.5.4.7 Feed RSS

In questa sezione sono riportati i link a tutti i **feed RSS** messi a disposizione dal sito istituzionale del Senato della Repubblica, e che riguardano l'attività parlamentare, i documenti pubblicati, le notizie e i comunicati stampa. Tali feed, infatti, pur non rientrando nell'ambito degli open data, costituiscono comunque uno strumento utile a disposizione di chi sia interessato ad accedere in modo continuo e tempestivo alle informazioni pubblicate dal Senato, sia per la consultazione personale che per un'eventuale elaborazione automatica.

2.5.5 Gli atti parlamentari in formato aperto

Molti dei dati pubblicati dal Senato trattano l'**iter di atti parlamentari**, che, essendo alla base dell'attività ivi svolta, meritano di essere trattati con livelli di accessibilità e fruibilità pari (se non superiori!) a quelli dei dati che li riguardano. Sulla base di questa osservazione il Senato, parallelamente al progetto di portale open data, ha avviato nel 2012 un **progetto per la distribuzione di "open documents"**. Tale progetto verte sulla pubblicazione degli atti relativi all'attività parlamentare in formati che facilitino l'elaborazione automatica dei loro contenuti da parte di soggetti terzi. Il formato scelto per la rappresentazione degli atti è stato quindi il formato denominato Akoma Ntoso, standard

[Apertura dell'iter degli atti parlamentari](#)

internazionale per la marcatura XML di testi normativi adottato da diverse istituzioni internazionali tra cui anche il Parlamento Europeo (per ulteriori dettagli, vedi p. 69). Tale formato è stato concepito per rappresentare dei testi normativi rendendone massimamente usabile il contenuto ai fini di loro elaborazioni automatiche. Pertanto, a partire dalla XVII legislatura, tutti i testi dei disegni di legge presentati in Senato sono pubblicati - oltre che nei formati HTML,PDF, ed ePub - **anche nel formato XML Akoma Ntoso**. Nel medesimo formato sono pubblicati anche i dossier, prodotti dall'amministrazione del Senato con finalità di documentazione degli organi parlamentari e dei parlamentari stessi. È inoltre in corso un progetto per l'estensione dello standard Akoma Ntoso agli emendamenti dell'Assemblea.

La pubblicazione in formato aperto e machine-readable degli atti parlamentari apre la strada a numerosi possibili campi di applicazione (si pensi, a puro titolo di esempio, all'analisi automatica dei testi dei disegni di legge per il calcolo di statistiche sulle frequenze delle parole ovvero dei temi trattati) e fornisce un ulteriore contributo alla trasparenza e al controllo democratico delle istituzioni. Tale prospettiva inoltre – se condivisa a livello interistituzionale – apre la strada all'alimentazione automatica delle banche normative e dei lavori preparatori quale ad esempio il portale della normativa vigente "Normattiva" (vedi p. 69), e quindi anche al miglioramento – sotto il profilo del supporto tecnico – delle attività di semplificazione normativa.

2.5.6 Evoluzioni future

Prossimi passi: più dati e più collegamenti, semplicità e standardizzazione.

Nell'ottica di fornire un servizio di apertura dei dati e degli atti improntato agli standard e alle migliori pratiche in ambito di Web semantico, sono attualmente previste delle evoluzioni che riguardano sia i dati e gli atti pubblicati che la modalità di fruizione degli stessi. In particolare:

- **Estensione dei dataset pubblicati**, sia in termini temporali, con la pubblicazione dei dati di composizione riferiti alle legislature precedenti la XIII, sia relativamente ai contenuti, con particolare riferimento alle informazioni di catalogazione degli atti. I riferimenti alla classificazione TESEO dei disegni di legge attualmente pubblicati potranno essere estesi con la pubblicazione in formato SKOS/RDF dello stesso thesaurus;
- **Maggiore collegamento dei dati** ad altre fonti di dati aperti. Attualmente le uniche relazioni presenti sono quelle verso i dati aperti della Camera dei deputati. E' pensabile collegare i dati relativi a persone, luoghi e professioni a fonti esterne, quali DBpedia e GeoNames;
- **Semplificazione della navigazione** tra i dati, mediante meccanismi di anteprima dei dataset, che consentano di esaminarne i contenuti direttamente nel sito, senza necessità di scaricarli, e tramite l'introduzione di editor visuali che permettano di navigare tra i dati evidenziando le relazioni presenti tra di essi.
- **Ampliamento delle tipologie di atti pubblicate nello standard Akoma Ntoso** a partire dagli emendamenti presentati per la discussione in Assemblea e – a seguire – per altre tipologie di atti (in primis, resoconti).

2.6 I dati aperti sulle norme

2.6.1 Le leggi: dato aperto per eccellenza

Il corpus normativo di uno Stato è composto dalle norme di vario rango (Costituzione, leggi costituzionali, leggi ordinarie, leggi regionali, provvedimenti regolamentari di vario tipo) che esprimono l'insieme di principi e regole su cui si fonda il patto sociale alla base della convivenza civile.

Tutti sono tenuti a conoscere e a rispettare le norme, prodotte dal gruppo ristretto di persone che

detiene temporaneamente il potere legislativo sulla base della “delega” conferita attraverso le elezioni, e pertanto la possibilità che esistano dei cittadini privi di strumenti di accesso alle leggi sembrerebbe un paradosso. Nella realtà, **la conoscenza della legge è stata per lungo tempo appannaggio di una élite** e solo di recente, dopo un percorso lungo e accidentato, si è creata la diffusa consapevolezza di come il libero accesso alle leggi costituisca un diritto fondamentale; ancora negli anni 2000 - infatti - l'accesso per via telematica alle norme, in Italia e in altri paesi europei, era a pagamento, e inoltre il servizio, anche quando veniva erogato da istituzioni pubbliche, era concepito per rivolgersi ad una utenza professionale (su questo aspetto vedi il [Rapporto Mandelkern su Better Regulation](#)).

2.6.2 La rete delle leggi

Un corpus legislativo consiste in un insieme di **documenti testuali**. I testi delle leggi trattano gli argomenti più disparati ed evolvono continuamente: le disposizioni contenute in una legge si collegano a quelle contenute in leggi precedenti richiamandole, modificandole o abrogandole; pertanto i documenti legislativi contengono numerosi collegamenti a documenti preesistenti. D'altra parte, il contenuto di ciascun provvedimento, per effetto dei richiami modificativi provenienti da provvedimenti successivi, varia nel tempo dando luogo a più “versioni”, ciascuna con un arco temporale di validità.

Si può quindi immaginare il **corpus normativo rappresentato come un grafo orientato che evolve nel tempo**: ogni norma è un nodo in cui

Il grafo delle norme.

“entrano” archi che rappresentano i richiami provenienti da altre norme successive e da cui “escono” archi che rappresentano i richiami verso norme precedenti. L'applicabilità della legge, che di per sé contiene l'esplicitazione di principi generali, necessita della definizione di procedure e modalità operative di attuazione che vengono definite in provvedimenti di natura regolamentare - i provvedimenti attuativi - emanati successivamente da parte di altri organismi. Ancora, la giurisprudenza relativa agli esiti dei processi condiziona l'interpretazione delle leggi che costituiscono la base giuridica delle sentenze, determinando altre possibili relazioni dovute ai richiami sentenze-leggi e sentenze-sentenze per conformità o difformità. Il grafo quindi si estende ad altri nodi con altri archi che esprimono ulteriori relazioni.

Ne risulta che un cittadino che voglia comprendere il contenuto di una legge – accantonato l'ostacolo della complessità del linguaggio - deve percorrere il grafo, lungo gli archi che rappresentano le relazioni tra gli atti.

Appare evidente che il **corpus normativo** - o in senso più generale il corpus giuridico che include anche giurisprudenza e dottrina - **costituisce di per sé una “nuvola”**, fortemente connessa al suo interno, inserita all'interno della più ampia Linked Open Data cloud (su cui vedi p. 31).

2.6.3 Gli standard di identificazione e rappresentazione degli atti normativi

Negli ultimi 15 anni si è svolta una intensa attività di ricerca e di sperimentazione relativamente all'informatizzazione dei documenti normativi che ha riguardato in particolare diversi paesi europei, portando inoltre alla costituzione di una comunità internazionale informale; essa, nel corso di workshop annuali su “legislative XML” (si vedano ad esempio gli atti del [3° workshop on legislative XML](#), 6-8 aprile 2005), ha dato luogo ad un proficuo confronto tra le elaborazioni e sperimentazioni che si andavano via via sviluppando. All'interno di questa comunità, l'Italia ha costituito un punto di riferimento importante, avendo anticipato i tempi con le attività di standardizzazione e di sperimentazione all'interno del progetto Normeinrete. Le attività di standardizzazione - di cui nel seguito si descrivono gli esiti - hanno prodotto risultati importanti, grazie ai quali oggi è disponibile un bagaglio di strumenti e di metodologie che rendono più facile applicare i principi degli Open Data e dei Linked Open Data ai documenti legislativi.

Il modello IFLA-FRBR

I più recenti tra gli standard di identificazione degli atti normativi fanno riferimento al **modello IFLA-FRBR** (Functional Requirements for Bibliographic Records), che, concepito per la catalogazione di risorse bibliografiche, risulta molto efficace ai fini della distinzione dei diversi livelli in cui si articolano le risorse informative legislative. Tale modello prevede che la descrizione di una risorsa possa riferirsi a **4 livelli di astrazione**:

1. work: identifica un'opera, una creazione intellettuale in astratto; nel nostro caso, la legge come concettualizzazione delle previsioni giuridiche contenute in tutte le sue versioni che si succederanno nel tempo;

2. expression: identifica ogni diversa versione di un'opera; quindi, per le leggi, ogni nuova versione originata da modifiche apportate da altre norme, vigente per un certo arco temporale o una traduzione di una versione in una lingua diversa dall'originale;

3. manifestation: una concreta realizzazione di una expression, che per un'opera in genere può riguardare il mezzo di riproduzione; nel caso delle leggi, si riferisce al formato di rappresentazione del testo (ad es: un file XML, PDF, etc.)

4. item: una singola copia di una manifestation; nel nostro caso, il file che contiene una copia della manifestation di una expression.

Per approfondimenti, si può consultare Joao Alberto de Oliveira Lima, ["An Adaptation of the FRBR Model to Legal Norms"](#) in Proceedings of the V Legislative XML Workshop, 2006, p. 53-65.

2.6.3.1 Gli standard ufficiali italiani

In Italia esistono **due standard tecnici relativi alla rappresentazione digitale degli atti normativi**, pubblicati in Gazzetta Ufficiale come circolari dell'Autorità per l'informatica nella Pubblica Amministrazione (AIPA, oggi confluita nell'Agenzia per l'Italia Digitale, AgID).

Gli standard italiani.

Tali standard definiscono aspetti tecnici e metodologici che riguardano rispettivamente l'identificazione degli atti (circolare AIPA CR/35 -

"Assegnazione dei nomi uniformi ai documenti giuridici" - G.U. n. 262 del 10 novembre 2001) e la rappresentazione dei testi normativi e dei relativi metadati (circolare AIPA CR/40 - **"Formato per la rappresentazione elettronica dei provvedimenti normativi tramite il linguaggio di marcatura XML"** - GU n. 102 del 3 maggio 2002).

I due standard sono stati definiti in maniera partecipata da gruppi di lavoro, costituiti all'interno del progetto Normeinrete, a cui hanno preso parte esperti di Camera e Senato, di diverse Amministrazioni centrali e regionali e di università ed istituti di ricerca.

Il modello di Normeinrete e i Linked Open Data

Normeinrete è stato un progetto intersettoriale, avviato nel 1999 con finanziamento dell'AIPA su proposta del Ministero della giustizia, finalizzato all'attuazione di un **modello federato per l'informatizzazione delle norme** ed ideato nella prospettiva di una completa automazione dei processi collegati agli atti legislativi – dalla redazione, alla gestione informatizzata delle modifiche, all'accesso libero per i cittadini. Sviluppato con approccio incrementale, nelle prime fasi ha prodotto gli standard di identificazione e marcatura degli atti, su cui si basava l'interoperabilità all'interno della "federazione", ed un portale di accesso che consentiva di effettuare ricerche su tutti i siti aderenti (di oltre 50 amministrazioni, centrali e regionali) attraverso un'unica interfaccia. Sono stati inoltre realizzati dei [pacchetti software opensource](#) di supporto all'utilizzo degli standard, per il drafting normativo e per il riconoscimento automatico all'interno di testi di leggi dei riferimenti normativi e la costruzione del relativo ID standard, tuttora utilizzati. La prospettiva immaginata era quella di costruire uno scenario in cui ogni autorità emittente di provvedimenti normativi pubblicasse, con valore legale, gli atti di propria competenza sul proprio sito istituzionale. Un modello di questo tipo consente di realizzare un sistema distribuito dotato, grazie agli standard, di un **alto livello di integrazione**, ma al tempo stesso **rispettoso dell'autonomia di ciascuna istituzione** e dell'indipendenza delle rispettive scelte tecnologiche implementative; elemento, quest'ultimo, rilevante in ambito normativo, considerata la potestà legislativa delle Regioni e la dimensione sovranazionale dovuta dall'appartenenza all'Unione Europea. L'interoperabilità all'interno della "federazione di siti" basata sulla

standardizzazione degli identificatori (URI) e sulla definizione di uno standard di formato degli atti che consentisse di rappresentare anche aspetti semantici tra cui quelli connessi alle relazioni tra gli atti, dava luogo ad un **modello del tutto analogo a quello su cui oggi si basano i Linked Open Data**. La portata innovativa del progetto ha riguardato anche le modalità di attuazione - basate su un forte coinvolgimento “dal basso” e sulla definizione partecipata di standard aperti. Soprattutto per quanto riguarda i risultati nel campo della standardizzazione, il progetto ha costituito un punto di riferimento a livello internazionale, influenzando iniziative di altri paesi sviluppatesi negli stessi anni e ispirandone altre (come, ad esempio, il progetto UNDESA per i Parlamenti africani, nell’ambito del quale è stato definito lo standard Akoma Ntoso – su cui vedi p. 69).

2.6.3.1.1 L'identificazione degli atti con lo standard URN

Lo standard per la identificazione degli atti è costruito sulla base della specifica [URN IETF](#), ed assume il namespace URN:NIR.

URN: un nome che identifica (univocamente) una risorsa all'interno di un dato dominio.

L'URN è costruito concatenando, secondo una ben definita grammatica, i dati caratteristici dell'atto che lo identificano univocamente (tipo di provvedimento, anno, numero); livelli di dettaglio ulteriori, non obbligatori, consentono di distinguere tra versioni diverse del medesimo atto (ad es. versioni relative a testi con diverse date di vigenza per modifiche successive).

Le caratteristiche più significative del **sistema di identificazione degli atti definito in Normeinrete** sono:

1. **la possibilità di costruire l'URI in maniera automatica**, in quanto generato sulla base di dati noti, utilizzati comunemente per identificare l'atto (tipo di provvedimento, anno, numero);
2. **il disaccoppiamento tra l'URI e la locazione fisica** di una rappresentazione dell'atto: l'URI/URN identifica l'atto “in astratto” (a livello di expression, nella gerarchia FRBR) assicurando la persistenza. Di contro, è sempre necessario un risolutore che associ all'URN una o più locazioni fisiche presso le quali sono disponibili le diverse versioni dell'atto.
3. **la possibilità di realizzare dei parser in grado di riconoscere automaticamente la presenza di riferimenti normativi** all'interno di un testo libero con un bassissimo margine di errore, e di creare automaticamente l'URN e il relativo link ipertestuale verso la legge citata, invocando il risolutore, senza interventi redazionali sulla pagine Web da cui si origina il link.

Le caratteristiche evidenziate semplificano la gestione dei link e ne assicurano la persistenza: al variare della locazione di un atto dovrà essere modificato solo il risolutore mentre tutti i siti che contengono link a quell'atto non vengono influenzati dalla modifica e non presentano il rischio di link “rotti”; meccanismo analogo ai [Domain Name Systems](#).

2.6.3.1.2 Lo standard basato su XML per la rappresentazione degli atti

Lo standard XML per la rappresentazione del contenuto degli atti definito nell'ambito di Normeinrete consente:

1. la **marcatatura strutturale dell'atto**, con un livello di granularità che consente di identificare, e quindi riferire, tutte le componenti della struttura del testo (formula iniziale, libro, parte, titolo, capo, sezione, paragrafo, articolo, comma e numeri/lettere interni al comma);
2. la **possibilità di associare metadati al documento o a sue parti**, quali descrittori (ad es. estremi della pubblicazione ufficiale, eventuali ripubblicazioni, URN, keyword, nome giornalistico dell'atto – ad es. “legge Fornero”), lavori preparatori o altri metadati relativi a schemi che è possibile associare;
3. la **marcatatura, all'interno del documento, di porzioni di testo non identificabili**

come elementi strutturali ma rilevanti per dal punto di vista semantico (ad es. citazioni ad altre norme con la qualificazione del riferimento – modificativo, abrogativo, ecc. – disposizioni caratterizzanti l'atto e simili)

4. la **rappresentazione, all'interno del documento, delle variazioni del testo intervenute nel tempo** con associate le relative date di validità, attraverso l'apposizione di elementi di marcatura che consentono la rappresentazione cosiddetta "multivigente".

Le finalità sottostanti alle scelte operate nel definire lo standard si riconducono a:

1. **consentire la realizzazione di un sistema di ricerca unitario** in grado di offrire funzionalità avanzate ed un elevato grado di precisione anche operando su una base documentale distribuita, con nodi gestiti da soggetti diversi con sistemi informatici disomogenei;
2. **automatizzare la creazione dell'ipertesto distribuito delle norme**, attraverso le citate possibilità di parsing e di costruzione automatica degli URN per la creazione dei link ipertestuali;
3. la possibilità, attraverso la rappresentazione del testo multivigente, di **realizzare applicazioni in grado di ricostruire dinamicamente il testo vigente** in funzione di una data richiesta;
4. la possibilità di **realizzare sistemi di supporto al drafting normativo e alla costruzione dei testi consolidati**. Tale opportunità rende possibile ipotizzare uno scenario in cui l'atto, dal momento della sua creazione, si va arricchendo di metainformazioni durante le diverse fasi in cui si articola il suo ciclo di vita, attraversando i processi che ne determinano l'evoluzione conservandone la memoria all'interno dello stesso documento. Il documento, essendo stato opportunamente marcato in ogni fase, potrà dinamicamente esporre a richiesta ciascuna delle versioni vigenti in un certo periodo del proprio ciclo di vita.
5. **consentire e promuovere il riutilizzo**. La marcatura, attuata come strumento per garantire un servizio pubblico di accesso ai cittadini, presenta significativi vantaggi anche per i soggetti privati interessati al riutilizzo, anche a scopo commerciale, potendo disporre di testi arricchiti di metadati, dotati di struttura e quindi più facilmente elaborabili per la creazione di servizi a valore aggiunto.

2.6.3.2 Altri standard di identificazione e marcatura XML degli atti normativi

Successivamente, sono stati elaborati **altri standard relativi alla identificazione e alla marcatura XML delle norme, standard che estendono e apportano miglioramenti a quelli elaborati nel progetto Normeinrete**, alcuni dei quali particolarmente rilevanti per il grado di diffusione che hanno (o che si prevede avranno nel prossimo futuro) in Italia e nel mondo.

2.6.3.2.1 URN:Lex

Standard di identificazione degli atti, URN:LEX costituisce una **diretta estensione dello standard URN:NIR**, di cui mantiene le regole fondamentali, e quindi le caratteristiche di persistenza e le possibilità di automatismo nella costruzione dell'identificativo, ma ne estende le potenzialità consentendo di identificare provvedimenti di diversi stati o giurisdizioni e di specificarne la lingua. URN:Lex assicura piena compatibilità con il modello FRBR ed è attualmente sottoposto come [proposta di Internet standard all'IETF](#).

2.6.3.2.2 ELI, lo European Legislation Identifier

Standard di identificazione degli atti, proposto dalla UE nelle [Conclusioni del Consiglio che invitano all'introduzione dell'identificatore della legislazione europea \(ELI\)](#).

La specifica di ELI prevede l'utilizzo di «URI HTTP» (IETF [RFC 6570](#)) e definisce criteri e

componenti da adottare per la costruzione di identificativi da parte di ciascuno Stato membro, tenendo conto dei requisiti specifici della propria lingua. I criteri di costruzione dell'ELI, analogamente a URN:NIR e URN:LEX, prevedono una concatenazione dei dati caratterizzanti il provvedimento ma, al fine di favorire la compatibilità con i diversi sistemi nazionali, tutti i componenti sono facoltativi e non hanno un ordine prestabilito. Per favorire l'interoperabilità tra i diversi sistemi di informatica giuridica, ELI è corredato di un set di metadati che gli Stati Membri sono invitati ad utilizzare, in combinazione con i propri set di metadati nazionali. Tra i metadati di cui si raccomanda l'esplicita rappresentazione – oltre il tipo di atto, l'argomento, l'autorità emittente con le sue articolazioni e le diverse date significative per l'atto – sono comprese diverse relazioni tra atti: *cites*, *transpose*, *transposed_by*, *related_to*, *changed_by*, *based_on*, *basis_for*, *consolidates*.

2.6.3.2.3 Akoma Ntoso

Akoma Ntoso, espressione della lingua Akan dell'Africa occidentale che richiama il collegamento tra i popoli (letteralmente "cuori-legati"), è il linguaggio di marcatura degli atti normativi basato su XML e definito nell'ambito di un progetto finalizzato alla informatizzazione dei Parlamenti

Akoma Ntoso, il linguaggio di marcatura degli atti normativi basato su XML.

africani nel contesto del Programma "Africa i-Parliament", nato su iniziativa del Dipartimento Affari Economici e Sociali delle Nazioni Unite (UNDESA). Il progetto, che si è avvalso inizialmente del supporto del CNIPA e poi della collaborazione di diversi esponenti dei gruppi di lavoro che avevano contribuito a definire gli standard di Normeinrete, ha apportato dei miglioramenti allo standard pervenendo alla definizione di un nuovo **linguaggio di marcatura basato su XML più flessibile e di maggiore potenza espressiva**, in quanto è in grado di rappresentare con un unico schema un ampio spettro di tipologie di provvedimenti, incluse le proposte di legge, i report d'aula e di commissione e la giurisprudenza. Attualmente è utilizzato anche dal Parlamento Europeo, e in diversi stati tra cui Svizzera, Brasile, California, Uruguay, Cile. In Italia è utilizzato dal Senato per i testi dei disegni di legge, delle relazioni delle Commissioni e dei messaggi legislativi.

2.6.4 L'accesso agli atti normativi

2.6.4.1 Legislazione statale e regionale

In seguito all'istituzione del Programma per l'informatizzazione della normativa vigente con l'art. 107 della legge n. 388/2000, poi regolamentato con il DPCM 24 gennaio 2003 (che ha istituito un Comitato Guida composto da Senato, Camera e Presidenza del Consiglio, affidando al CNIPA la cura degli aspetti informatici), **Normeinrete ha ceduto il passo al progetto Normattiva**. Esso ha mutuato parte dei risultati conseguiti soprattutto per quanto riguarda gli standard; per quanto riguarda l'architettura, invece, Normattiva ha preferito adottare inizialmente un approccio centralizzato, concentrando gli sforzi sulla creazione dei testi vigenti della normativa statale.

Da Normeinrete a Normattiva.

Normattiva

Attualmente Normattiva offre funzionalità di ricerca sugli atti normativi numerati statali pubblicati in Gazzetta Ufficiale, rendendo disponibili funzionalità di ricerca sui testi vigenti alla data indicata dall'utente. Grazie all'adozione dello standard URN:NIR risulta estremamente **facilitato il processo di predisposizione di link da qualsiasi sito per la visualizzazione di atti normativi**: è infatti sufficiente utilizzare gli estremi dell'atto, ed invocare il risolutore indicando opportunamente il tipo di provvedimento come prescritto dallo standard URN:NIR (ad esempio, per un decreto legge: <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legge:AAAA-MM-GG;NNN>).

Normattiva utilizza inoltre il [software per il parsing normativo](#) rendendo disponibile una funzionalità "Attiva riferimenti normativi" per **trasformare il testo semplice contenente il riferimento ad un atto in un link all'atto stesso**.

Sono attualmente allo studio, in collaborazione con la Conferenza dei Presidenti delle Assemblee legislative delle regioni e delle provincie autonome, alcune **linee di evoluzione del portale** orientate a riprendere un modello architetturale di tipo federato: in tal modo sarà possibile offrire ai cittadini un punto unico di accesso alla normativa regionale e statale preservando al tempo stesso l'autonomia delle Regioni nella pubblicazione della normativa di propria competenza (per una panoramica storica sui progetti d'informatizzazione del patrimonio normativo italiano si può consultare Carlo Marchetti, *Il Senato della Repubblica e l'evoluzione del sistema informativo parlamentare* in *L'informatica giuridica in Italia*, a cura di G. Peruginelli, M. Ragona, ESI, 2014).

Il sistema Normattiva consente di effettuare il **download di singoli di documenti**, rendendo disponibili diversi formati tra cui XML-NIR, soltanto in maniera interattiva. Non sono pertanto disponibili per il download collezioni di norme statali.

Le Regioni dispongono di propri sistemi di pubblicazione degli atti normativi. Molte di esse hanno preso parte a NormeInrete e dispongono quindi di testi identificati da URN e marcati in XML-NIR. Anche per le Regioni sono disponibili funzionalità per il download di singoli provvedimenti, ma non per collezioni.

2.6.4.2 Atti parlamentari

I siti Web di Camera e Senato.

I siti delle istituzioni parlamentari, Camera dei deputati e Senato della Repubblica, rendono **disponibili alla ricerca e all'accesso tutti gli atti parlamentari**; tra essi, i disegni di legge del Senato sono disponibili anche

nel formato definito dallo standard XML Akoma Ntoso sopra descritto. Le istituzioni parlamentari hanno inoltre realizzato siti Open Data ad hoc, rispettivamente dati.camera.it e dati.senato.it, che rendono disponibili molte informazioni relative alle attività parlamentari e ai soggetti che vi prendono parte anche rappresentate nella forma e con le modalità dei Linked Open Data e pertanto arricchite da informazioni sulle relazioni tra entità. Tra queste, sono disponibili come Linked Open Data anche l'elenco dei Disegni di Legge presentati alla Camera e al Senato, gli iter legislativi, le loro variazioni e i relativi interventi dei parlamentari in Aula e nelle Commissioni, scaricabili in dataset omogenei e potendosi anche selezionare il range di date dei documenti in essi contenuti.

2.6.5 Open data e Linked Open Data normativi

2.6.5.1 Open data normativi

Open Data e Linked Open Data.

L'applicazione dei principi dell'Open data nel contesto delle informazioni a carattere normativo può riguardare la disponibilità di collezioni di atti normativi o di Linked Open Data che rendano accessibili in forma di triple

RDF (vedi sopra p. 31) le relazioni tra atti.

La disponibilità di intere collezioni di testi in formato aperto disponibili per il riutilizzo, corredate di opportune licenze d'uso e di metadati, non è attualmente presente per la legislazione italiana. Tale opportunità è invece prevista per il contenuto della banca dati EUR-Lex, che prevede la possibilità di scaricare e riutilizzare gratuitamente il contenuto della banca dati, reso disponibile al download in XML dopo aver effettuato una richiesta tramite un modulo on-line.

Si è prima illustrato (vedi p. 65 e ss.) come gli elementi propri degli atti legislativi rendano il **corpus normativo "naturalmente" idoneo a essere rappresentato come Linked Open Data**; e come (vedi p. 67 e ss.) l'adozione degli standard di marcatura XML e di identificazione, in grado di riferire un intero atto o una sua parte tramite URI, consentano di esprimere agevolmente relazioni tra atti – o tra parti di essi – e implementare il modello dei Linked Open Data.

Ognuno dei metadati associati a una norma ed ognuna delle relazioni esistenti tra due atti può dare origine ad una tripla RDF il cui insieme implementa la **"nuvola giuridica"** prefigurata

inizialmente. Le innumerevoli altre triple che esprimono relazioni tra gli atti normativi e risorse informative di altri domini – come ad esempio entità corrispondenti a concetti, luoghi, persone o eventi pubblicati da DBpedia – ne determinano l'inserimento all'interno della più ampia Linked Open Data Cloud (su cui vedi sopra, p. 30 e p. 45).

2.6.5.2 Metadati

Agli open data normativi si applicano le considerazioni che possono farsi per gli open data in genere relativamente ai metadati (ovvero, alle informazioni che descrivono un insieme di dati).

I **metadati fondamentali da associare agli atti legislativi** [Metadati degli atti normativi](#) comprendono:

1. l'**autorità emittente**;
2. il **tipo di atto** (legge, decreto legislativo, decreto-legge, legge regionale, etc.);
3. l'**ambito territoriale di applicabilità**;
4. la **classificazione per materia**, secondo uno o più schemi;
5. le **date d'interesse per l'atto o per le sue parti** (data di pubblicazione, data di entrata in vigore, data di abrogazione, etc.);
6. la **lista dei documenti correlati** (atti modificanti e modificati, proposte di legge da cui il provvedimento ha avuto origine, giurisprudenza correlata, pronunce della Corte Costituzionale che lo riguardano, relazione con direttive europee eventualmente recepite dall'atto, etc.);
7. gli **alias con cui il documento conosciuto**;
8. i **riferimenti di pubblicazione ufficiale**;
9. il **link alle locazioni fisiche su cui sono disponibili copie dell'atto** (manifestation) e CC-BY 3.0 Unported **informazioni sul relativo editore**.

Ai fini di favorire l'interoperabilità semantica è necessario che per la rappresentazione di metadati si utilizzino il più possibile **vocabolari già esistenti e condivisi**, come ad esempio [Dublin Core Metadata Element Set](#) per la descrizione delle informazioni di tipo bibliografico.

Insieme a questi, potranno poi essere integrati vocabolari specifici: i principali tra di essi sono descritti brevemente nei paragrafi che seguono.

2.6.5.2.1 Schemi di classificazione

Un elemento importante per le collezioni documentali legislative è costituito dallo **schema di classificazione adottato per i metadati descrittivi**.

Nell'attuale scenario non esiste uno schema adottato universalmente; **i due principali sono Teseo ed EuroVoc**. Entrambi sono attualmente rappresentati in RDF/SKOS, anche se Teseo non è ancora disponibile pubblicamente in tale formato. La possibilità di rappresentare entrambi gli schemi in RDF/SKOS offre la possibilità di stabilire collegamenti tra Teseo ed Eurovoc, creando i presupposti per un'eventuale conciliazione tra i due schemi, con le conseguenti possibilità di integrazione anche multilingua tra sistemi di ricerca di atti parlamentari.

2.6.5.2.2 TESEO

TESEO (TEsauro SEnato per l'Organizzazione dei documenti parlamentari) è uno **schema di classificazione usato nelle più importanti banche dati del Senato**. Il thesaurus, finalizzato alla classificazione per materia di atti parlamentari e legislativi che abbracciano tutta la realtà sociale, non è specializzato su discipline specifiche e comprende oltre 3.600 descrittori suddivisi in 46 aree tematiche (top terms), organizzati secondo la struttura logica della Classificazione Decimale Universale (CDU).

Il sistema di relazioni gerarchiche in cui sono strutturati i descrittori consente di evidenziare, per

ciascuno di essi, la presenza di termini più ampi (BT = Broader Term), più specifici (NT = Narrower Term), sinonimi (UF = Use For), o eventuali termini associati (RT = Related Term).

2.6.5.2.3 EuroVoc

EuroVoc è un **thesaurus multilingue e pluridisciplinare che comprende la terminologia dei settori d'attività dell'Unione europea**, con un'attenzione particolare per l'attività parlamentare. È disponibile nelle 23 lingue ufficiali dell'Unione europea (bulgaro, ceco, croato, danese, estone, finlandese, francese, greco, inglese, italiano, lettone, lituano, maltese, neerlandese, polacco, portoghese, rumeno, slovacco, sloveno, spagnolo, svedese, tedesco e ungherese), nonché nella lingua di un paese extra UE (serbo).

Il thesaurus EuroVoc viene utilizzato dal Parlamento europeo, dall'Ufficio delle pubblicazioni, da parlamenti nazionali e regionali in Europa, come anche da amministrazioni nazionali e utenti privati di paesi europei e non.

2.7 Un confronto internazionale

In questo paragrafo si presentano i risultati di una rassegna condotta rispetto alla pubblicazione di dati aperti da parte dei parlamenti, a livello internazionale. Le riflessioni svolte alla luce dei risultati di tale rassegna tendono ad evidenziare tanto i principali **aspetti pratici** che possono agevolare il riutilizzatore nella ricerca e fruizione dei dati, quanto le **buone pratiche** attualmente portate avanti in questo senso dai diversi parlamenti.

2.7.1 Nota metodologica sull'analisi dell'offerta dei parlamenti

Il confronto è stato condotto rispetto ad un **numero scelto di parlamenti** europei ed extra-europei. L'ambito dell'analisi, che esprime una presenza maggioritaria di parlamenti europei, è stato da un lato orientato nel senso di includere parlamenti afferenti a quegli ordinamenti che notoriamente rappresentano un modello di riferimento in quanto profittevoli 'praticanti di lungo corso' del paradigma Open Data (tra cui Stati Uniti e Regno Unito), ma non ha pretese di esaustività, nella consapevolezza che un confronto completo delle realtà presenti nello scenario internazionale avrebbe richiesto mezzi e tempi che eccedevano le disponibilità del presente studio. Il numero iniziale di parlamenti considerati è stato ampliato, ed i rilievi che seguono sono stati rifiniti, alla luce delle risposte ad un **questionario** sull'impiego di social media e open data da parte delle istituzioni parlamentari - distribuito in vista del [seminario CERDP From e-Parliament to smart-Parliament](#), organizzato da Camera dei deputati e dal Senato della Repubblica il 12-13 giugno 2014, e messo a disposizione per la stesura del presente Rapporto. Le risposte fornite al questionario sono state oggetto di ulteriore verifica tramite **consultazione diretta dei siti Web** delle istituzioni parlamentari campione; in questo senso, si è volutamente assunto un punto di vista non privilegiato, ossia quello del "semplice" cittadino utente-riutilizzatore che si appresti a navigare il sito istituzionale del proprio parlamento (e quello dei parlamenti di altri paesi) per vagliare la presenza di dati e informazioni.

Rispetto ad ogni parlamento è stata verificata la **disponibilità o meno di informazioni e documenti** in forma di Open Data, in accordo con l'Open Definition (ovvero secondo modalità che ne consentano il libero accesso e riutilizzo), esaminando inoltre il formato impiegato per pubblicare dati e la licenza ad essi associata; in mancanza di licenza esplicita, si è cercato comunque - per quanto possibile - di risalire allo status giuridico dell'informazione sotto il profilo del diritto d'autore, verificando in ultima istanza se essa sia effettivamente riutilizzabile e a quali condizioni. Considerate le peculiari funzioni e la struttura dell'istituzione Parlamento, si è ritenuto di condurre la verifica ricomprendendo i dati eventualmente disponibili entro **quattro tipologie di atti o documenti**: Composizione del parlamento (gruppi, commissioni, etc.); Informazioni sui parlamentari e la loro attività; Atti e processo normativo, votazioni; Documenti non legislativi (es. i

prodotti degli uffici studi); Budget, spese e altre informazioni contabili.

È possibile consultare le informazioni di cui sopra, reperite per ognuno parlamenti esaminati, nel Wiki <http://dati-parlamento.nexacenter.org>; in fondo al capitolo è disponibile inoltre una tabella comparativa (p. 77). Va precisato che, tanto nelle singole pagine quanto nella rilevazione tabellare, ai parlamenti che non pubblicano dati aperti sono stati equiparati quei parlamenti che, non avendo esplicitato chiaramente le condizioni di accesso e riutilizzo rispetto all'informazione presente sul proprio sito Web, non mettono in condizione l'utente-riutilizzatore di sapere con certezza se sia possibile condividere e riutilizzare i contenuti pubblicati. In quest'ultimo caso, per quanto possibile, si è tentato di ottenere un chiarimento via e-mail presso i contatti indicati nei diversi siti istituzionali circa l'effettiva possibilità di riutilizzo (avendosi ottenuto riscontro da parte di Estonia, Georgia e Nuova Zelanda).

Le rilevazioni effettuate mediante consultazione dei siti Web parlamentari sono **aggiornate al 28 novembre 2014**.

E' bene tuttavia segnalare che l'estrema attenzione al tema degli Open Data manifestata da parte dei parlamenti e la viva dinamicità che caratterizza le loro iniziative nel settore fanno ritenere estremamente probabile che lo scenario internazionale sia di fatto soggetto a rapide evoluzioni, sia in termini di scelte tecnologiche operate sia in termini di dimensione e tipologia della offerta di Open Data parlamentari.

2.7.2 Differenti modalità di pubblicazione dei dati

Al di là della principale demarcazione tra parlamenti che pubblicano e parlamenti che non pubblicano dati in modalità 'open', un primo rilievo palese che emerge dalla rassegna condotta riguarda le modalità adottate dai parlamenti nel mettere a disposizione i propri dati.

In alcuni paesi sono stati infatti lanciati **progetti volti alla pubblicazione di open data parlamentari attraverso siti Web dedicati** come dati.camera.it e dati.senato.it in Italia (varati, rispettivamente, dalla Camera dei deputati e dal Senato della Repubblica), data.riksdagen.se in Svezia, data.stortinget.no in Norvegia, data.parliament.uk nel Regno Unito. Tipicamente, questa modalità sottende una serie di decisioni propedeutiche alla scelta esplicita di imbastire un progetto Open Data, relative in particolare al formato, alle soluzioni tecnico-informatiche, ai termini legali e alle condizioni con cui sono messi a disposizione i dati, che per l'appunto si accordano a (quando non ispirano direttamente) le buone pratiche e alle raccomandazioni divenute prassi ormai consolidate presso la maggioranza di quanti pubblicano e riutilizzano dati aperti, nonché presso le community, gli attivisti e gli esperti del settore.

In altri casi, viene adibita una **sezione apposita del sito Web istituzionale** mediante la quale abilitare il download di dati, come nella **sezione bulk data** messa a disposizione dal GPO's Federal Digital System statunitense; oppure, le **informazioni vengono raggruppate ed esposte, a seconda del loro contenuto, presso varie sezioni del sito Web**, adibite ad illustrare la composizione, le funzioni e le competenze del parlamento o di un suo singolo ramo.

Soprattutto in quest'ultimo caso – e in particolare laddove fosse mancata una fase di studio e di progettazione vere e propria, finalizzata alla realizzazione di un processo strutturato di messa a disposizione di dati aperti – la modalità scelta non sempre si accorda in toto con le prassi ottimali o le definizioni individuate dagli standard di pubblicazione degli Open Data (5 stelle di Tim Berners-Lee; Open Data Definition). Questa eventualità non è infrequente dal momento che, quando le informazioni sono 'semplicemente' pubblicate nello stato in cui si trovano, può accadere che si trascurino quegli aspetti capaci di rendere davvero efficace la pubblicazione di dati nell'ottica del loro riutilizzo (come, ad esempio, scelte relative all'apertura tecnica e giuridica dei dati, l'adozione di una licenza aperta o di una nota legale che espliciti le condizioni di riutilizzo, la realizzazione di

guide o altro materiale informativo a beneficio dell'utente-riutilizzatore), sovente confidandosi nell'esistenza di una generica - pur se opportuna e desiderabile in se stessa - previsione di legge che esclude l'operatività del diritto d'autore su particolari tipologie di informazione pubblica o che le ascrive direttamente nel pubblico dominio.

2.7.3 Lo status giuridico dei dati pubblicati: un'esigenza di certezza

Come si è già accennato nel primo capitolo, l'**incertezza** rappresenta un deterrente primario alle possibilità generative insite nel reimpiego dell'informazione, poste a premessa e promessa del concetto di Open Data, inteso appunto quale riutilizzo inatteso ed ulteriore di dati preesistenti. Questa osservazione risulterà evidente (ed in fondo banale) qualora si consideri che, tanto nel riutilizzo dell'informazione quanto in qualsivoglia altra forma di traffico o atto umano, prima ancora che giuridico, l'esigenza insopprimibile che guida ogni azione è appunto la certezza in ordine ai presupposti e alle conseguenze di tale agire.

Note legali e termini d'uso.

In questo senso, si deve constatare come l'eterogeneità riscontrata circa le modalità di pubblicazione dei dati si rispecchi nelle differenze apprezzate rispetto alla predisposizione di **avvisi, disclaimer o note informative** che illustrano il regime giuridico sui dati pubblicati e le eventuali condizioni poste per la fruizione e circolazione degli stessi. Mentre la maggioranza dei parlamenti considerati esplicita i termini d'uso o include una nota sul copyright relativo ai contenuti del sito Web, in alcuni casi non se ne è riscontrata la presenza (così, ad esempio per i parlamenti di Andorra, Nuova Zelanda, Serbia e Slovacchia). In altri casi, invece, pur essendo reperibili termini d'uso e/o una copyright notice, essi non risultano poi chiari nell'esplicitare se i dati siano effettivamente aperti (così in Bosnia, Repubblica Ceca e Slovenia).

Certezza giuridica e chiarezza terminologica.

L'assenza di tali strumenti informativi o la loro poca chiarezza risulta decisamente deleteria per ogni istituzione parlamentare che si sia impegnata nel processo di messa a disposizione del proprio patrimonio informativo; ciò proprio perché si rischia così di minare in radice l'efficacia stessa – e quindi ogni ricaduta positiva – della scelta che si è operata nel senso dell'apertura. La mancanza di chiarezza circa il regime delle informazioni e dei dati, pur di per sé disponibili online, rappresenta infatti uno scoglio pratico per qualsiasi riutilizzatore accorto che pretenda di avere un minimo di **certezza rispetto allo status legale dei dati**, prima di impiegarli. Ciò è ancor più vero per i ri-utilizzatori c.d. 'cross-border', i quali, oltre allo scoglio costituito dalla lingua, verosimilmente non sono a conoscenza della normativa nazionale adottata in un ordinamento straniero in tema di accesso alle informazioni pubbliche, dell'applicabilità o meno di norme relative al diritto d'autore sull'informazione pubblica e della possibilità o meno di riutilizzare i dati messi a disposizione dal parlamento secondo le norme relative alla Public Sector Information, come recepite nazionalmente. Risulterà evidente, quindi, quanto la necessità di scovare e ricostruire da sé il regime giuridico complessivo combinando assieme le note di copyright, i termini di utilizzo del sito Web e il quadro normativo di riferimento definito da fonti legislative - nella peggiore delle ipotesi non richiamate nelle note legali - sia costoso e dissuasivo.

Sempre a beneficio della chiarezza verso gli utenti, qualora siano già presenti, può essere utile effettuare un esercizio di **aggiornamento e coordinamento delle note legali** del sito: includendovi cioè un riferimento esplicito al quadro normativo circa l'accesso e riutilizzo dell'informazione (copyright - o assenza di copyright – sull'informazione; norme di recepimento della Direttiva PSI e norme correlate in tema di accesso e riutilizzo, come il CAD e il Decreto Trasparenza italiani) ed alle scelte operate rispetto alle modalità di apertura (ad es., adozione di un modello di licenza standard per l'impiego dei dati); come pure richiamando nelle note legali del sito Web del parlamento o del ramo parlamentare (ad es., www.camera.it) eventuali condizioni per il riutilizzo previste all'interno delle note legali del servizio Open Data, se fornito su un sito Web autonomo (ad es., dati.camera.it) e nel caso in cui le note legali del primo sito Web non menzionino le condizioni previste dalle note legali del secondo.

Altrettanta cura andrebbe prestata rispetto alla **terminologia** che si intende impiegare: ad esempio, per illustrare il c.d. divieto di misrepresentation (cioè di fare un riutilizzo dell'informazione tale da trarre in inganno terzi o che l'informazione di partenza venga travisata) [Svizzera](#) e [Spagna](#) adottano una terminologia potenzialmente fuorviante che rischia di far ritenere che sia vietato anche modificare l'informazione tout court o quel tanto necessario al fine di creare un'opera derivata (ciò che si verifica – banalmente – ogni volta che il dato non sia riprodotto tale e quale; ad esempio poiché inserito in una diversa banca dati assieme ad altri dati, o poiché maneggiato ulteriormente al fine di trarre da esso inferenze, unitamente ad altri dati e informazioni). Oltretutto in entrambi gli esempi risulta già prevista un'autonoma condizione che vieta di alterare il senso o di lasciare intendere che l'informazione riutilizzata abbia natura ufficiale.

2.7.4 Reperibilità e fruibilità dei dati

Si è accennato poco sopra come proprio la lingua rischi di tramutarsi involontariamente in un ostacolo alle possibilità di riutilizzo. Considerare l'eventualità di **fornire anche una versione in inglese** del sito Web

Versione inglese e look & feel del sito.

relativamente alla sezione Open Data, o per lo meno rispetto ad alcune pagine-chiave, come quelle descrittive della funzione del servizio, dei contenuti e delle feature offerte, ridurrebbe di molto le difficoltà legate alla consultazione delle risorse. Comprensibilmente, non si può pretendere che siano oggetto di traduzione anche i singoli dati e contenuti (proposte di legge, votazioni, resoconti delle sedute, documentazione tecnica etc.), ma l'ambizione di essere **foreign user-friendly** rappresenta di certo una buona pratica; soprattutto se si ambisce a porsi nell'ottica di agevolare e stimolare l'impiego della propria base di dati presso la più vasta platea di riutilizzatori. Quanto detto risulterà evidente mediante la diversa impressione che si ricava approdando rispettivamente alle homepage della piattaforma open data [svedese](#) e di quella [norvegese](#).

Anche lo stesso **livello grafico-strutturale** può essere valorizzato rendendolo tanto più efficace quanto più visibile ed esplicito: il posizionamento di link testuali o icone che puntano alla sezione open data andrebbe operato in modo da far risaltare quest'ultima il più possibile. L'ideale sarebbe rappresentato da una voce "Open Data" da inserire direttamente nel sommario del sito istituzionale (la quale, constando di un termine inglese risulta una voce evidente, autoesplicativa e ancor più chiara di [data.riksdagen.se](#) o [dati.senato.it](#)); un link o un'icona posizionato in un angolo, in testa o in calce al sito Web, potrebbe invece correre il rischio di passare inosservata al visitatore meno attento.

Infine, nuovamente per le opportunità sopra esposte legate alla facilità di consultazione, non pare disprezzabile la scelta operata in alcuni paesi di riunire le sezioni relative ai due rami all'interno di un singolo sito relativo al parlamento (così, tra gli altri, in Austria, Australia, Svizzera e Regno Unito). Dal momento che modificare radicalmente in tal senso la scelta architettonica comporterebbe verosimilmente un iter procedurale non banale, da sancirsi inoltre con accordi e formalità pratico-operative, risulta probabilmente più ragionevole lo sforzo di **facilitare al massimo l'accessibilità alle informazioni pubblicate o alle sezioni open data a partire da tutti i siti istituzionali esistenti** (sito del parlamento, sito dei rami del parlamento, eventuali siti open data dedicati), provvedendo anche a segnalare – come già detto – nelle note legali di ognuno di essi le condizioni previste per il riutilizzo di dati e informazioni.

2.7.5 Buone pratiche

In questo paragrafo vengono illustrate alcune buone pratiche nella messa a disposizione dei dati, corredate dall'implementazione concreta che si è riscontrata in essere presso vari parlamenti.

Realizzare una **sezione del sito o una piattaforma Open Data ad hoc** entro cui raccogliere ed organizzare tutti i dati e le informazioni che si intendono aprire. Questa opzione presuppone innanzitutto una scelta consapevole; la quale (generalmente ed auspicabilmente) implica l'esame e

l'adozione di modalità espositive maggiormente strutturate – soprattutto dal punto di vista tecnico – rispetto alla 'mera' possibilità di fruizione dei contenuti, in qualsiasi modo siano stati pubblicati, che dovesse discendere unicamente da una previsione di legge o dall'assenza di vincoli normativi (come nel caso di informazioni non soggette a copyright).

- Il sito Open Data del parlamento svedese pubblica dati in formato XML, CSV, JSON, TXT e HTML tramite [API](#). Il sito Open Data del parlamento norvegese data.stortinget.no pubblica dati in formato XML e JSON mettendoli a disposizione ai termini della [Norwegian Licence for Open Government Data](#). Il parlamento svizzero fornisce un'[interfaccia machine-readable aperta](#) in cui i dati sono disponibili in formato XML, JSON, XSD alle [condizioni](#) previste per l'uso del Web service. I portali Open Data italiani della [Camera dei deputati](#) e del [Senato della Repubblica](#) si distinguono per l'impiego di [Linked Open Data](#) e di [ontologie formali](#) per descrivere i dati che vengono pubblicati in formato RDF.

Predisporre una **versione inglese** della pagina o delle pagine che illustrano i contenuti della sezione o del sito Open Data.

- Il sito Open Data del Riksdagen svedese fornisce una [pagina introduttiva in inglese](#) contenente sintetica descrizione delle condizioni per il riutilizzo, delle tipologie di dati disponibili, delle API e dei contatti dei responsabili del sito. Sebbene non sia disponibile una versione inglese delle altre parti del sito che ne agevoli la consultazione rispetto alle relative sezioni (Data, Dokumentation), si tratta comunque di una scelta apprezzabile, per agevolare e quindi potenzialmente ampliare la base di visitatori, scelta che soprattutto tiene conto dell'esistenza di eventuali riutilizzatori o utenti stranieri.

Fornire una **guida utente**, contenente descrizione dei dataset e relativa documentazione tecnica, che accompagni l'utente nei diversi passi necessari per il download dei dati.

- L'US Government Printing Office mette a disposizione una [User Guide](#) per i bulk data sulle proposte di legge dell'House of Representatives, forniti in XML. Il portale Open Data norvegese fornisce una [guida](#) contenente la descrizione della base dati e ulteriori ragguagli tecnici (disponibile solo in norvegese). Il parlamento svizzero ha realizzato un [documento informativo](#) sull'utilizzo del proprio Web service sui dati aperti.

Corredare il sito di una **nota legale esaustiva** e comprensiva di tutte le informazioni rilevanti per il riutilizzatore: ad es., status giuridico dell'informazione, con particolare riguardo per la normativa sul diritto d'autore; termini di riutilizzo dell'informazione in accordo con la normativa sulla PSI e altre disposizioni relative ad accesso e riutilizzo; menzione della licenza eventualmente adottata per condividere dati e informazioni, recapiti del soggetto o dell'ufficio responsabile per il riutilizzo dell'informazione (e-mail; contatto telefonico).

- Il sito dell'Oireachtas irlandese contiene una [copyright and re-use note](#) che regola il riutilizzo delle informazioni (richiamandolo esplicitamente anche nel titolo della nota sul copyright) e che fornisce un quadro completo circa le condizioni previste, la normativa di riferimento, lo status di ufficialità dei dati pubblicati ed il personale incaricato di fornire informazioni sul servizio.

Esplicitare una **policy sui dati personali** che avvisi il riutilizzatore rispetto alle responsabilità inerenti al trattamento di dati personali che dovessero essere presenti nei dataset, richiamando ad un riutilizzo conforme alle disposizioni vigenti in materia.

- Il sito Open Data del Riksdag svedese include tra i propri [Terms of use](#) una policy esplicita sull'impiego di dati personali secondo la quale il riutilizzatore di dati contenenti dati personali deve assicurarsi di rispettare le disposizioni previste dal Personal Data Act (1998:204).

Collegare l'iniziativa di apertura dei dati ad obiettivi strategici più ampi che mirino non solo al conseguimento di benefici socio-economici esterni in termini di conoscenza, trasparenza,

inclusività e innovazione, ma anche di ricadute positive interne che influiscano sull'efficacia, efficienza e semplificazione dei processi.

- Il Riksdag svedese ha sviluppato una [Open Data Strategy](#) volta a sfruttare le potenzialità anche internamente: il sito Web del Riksdag e l'applicazione Kammarappen per iPhones (disponibile per i membri del parlamento e che tiene traccia di attività e procedure) sfruttano gli open data. Anche l'Intranet del Riksdag sfrutta in parte i dati aperti del Riksdag ed è stato sviluppato un procedimento standard per la pubblicazione dei dati affinché questi siano resi disponibili sul sito Open Data, recuperabili attraverso motore di ricerca ed esposti inoltre sul sito istituzionale del Riksdag [fonte: risposte questionario CERPD].

Promuovere e sostenere l'iniziativa di apertura dei dati attraverso **hackathon e occasioni di confronto** tra il parlamento, la comunità di riutilizzatori, le imprese e i cittadini.

- In Italia, la Camera dei deputati ha organizzato nel 2014 l'hackathon [Code4Italy@Montecitorio](#) per fornire un contesto entro cui ampliare la conoscenza sui dati aperti parlamentari resi disponibili a partire dal 2011 e per promuoverne il riutilizzo da parte della comunità di sviluppatori e di tutti i soggetti interessati. In Norvegia, il servizio Open Data [data.stortinget.no](#) è stato promosso all'interno di un hackathon organizzato dall'Agency for Public Management and eGovernment (Difi) [fonte: risposte questionario CERPD].

2.7.6 Approfondimenti utili

[Openingparliament.org](#), il forum di organizzazioni specializzate in monitoraggio parlamentare, ha diffuso una [survey sulla pubblicazione di dati aperti relativi alle votazioni](#) da parte delle Camere elettive di duecento giurisdizioni, realizzata da Kamil Gregor – data analyst presso [KohoVolit.eu](#) e l'Università di Masaryk.

2.7.8 Tabella riassuntiva

Paese	Tipologia di Licenza?	Quali informazioni?	In quali formati?	Dove si trovano i dati?
Austria	Licenza aperta standard	Atti e processo normativo	XML RSS	https://www.data.gv.at/auftritte/?organisation=parlament
Irlanda	Licenza aperta nazionale	Composizione del parlamento Gruppi Commissioni Atti e processo normativo Budget Spese e altre informazioni contabili Votazioni Informazioni sui parlamentari e la loro attività	PDF HTML File di testo	http://www.oireachtas.ie/parliament/

Paese	Tipologia di Licenza?	Quali informazioni?	In quali formati?	Dove si trovano i dati?
Italia	Licenza aperta standard	Composizione del parlamento Gruppi Commissioni Atti e processo normativo Votazioni Documenti non legislativi Informazioni sui parlamentari e la loro attività	RDF XML CSV JSON HTML	
Norvegia	Licenza aperta nazionale	Atti e processo normativo Votazioni Composizione del parlamento Gruppi Commissioni	JSON XML	http://data.stortinget.no/
Olanda		Atti e processo normativo	HTML PDF XML ODF	https://zoek.officielebekendmakingen.nl/zoeken/parlementaire_documenten
Regno Unito	Licenza aperta nazionale	Composizione del parlamento Gruppi Commissioni Informazioni sui parlamentari e la loro attività Atti e processo normativo Votazioni Documenti non legislativi Budget Spese e altre informazioni contabili	HTML XML PDF	http://www.data.parliament.uk/
Stati Uniti		Atti e processo normativo	XLM	http://www.gpo.gov/fdsys/bulkdata
Svezia		Informazioni sui parlamentari e la loro attività Atti e processo normativo Votazioni Documenti non legislativi	JSON XML SQL HTML TXT CSV	http://data.riksdagen.se/

Paese	Tipologia di Licenza?	Quali informazioni?	In quali formati?	Dove si trovano i dati?
Svizzera		Composizione del parlamento Gruppi Commissioni Votazioni Atti e processo normativo	XML JSON XSD	http://www.parlament.ch/e/dokumentation/web-services-opendata/Pages/default.aspx

I seguenti Parlamenti sono stati monitorati, ma non pubblicano Dati Aperti (ai sensi della Open Definition) oppure pubblicano dati senza esplicitarne chiaramente le condizioni di accesso e riutilizzo: Albania, Andorra, Australia, Bosnia, Canada, Estonia, Finlandia, Francia, Georgia, Germania, Grecia, Lituania, Nuova Zelanda, Portogallo, Repubblica Ceca, Russia, Serbia, Slovacchia, Slovenia, Ungheria.

Per consultare le schede relative ai singoli parlamenti si rinvia al wiki Open Data in ambito parlamentare, disponibile all'indirizzo <http://dati-parlamento.nexacenter.org/wiki>.

Cap. 3. La domanda di dati aperti parlamentari

Il capitolo evidenzia i principali aspetti legati alla domanda di dati parlamentari unitamente alle impressioni e ai suggerimenti di riutilizzatori di dati parlamentari, ricercatori, esperti e *communities* di open data, come raccolti nel corso di una serie di audizioni informali tenutesi presso i Servizi Informatica di Camera e Senato.

3.1 Il riutilizzo di dati parlamentari: opportunità e aspetti critici

Come è stato evidenziato da soggetti riutilizzatori di dati parlamentari in [Metriche e confronti](#) accordo con la loro esperienza diretta, un momento chiave nel processo di valorizzazione e riutilizzo è costituito dalla 'traduzione' della mole di dati disponibili attraverso **metriche accessibili agli utenti e organi di informazione**, aspetto determinante al fine di consentire **confronti e comparazioni** dotati di oggettività. Come nel caso dei dati relativi alla partecipazione dei parlamentari alle votazioni, veicolati ad esempio dal progetto [Open Parlamento](#) dell'associazione [Open Polis](#) – a partire da dati messi a disposizione Camera e Senato – ed illustrati nel dossier [Indice di Produttività Parlamentare](#), esistono **criticità nel passaggio dal dato pubblicamente accessibile al dato pubblicato a valle**, le quali si sostanziano nella forma di un "trade-off". Infatti, la semplice esposizione di dati grezzi – al fine, ad esempio, di permettere di valutare la produttività dei parlamentari – assicura, da un lato, una certa oggettività, ma rischia di consentire valutazioni solo parziali, e/o, di influenzare il comportamento dei parlamentari al fine di migliorare il proprio "posizionamento" secondo uno specifico criterio. Al tempo stesso, la creazione di indici complessi di valutazione dell'attività parlamentare rischia di influenzare il giudizio "a monte".

La disponibilità di **dati grezzi** rappresenta un elemento prezioso per [Dati in tempo reale](#) riequilibrare l'asimmetria informativa tra cittadini e istituzioni. A tal fine, pare necessaria l'attività di soggetti che elaborino tali dati - solitamente complessi - al fine di renderli comprensibili ad una fascia più larga di cittadini interessati. Dal momento poi che lo scopo di un riutilizzatore potrebbe essere quello di fornire servizi di aggiornamento in tempo reale, **reperire dati tempestivamente** (quando non - appunto - in tempo reale) diviene un prerequisito indispensabile per il funzionamento del servizio stesso. Ciononostante, le versioni ufficiali di alcuni dei dati generati dal Parlamento non vengono messe a disposizione in tempo reale per questioni burocratico/formali. È il caso, ad esempio, del dettaglio delle votazioni parlamentari, immediatamente disponibile (in HTML), ma convalidato ufficialmente solo dopo il termine della seduta.

3.2 “Open Community Data”

L'**apertura del dato**, che di per sé costituisce il culmine di una serie di attività interne da parte del suo detentore, rappresenta per il potenziale riutilizzatore l'**inizio di un processo** fatto di condivisione e dialogo, e non dovrebbe restare un evento *una tantum* di cui poi disinteressarsi: nel momento in cui soggetti terzi riutilizzano un dato, infatti, si apre una forma di interlocuzione tra chi lo ha messo a disposizione e chi ne fruisce a valle, potenzialmente in grado di determinare **ricadute positive** per la stessa amministrazione pubblica (come nel caso di rilevamento di errori o di raffinazione del dato; il quale, proprio perché messo in condizione di circolare, può ritornare in tale nuova veste nella disponibilità della P.A.). In questo senso, allora, e dal momento che il dato 'perfetto' non esiste, il riutilizzo dei dati costituisce anche e soprattutto uno strumento distribuito per migliorarne la qualità e depurarli da possibili errori.

Nella pratica, alcuni riutilizzatori di dati aperti hanno effettivamente allacciato un **dialogo informale** con le istituzioni parlamentari; questo, se [Condivisione e dialogo: un processo stabile](#)

da un lato agevola evidentemente l'auspicabile coinvolgimento tra i vari attori di open data, può dall'altro prestare il fianco a potenziali critiche da parte di altri riutilizzatori che non avessero a beneficiare dei medesimi canali comunicativi. Pertanto, per aumentare la trasparenza e garantire le medesime opportunità di interlocuzione a tutti i possibili riutilizzatori, pare auspicabile individuare modalità operative che istituiscano **processi di coinvolgimento ampio** tra chi pubblica i dati e chi li riutilizza; negli Stati Uniti, ad esempio, la House of Representatives del Congresso organizza specifici momenti di incontro tra stakeholders e soggetti di volta in volta interessati. Analoghi meccanismi virtuosi di collaborazione sarebbero opportuni al fine di espandere il patrimonio di asset relazionali (banalmente, ad esempio, abilitando uno spazio per la segnalazione di errori a fianco dei dataset scaricabili). Lo sforzo dovrebbe essere quindi volto a favorire processi culturali che innovino le modalità di partecipazione ai processi governativi, ragionando non solo in termini di Open Government Data, ma anche di **Open Community Data**: pensare cioè al pubblico come ad una comunità, promuovendo un processo dialettico tra soggetti pubblici e privati che conduca ad affiancare ai dati pubblici dati provenienti da altre fonti. In Italia stanno ad esempio emergendo forme di iniziativa privata, come [Monithon](#), che divulgano i dati aperti presso scuole e cittadinanza in modo da 'vivificare' il dato, portandolo direttamente all'interno della comunità (ad esempio, per trovare riscontro su come siano state impiegate le risorse investite in un'opera pubblica e che effetto essa abbia prodotto secondo la percezione dei cittadini).

3.3 Le priorità per i riutilizzatori specializzati in dati parlamentari

Desiderata dei riutilizzatori di dati parlamentari.

Nelle preferenze di quanti già riutilizzano e diffondono i dati parlamentari attualmente disponibili, sarebbe auspicabile poi che **si aprissero anche** ulteriori tipologie di dati, in particolare:

- **Dati relativi ai lavori delle Commissioni.** Di una parte rilevante del processo e delle attività delle Commissioni non si riescono a reperire dati (ad esempio sulle presenze e votazioni). La difficoltà maggiore consiste nel fatto che in Commissione non è previsto il voto elettronico: si procede al voto per alzata di mano e viene redatto un verbale rispetto al totale di voti favorevoli e contrari, senza che si registri chi abbia singolarmente votato a favore o contro. All'esterno pervengono soltanto delle ricostruzioni giornalistiche, con inevitabili ricadute in termini di **trasparenza e accountability** complessive. Spesso si giustifica la mancata disponibilità di questo tipo di dati in ragione del fatto che la Commissione rappresenta tipicamente il luogo della mediazione e che quindi ragioni di opportunità suggerirebbero una certa cautela nella diffusione di informazioni; del resto, non si intenderebbe ovviamente pervenire all'estremo opposto, prevedendo forme di documentazione pervicaci come un resoconto stenografico o una diretta streaming, ma piuttosto rendere conoscibile chi siano i presenti e come votino. Questo tipo di informazione non dovrebbe di per sé perturbare il normale andamento delle scelte e del dialogo politico. Anche per questa tipologia sarebbe auspicabile ottenere dati aperti tempestivi (ad esempio **sugli ordini del giorno, sui lavori e sulla programmazione**).
- **Dati sul funzionamento della “macchina” istituzionale.** In particolare, **bilanci e risorse utilizzate** nell'amministrazione: **entrate, spese, rendicontazioni dei gruppi parlamentari**, fino ad arrivare eventualmente alla **dichiarazione patrimoniale dei parlamentari**. Quest'ultimo dato è già disponibile online, ma esiste una barriera rappresentata dal formato in cui esso viene di solito raccolto e prodotto: tipicamente, la scansione in PDF. Dal momento che lo scopo della pubblicazione è quello di stimolare la trasparenza e la conoscenza mediante la circolazione dell'informazione, tale scopo rischia di vanificarsi se i dati diventano sostanzialmente non processabili per via di un formato chiuso (il quale rende oltretutto costoso superare gli ostacoli tecnici inerenti alla sua più difficoltosa fruizione). Attualmente esistono dichiarazioni sullo stato patrimoniale che possono aiutare l'analisi; in questo senso, lo standard potrebbe essere quello previsto per la

dichiarazione dei redditi e per il bollettino patrimoniale.

- **Dati informativi a disposizione di Camera e Senato.** Ad esempio, i dati raccolti dai **Servizi Studi** e dall'**Ufficio Parlamentare di Bilancio**, che attua il controllo sul bilancio dello Stato. Sarebbe utile che i report e le analisi messe a disposizione dei parlamentari fossero accessibili anche ai cittadini; come pure che i dati su cui essi si basano fossero aperti e rielaborabili da soggetti terzi. Il **materiale informativo a supporto delle audizioni** è attualmente messo a disposizione in formato PDF (i dossier di documentazione sono disponibili sui siti della [Camera](#) e del [Senato](#)) e di recente è stata introdotta una modalità partecipativa, rispetto ai lavori in Commissione, nella forma di invio via e-mail di contributi e proposte, da parte dei cittadini i quali, qualora ritenuti meritevoli, sono pubblicati online ed entrano a far parte del materiale informativo. Esiste comunque, rispetto a questa tipologia di materiali informativi, una sorta di "zona grigia" legata al fatto che alcuni di essi vengono realizzati anche mediante contributi esterni; in tali casi, se anche il prodotto del Servizio Studi risulta di per sé un dato pubblico, il grado di apertura effettivo rischia di essere ridimensionato (o quanto meno di risultare incerto, in assenza di un processo di right clearance reso esplicito da parte degli uffici parlamentari) per via dello status giuridico rispetto agli apporti di terzi esterni, sui quali potrebbero insistere alcune restrizioni previste dalla normativa del diritto d'autore.

Si ravvisa inoltre l'utilità di realizzare un **sistema di monitoraggio dell'effettiva attuazione delle norme**. Il Parlamento potrebbe assumere una funzione informativa rispetto a quanto accade dopo che le norme sono state pubblicate nella Gazzetta Ufficiale; l'organo istituzionale demandato alla produzione normativa verrebbe cioè a monitorare anche questo aspetto (esistono peraltro uffici della Camera, del Senato e del Governo che in parte già si occupano di rilevare lo stato di effettiva attuazione delle norme). Si tratta forse di un'attività di tipo più che altro amministrativo, riguardando normative secondarie (ad es., decreti attuativi), ma sarebbe comunque auspicabile un coinvolgimento del parlamento stesso all'interno di un tale sistema.

Tra i soggetti che maggiormente promuovono l'apertura e il riutilizzo dei dati parlamentari, un ruolo di primo piano viene ricoperto dalle c.d.

Dichiarazione sull'apertura parlamentare.

Parliamentary Monitoring Organizations (PMO). In particolare, tra le iniziative concrete di maggior rilievo da esse promosse a livello globale, è doveroso menzionare la [Declaration on Parliamentary Openness](#): lanciata nel corso della World e-Parliament Conference - svoltasi a Roma nel 2012 - ad opera di [OpeningParliament.org](#) (il forum che riunisce organizzazioni specializzate in monitoraggio parlamentare provenienti da tutto il mondo), la **Dichiarazione sull'apertura parlamentare** si rivolge a parlamenti nazionali ed organi legislativi sub-nazionali e transnazionali sollecitandoli ad assumere impegni concreti per l'apertura dei dati in loro possesso e a promuovere forme di partecipazione e confronto con i cittadini.

Si noti infine come l'importanza della disponibilità di dati parlamentari in formato open e machine readable sia stata ribadita anche dal **National Democratic Institute** e dal **World Bank Institute** all'interno di una survey congiunta, [Strengthening Parliamentary Accountability, Citizen Engagement and Access to Information](#), che già nel 2011 documentava l'impiego di [Parliamentary informatics](#) presso il 40% circa delle PMO in tutto il mondo.

3.4 Le indicazioni di ricercatori e "communities"

3.4.1 Interoperabilità semantica

Oggetto di riflessione tra quanti svolgono attività di studio e ricerca sul tema degli open data risultano essere in particolare gli aspetti connessi all'**interoperabilità semantica**, descritta all'interno dello [European Interoperability Framework for European Public Services](#) come la possibilità di elaborare

Interoperabilità semantica e ontologie.

informazioni da fonti esterne o secondarie senza perdere il reale significato delle informazioni stesse nel processo di elaborazione. In effetti, generalmente, gli enti pubblici (e privati) si trovano a gestire basi di dati *in primis* quale naturale conseguenza della rispettiva funzione istituzionalmente svolta ed eventualmente come supporto interno alle proprie attività; per questo è comprensibile che, qualora essi decidano di aprire i propri dati al riutilizzo da parte di terzi, non abbiano sempre una chiara percezione della precipua potenzialità degli open data – ovvero, appunto, che **dati pubblicati da un determinato soggetto vengano incrociati con dati provenienti da altre fonti** – e di conseguenza non operino per agevolarla dal punto di vista tecnico. Tuttavia, l'interoperabilità semantica tra dati di provenienza differente costituisce un ingrediente fondamentale se si intende mettere gli utenti in condizione di fruire appieno dei contenuti (appunto, i dati) pubblicati in modalità aperta sul proprio sito Web: se si ambisce cioè a pubblicare open data di qualità, e considerando che imprese, utenti e cittadini riutilizzatori attingono ad informazioni rispetto a svariati contesti pubblicate su siti Web differenti, i dati in nostro possesso e quelli in possesso di altri soggetti non devono rimanere tecnicamente costretti all'interno di silos separati, incapaci di comunicare tra loro. Al fine di raggiungere questo particolare tipo di interoperabilità, sarebbe pertanto opportuno che all'atto della pubblicazione dei dati si associasse sempre un'**ontologia**, la quale consiste nella [rappresentazione formale e condivisa](#) che descrive i dati attribuendo un **significato univoco** ai concetti da loro espressi ed alle relazioni che caratterizzano il pertinente dominio di conoscenza.

Si noti, peraltro, che l'aspetto dell'interoperabilità semantica è stato oggetto di specifica trattazione all'interno delle linee guida sviluppate in materia dall'AgID (su cui si veda p. 15).

3.4.2 Formati e qualità dei dati

Linked Open Data e coinvolgimento degli utenti.

I rappresentanti di comunità Open Data con i quali ci si è confrontati rimarcano pressoché all'unisono la centralità dell'**impiego di Linked Open Data** al fine di abilitare appieno le potenzialità che i dati aperti sono capaci di generare, pur riconoscendo la complessità intrinseca al modello e la sua non immediata fruibilità da parte di qualsiasi utente (trattandosi pur sempre di un linguaggio "macchina-macchina"). Essi pongono inoltre l'accento sull'importanza di **coinvolgere gli utenti finali** quanto più possibile attraverso mezzi e forme di partecipazione diretta o indiretta, come, ad esempio: strumenti per maneggiare i linked data (quali vocabolari e ontologie descrittive dei dataset); esempi, tutorial, visualizzazioni e altri materiali divulgativi per includere nel circolo virtuoso abilitato dai dati aperti anche quanti non fossero in possesso di cognizioni tecniche approfondite; iniziative, come hackathon e storytelling, per diffondere presso un pubblico ampio lo stato dell'arte dell'apertura dei propri dati; aggregazione di una comunità di utenti vera e propria attorno al proprio modello di linked Open Data (abilitando contesti entro i quali raccogliere gruppi di discussione fornendo stimoli alla comunità che li regge - similmente, ad esempio, a quanto fatto dal consorzio W3C per promuovere l'utilizzo e la conoscenza rispetto agli standard tecnici da esso realizzati per il World Wide Web).

Attorno a questi due filoni principali, le communities promotrici dei dati aperti si sono fatte portavoce, nel corso delle interviste, di alcune **esigenze e opportunità specifiche**. In particolare:

- l'esigenza di **completezza nella tracciabilità del dato**. La non completezza nella tracciabilità del dato sulla produzione legislativa e nella ricostruzione dei vari stadi di produzione di esso può, infatti, rendere opaca tanto la provenienza del dato finale quanto la catena di passaggi che ne ha scandito i diversi stadi di produzione. Utilizzare standard di rappresentazione dei dati, ontologie e semantiche condivise per tutta la produzione dei dati sull'iter legislativo diventa quindi essenziale se si vuole non solo rendere fruibile il dato sul provvedimento legislativo così come promulgato, ma anche abilitare la conoscibilità e comprensione di tutta una serie di aspetti rilevanti circa l'iter legislativo sottostante: ad

esempio, dove e in che modo un dato emendamento sia andato ad incidere sulla norma finale oppure quale parlamentare abbia proposto detto emendamento, in modo da tracciare in concreto gli effetti dell'azione parlamentare sulla produzione normativa. Peraltro, bisogna tenere presente come la catena di tracciabilità dell'attività legislativa sia altamente condizionata dall'andamento dei lavori parlamentari e dalle diverse attività parlamentari. Mentre infatti i percorsi parlamentari per così dire “classici” sono quelli che meglio si prestano ad essere tracciati, in alcuni momenti può capitare che essi vengano marginalizzati dalla contingente situazione politica: si è riscontrato infatti che le proposte di legge ad iniziativa parlamentare rimangono ferme mentre vengono portati avanti disegni di legge promossi dal Governo, che l'attività emendativa nasconde – in alcuni casi – finalità ostruzionistiche e che l'iter si interrompa quando venga posta una questione di fiducia (con l'effetto oltretutto che i maxi-emendamenti prodottisi facciano perdere traccia del legame con l'emendamento originario). Inoltre, in caso di legislazione delegata, le Commissioni parlamentari sono in grado di intervenire solamente esprimendo un rapido parere. Questi aspetti non devono essere trascurati e vanno pure tenuti in considerazione, in quanto forniscono un correttivo laddove ci si accinga a leggere ed interpretare i dati sulla produzione legislativa.

- l'esigenza di **ridurre le difficoltà inerenti all'integrazione di dati espressi in formato eterogeneo**. Un approccio pragmatico consisterebbe nella definizione di una semantica unica, costruita a partire dall'osservazione dei dati stessi o dall'analisi automatica dei loro contenuti (eventualmente per mezzo di tecniche di processamento del linguaggio naturale o statistiche sulla frequenza delle parole). Il fine sarebbe quello della realizzazione di un'ontologia unica e ben documentata del processo legislativo delle due camere, comprendente i termini usati e la loro semantica (ad esempio, cosa significhi “atto” presso la Camera e presso il Senato; se il termine “accorpamento” abbia un significato analogo ad “assorbimento”? etc.); da accompagnarsi alla creazione di URI condivisi per le diverse triple RDF e alla previsione di un ID unico per ogni disegno di legge nei due rami del Parlamento. Si potrebbe inoltre valutare l'adozione dello standard di rappresentazione XML [Akoma Ntoso](#) per tutti i documenti di Camera e Senato. Da ultimo, potrebbe essere utile anche offrire strumenti di visualizzazione che forniscano un modello di verifica a chi intende riutilizzare i dati.
- l'opportunità di **aumentare l'efficacia ed efficienza degli uffici tecnici** del Parlamento nello svolgere le proprie attività. Le tecnologie Linked Open Data rappresentano un'opportunità interna di razionalizzazione delle risorse umane e tecnologiche; ciò può avvenire, ad esempio, offrendo ai Servizi Studi di Camera e Senato, alle segreterie e alle Commissioni strumenti conoscitivi basati su panorama dei dati aperti disponibili, necessari o rilevanti per l'esercizio delle rispettive funzioni. A tale fine si rende però necessario gestire un processo coordinato per la definizione e scelta di vocabolari, ontologie, tesauri condivisi che siano alla base dei modelli espositivi delle entità rappresentate nei dati, tanto a fini conoscitivi connessi con l'attività parlamentare quanto a fini di riutilizzo dei dati da parte dei cittadini. Inoltre, realizzare uno storytelling rispetto ai flussi di uso interno di dati.camera e dati.senato potrebbe aiutare a coglierne il potenziale, esprimerlo all'esterno ed accrescerne la consapevolezza presso i vari livelli dell'amministrazione.
- l'opportunità di **amplificare la conoscibilità delle iniziative open data di Camera e Senato** già intraprese. Un'iniziativa potrebbe essere rappresentata da una "Scuola Open Data" di Camera e Senato che educi gli utenti rispetto a quel che significhi e comporti manipolare dati aperti; oppure si potrebbe accompagnare la componente contenutistica – rappresentata dai dati pubblicati da Camera e Senato – con una componente formativa che a cadenza quotidiana o periodica provveda a fornire agli utenti informazioni e spunti su come si possano riutilizzare i dati.

3.4.3 Apertura di ulteriori tipologie di dati

Dati sul bilancio del Parlamento.

A completamento e supporto di quanto espresso dai riutilizzatori di dati parlamentari in merito all'opportunità di mettere a disposizione ulteriori dati sul funzionamento della “macchina istituzionale”, è stato ricordato come alcuni tra i primi dataset in formato Linked Open Data ad apparire in Italia, già nel 2011, furono quelli pubblicati nel contesto dell'iniziativa [Linked Open Camera](#), promossa dall'Associazione Linked Open Data Italia e realizzata partendo dalla base dati relativa al bilancio della Camera messa a disposizione dal Partito Radicale. Pur riferendosi ad un contesto che ha già sperimentato la spinta di iniziative dal basso, sarebbe non di meno auspicabile che a curare la pubblicazione di questo tipo di dati provvedessero direttamente Camera e Senato, eventualmente nella forma di un **bilancio Linked Open Data unitario del Parlamento**, esponendo i rispettivi dati di budget delle rispettive camere.

3.5 La domanda di dati (aperti) da parte del Parlamento

La domanda di dati proveniente dai parlamentari e dai Servizi parlamentari di documentazione.

Nell'ambito della analisi della domanda di dati abbiamo voluto prendere specificamente in considerazione la **domanda proveniente dall'interno stesso del Parlamento**, e cioè dai parlamentari e dagli uffici coinvolti nella attività legislativa. L'attività legislativa prevede infatti, per essere accurata ed efficace, l'accesso e l'elaborazione di una grande quantità di informazioni della più varia natura. Di alcune di queste informazioni abbiamo già parlato nel capitolo dedicato all'offerta, in particolare quando abbiamo dedicato una sezione alla accessibilità e riutilizzabilità degli atti legislativi e del corpo normativo. Ma molte altre non sono attualmente accessibili mediante i sistemi informativi di Camera e Senato.

Si pensi ad esempio ai dati statistici sui più diversi fenomeni economici e sociali. O ai dati finanziari della contabilità nazionale e locale. O ancora ai dati di ricerche effettuate su specifici fenomeni ambientali e imprenditoriali. O, più in generale, a dati pubblici non ancora resi disponibili dalle amministrazioni che li detengono.

Il fatto che tali dati siano oggi solo in parte direttamente accessibili da parte dei parlamentari mediante i sistemi informatici di Camera e Senato non significa però che essi non siano utilizzabili nel corso della attività legislativa. Alla loro raccolta ed elaborazione sono infatti predisposti alcuni uffici parlamentari di lunga tradizione e riconosciuta reputazione: si tratta dei Servizi parlamentari di documentazione, in particolare dei Servizi Studi di Camera e Senato, che producono rapporti sui più diversi temi oggetto della attività legislativa.

Questi Servizi di fatto svolgono la funzione di trasformare la conoscenza esterna al Parlamento, nei più di diversi settori tematici, in informazione direttamente fruibile dai parlamentari e dagli uffici che svolgono attività legislativa. Sempre di più questi Servizi si confrontano con l'opportunità di accedere ed elaborare dati digitalizzati detenuti dal settore pubblico e/o privato e rilevano la necessità di poter utilizzare come open data dati ancora non ancora disponibili, o, in alcuni casi, ancora non digitalizzati.

Le interviste effettuate con i Servizi Studi di Camera e Senato consentono quindi di tracciare una prima mappa della domanda di dati aperti da parte del Parlamento con una duplice finalità: prevedere ove possibile una diretta accessibilità tramite i sistemi informativi dell'amministrazione e ampliare, anche per questa via, la disponibilità di open data pubblici forniti ai cittadini dall'istituzione parlamentare.

Si tratterebbe in questo caso non di dati relativi in senso stretto alla attività parlamentare, ma del più largo insieme dei dati utilizzati dai parlamentari nel corso delle attività di discussione e produzione di leggi.

Nei paragrafi successivi si descrivono i principali contenuti emersi proprio nel corso delle

interviste effettuate con i Servizi Studi di Camera e Senato (laddove le indicazioni ricevute siano state analoghe, si farà riferimento genericamente ai Servizi di documentazione parlamentari).

3.5.1 Le attività dei Servizi di documentazione parlamentari

I Servizi di documentazione parlamentari curano l'assistenza tecnico-documentaria agli organi parlamentari finalizzata a fornire la completa **base conoscitiva** sui temi intorno a cui si svolge l'attività parlamentare. Inoltre, su richiesta e su temi specifici, svolgono analoghe attività per i gruppi parlamentari ed i singoli membri del Parlamento. I Servizi di documentazione parlamentari non forniscono pareri, bensì **supporto documentale**, curando l'eshaustività e l'accuratezza del quadro informativo e delle fonti impiegate, senza tralasciare quindi alcuna fonte pertinente e - se necessario - operando comparazioni tra diverse fonti, in modo che il ragionamento e l'esito risultino quanto più evidenti, logici, obiettivi e neutrali possibile.

Per la Camera, le attività di cui sopra sono svolte dal **Servizio Studi della Camera dei Deputati**, assieme agli altri **Servizi di documentazione**.

I Servizi di documentazione della Camera dei Deputati.

Tra le strutture specializzate impegnate anch'esse nella cosiddetta fase ascendente del procedimento legislativo: il **Servizio del Bilancio**, che cura le verifiche degli impatti quantitativi e di copertura finanziaria degli interventi legislativi sul Bilancio dello Stato; l'**Ufficio Rapporti con l'Unione Europea**, che produce documentazione sulle politiche da adottarsi nel contesto europeo; alcuni settori della **Biblioteca**, che effettuano analisi comparate della legislazione straniera.

Analogamente a quanto descritto sopra, il **Servizio Studi del Senato della Repubblica** fornisce supporto documentale di livello generale al Senato: è **produttore di informazione** soprattutto per quanto riguarda l'analisi tecnico-giuridica, **ma è anche ricercatore-intermediatore di informazione non propria**. In aggiunta ad esso, svolgono funzioni affini il **Servizio Bilancio**; il **Servizio degli affari europei ed internazionali** e il **Servizio per la qualità degli atti normativi**.

I Servizi di documentazione del Senato della Repubblica.

La natura e le finalità delle attività dei Servizi di documentazione parlamentari sono tali da richiedere che il **patrimonio di conoscenza** su cui essi operano analisi e ricerche – per fornire il supporto agli organi parlamentari – sia **ampio e articolato**, poiché l'attività legislativa può focalizzarsi su qualsiasi ambito o tema, ma anche **costantemente aggiornato e immediatamente disponibile**, per rispondere in modo professionale e tempestivo alle esigenze informative di commissioni, organi e deputati.

Natura e finalità dei Servizi di documentazione parlamentari.

Dalle interviste con i Servizi della Camera dei Deputati, è emerso come questi ritengano particolarmente qualificante una attività quotidiana di **monitoraggio e classificazione delle informazioni** e dei documenti di molteplici fonti, a prescindere da specifiche richieste e necessità contingenti, al fine di predisporre internamente in modo progressivo e preventivo la base di conoscenza ampia e aggiornata a cui attingere in base alle specifiche esigenze e nei tempi compatibili con le scadenze dettate dai lavori parlamentari.

Rispetto alle tipologie di fonti monitorate e di analisi effettuate, in particolare dalle interviste con i Servizi del Senato della Repubblica, è emerso che l'attività principale consta dell'**analisi di testi giuridici**: ad esempio, il Servizio Studi del Senato stima che questo tipo di attività rappresenti il 50% del volume totale di attività svolte.

La "Piattaforma Aperta per la Documentazione"

Per la gestione della base di conoscenza necessaria, gli uffici della Camera dei Deputati menzionati in questa sezione utilizzano una infrastruttura digitale recentemente realizzata, denominata **PAD**, acronimo di **Piattaforma Aperta per la Documentazione**.

Entro questa infrastruttura sono stati digitalizzati in un unico **repository RDF** tutti i contenuti degli archivi e sono state codificate in **SKOS** le griglie di classificazione degli uffici – che costituiscono il thesaurus Theca – e il thesaurus **Eurovoc**, consentendo la gestione delle relazioni tra voci di thesaurus diversi.

PAD si sviluppa in tre aree funzionali: una relativa agli archivi documentali, una per la gestione del semantic enrichment e la classificazione dei contenuti, ed una per la produzione partecipata della documentazione.

Le attività quotidiane di **spoglio delle fonti informative**, che alimentano gli archivi documentali di PAD, vengono effettuate in minima parte con le attività di alimentazione manuale da parte degli operatori, in massima parte attraverso l'automatico scarico in RDF nel repository Linked Open Data di informazioni e documenti interni alla Camera dei deputati e attraverso operazioni semiautomatiche che con servizi di *scraping* acquisiscono contenuti da fonti esterne selezionate, quali ad esempio la [Gazzetta Ufficiale](#), la [Corte Europea di Giustizia](#), l'[Official Journal of European Union](#), lasciando a cura del personale interno le eventuali integrazioni manuali, le annotazioni e le attività di selezione e classificazione semantica, ripartite per settori tematici di competenza.

Le attività redazionali per la produzione dei **vari prodotti editoriali** – dossier, note, newsletter, ricerche, collane specifiche – vengono effettuate con le funzioni del sottosistema di produzione dei documenti, che offre servizi evoluti di Web editing integrati con strumenti di word processing, supporto alla redazione partecipata e collaborativa di documenti, strumenti di mash-up di contenuti del repository per includerle nei prodotti documentali, funzioni di pubblicazione e distribuzione dei documenti digitali prodotti in HTML, PDF, RDF, eBook.

Tutti i contenuti del repository, siano quelli degli archivi, siano i singoli elementi informativi dei prodotti documentali, possono esser marcati con tag e con le voci dei diversi tesauri codificati in SKOS e gestiti dal sottosistema di gestione semantica, in modo da arricchire il valore semantico attribuito ad ogni singola informazione e ampliare le capacità di ricerca e accesso nella base dati.

3.5.2 L'approntamento della documentazione al servizio delle attività parlamentari

Le attività di produzione documentale dei Servizi della Camera

Le attività di produzione documentale svolte dal Servizio Studi sono strettamente legate all'andamento dell'attività parlamentare, come disposto dallo stesso [Regolamento della Camera dei deputati, all'art. 79, comma 4](#), che assegna ad esso un ruolo fondamentale nell'**istruttoria legislativa**:

ogni punto all'ordine del giorno relativo all'esame di una proposta di legge necessita di un dossier di documentazione. Pertanto il Servizio Studi deve essere pronto a svolgere questo compito istruttorio con competenza nelle specifiche materie e tempestività rispetto alla programmazione dei lavori parlamentari. L'attività della Camera si dipana attraverso 14 Commissioni permanenti, ciascuna competente su specifiche materie, alle quali corrispondono 14 Dipartimenti del Servizio Studi.

Per ogni provvedimento è previsto l'obbligo di mettere a disposizione un dossier alla Commissione e a ogni parte politica coinvolta, affinché si possa avviare l'esame del testo della proposta di legge. I dossier sono inoltre messi a disposizione dei cittadini sul [sito Web della Camera](#) assieme alle informazioni dell'iter dei progetti di legge. Inoltre il Servizio Studi mette a disposizione dei deputati e dei cittadini parte della sua conoscenza sugli argomenti su cui si focalizza l'attività parlamentare, curando i contenuti di una sezione del sito [www.camera.it](#) accessibile dall'homepage, destinata a illustrare **i temi dell'attività parlamentare**, connessi con i provvedimenti all'ordine del giorno, suddivisi per aree tematiche e commentati con schede di documentazione articolate a diversi livelli di dettaglio e corredate di link a fonti e contenuti correlati.

Il Servizio Studi opera anche **su richiesta di singoli deputati o organi** che possono essere interessati a ricerche su specifici argomenti non necessariamente legati alla attività parlamentare all'ordine del giorno. La documentazione prodotta per rispondere alle esigenze conoscitive di questo tipo non viene pubblicata ma viene fornita unicamente al soggetto richiedente. Solo nel caso in cui l'oggetto della ricerca acquisisca un rilievo nell'ambito del lavoro di un organo parlamentare ne viene disposta la pubblicazione sul sito.

Il Servizio Studi produce materiale documentale **d'ufficio e su domanda**. Nel primo caso si tratta

di attività che soddisfa le necessità informative rispetto all'ordine del giorno delle assemblee e delle commissioni, a servizio dei senatori, di organi e gruppi. Questo tipo di attività è pianificata in relazione alle attività del Parlamento e del Governo e, in questo senso, l'articolazione del Servizio Studi per uffici riproduce quella delle Commissioni, che a sua volta è riconducibile a quella dei Ministeri. Nel secondo caso, si tratta di attività svolta in seguito a specifiche richieste di senatori o di organi del Senato: tipicamente, tali domande richiedono un livello di elaborazione minore, non passano per gli uffici se non per una forma di controllo conclusivo, e sono trattate da un pool selezionato all'interno del Servizio Studi.

Le attività di produzione documentale dei Servizi del Senato

3.5.2.1 Le fonti dei Servizi di documentazione parlamentari

Indipendentemente dalla origine dell'esigenza di documentazione, la produzione dei Servizi di documentazione muove da una **base di conoscenza** di fonti documentali e informative precostituita e organizzata attraverso la quotidiana attività di ricerca, monitoraggio informativo, spoglio e classificazione, che consente di rispondere tempestivamente e con il livello di approfondimento necessario alle esigenze informative. Ad esempio, il Servizio Studi della Camera dei Deputati si basa in primis su **fonti di tipo legislativo/normativo**: in particolare, tutte le fonti generali come la [Gazzetta Ufficiale](#), la [Gazzetta Ufficiale dell'Unione Europea](#), i Bollettini Regionali, Decreti Ministeriali, codici, circolari, e banche dati (ad es., [InterLex](#)). Vengono consultate correntemente anche fonti di dottrina di giurisprudenza, oltre a fonti Web che forniscono i testi vigenti di normativa e ovviamente al sito [Normattiva](#), ai cui contenuti la Camera dei deputati contribuisce direttamente nell'ambito della funzione legislativa.

Per il supporto istruttorio alle attività di Commissioni non prettamente tecnico-giuridiche vengono consultate anche **fonti istituzionali di altro tipo**: ad esempio, in caso di finanza pubblica vengono reperite informazioni presso la [Ragioneria Generale dello Stato](#), [Istat](#), [Agenzia delle Entrate](#), [Corte dei Conti](#). A livello regionale, il Servizio Studi si avvale della cooperazione tra l'Osservatorio sulla legislazione della Camera dei deputati e le amministrazioni delle Assemblee regionali, nonché del supporto dell'[ISSIRFA](#) (Istituto di Studi sui Sistemi Regionali Federali e sulle Autonomie).

Sebbene i Servizi di documentazione parlamentari siano in possesso di strumenti privilegiati, quali abbonamenti a banche dati ed accesso a materiale ad accesso ristretto (ad es. quello del prodotto dal CERDP - il centro di ricerca per la ricerca e documentazione parlamentare), il lavoro dei Servizi si basa anche e in parte non trascurabile su **dati accessibili a tutti i cittadini**.

Anche i dati aperti tra le fonti utilizzate.

Nel reperire le informazioni, si privilegiano naturalmente le **fonti Web istituzionali**, la cui ufficialità solleva da un controllo sull'affidabilità della fonte (esempi tipici di siti Web largamente utilizzati sono quelli, già citati, dell'Istat o della della Ragioneria Generale dello Stato, o ancora le statistiche del Ministero della Giustizia e il sito Webstat giustizia).

All'occorrenza, vengono consultate anche altre **fonti non istituzionali reperibili su Internet** (inclusi Wikipedia o blog di vario genere), che i Servizi non considerano di per sé meno importanti: ai fini della attività d'indagine non si ritiene opportuno escludere a priori alcuna fonte informativa in ragione del diverso peso ed autorevolezza rispetto ad altre fonti, essendo in linea di principio tutte potenzialmente utili. Ad esempio, il settore relativo alla politica internazionale presenta una estrema varietà e diversità di fonti che possono dover essere prese in considerazione per costruire un quadro informativo esaustivo e neutrale: in taluni casi, ad esempio, vengono consultate anche fonti tipicamente informali, quali i blog di attivisti. Rispetto a questo tipo di fonti ovviamente i Servizi mantengono un approccio estremamente cauto, al fine di garantire e preservare la necessaria autorevolezza e neutralità delle fonti che costituiscono la base informativa della documentazione prodotta. Tipicamente, queste fonti rappresentano un'informazione per arrivare all'informazione: servono cioè per ottenere un primo livello di informazioni, soprattutto nel caso in cui ci si trovi di fronte a contesti specialistici, in modo da vagliare e rintracciare più rapidamente

fonti pertinenti e autorevoli.

3.5.2.2 L'utilizzo di dati di altri enti

L'esperienza della Camera.

In alcuni casi, il Servizio Studi della Camera ha riportato una certa **difficoltà ad accedere ai dati di altri enti**: ad esempio ai dati dell'[INPS](#), di alcune agenzie fiscali, della parte non pubblica delle banche dati della [Protezione Civile](#). Oltretutto, a fronte della richiesta di dati viene sovente opposto il dettato della determinazione commissariale INPS n. 60 del 11 marzo 2010 che stabilisce che la richiesta di informazioni ad enti pubblici vada inoltrata in forma scritta ed accolta dietro il pagamento di una tariffa (legata ai costi di gestione di recupero del dato, anche se in formato digitale). Proprio in conseguenza di questi episodi, il Comitato per la vigilanza sulla biblioteca e la documentazione - organo interno dell'Ufficio di Presidenza della Camera dei deputati - ha richiesto al Servizio Studi di produrre un elenco di enti cui esso non riesce ad accedere. Si noti del resto come, rispetto al comparto della finanza pubblica, la [Legge di contabilità 31 dicembre 2009, n. 196](#) preveda che le relative banche dati di amministrazioni pubbliche e ogni altra fonte informativa rilevante gestita da soggetti pubblici debbano essere accessibili alla Camera e al Senato (vedi il Quadro di sintesi e l'art. 6); in questo senso, il Servizio trova infatti accesso ai dati sulla disponibilità dei capitoli di spesa in possesso della Ragioneria Generale.

Il Servizio Studi reperisce ed impiega l'informazione senza procurarsi un'autorizzazione per tali fini: il rapporto con altri enti viene considerato di natura istituzionale e pertanto **non viene effettuata una richiesta motivata e formale di accesso ai dati**. Rispetto allo **status giuridico dei dati** e alla right clearance su di essi, ritiene che una volta che l'ente o autorità abbia fornito loro i dati, essi siano pubblici e pertanto pubblicabili dal Servizio Studi secondo le modalità da questo scelte (e quindi anche come dati aperti). Oltretutto, una volta che i dati siano confluiti all'interno di un dossier o altra documentazione e questa sia stata presentata alla Commissione competente, essi diventano pubblici nel momento in cui sono adottati dalla Commissione stessa con delibera.

L'esperienza del Senato.

Il Servizio Studi del Senato, rispetto al reperimento delle informazioni, rileva una **mancanza di corrispondenza tra dati di fonti diverse**, mentre altre volte incorre nelle difficoltà connesse al passaggio **da un dato "sporco" ad uno "pulito"** come in caso di informazioni utili ma che vanno "verificate" (ad es., perché reperite su siti di cui non è certa l'affidabilità) o in caso di informazioni formalmente non accessibili ma che tuttavia sono state fatte circolare all'esterno (come nel caso di una circolare di ragioneria riservata che venga pubblicata online, non esperibile da parte del Servizio all'interno della propria documentazione ufficiale). In quest'ultimo caso, proprio la chiusura del dato a monte rischia di creare potenzialmente un'impasse rispetto alla funzione svolta dal Servizio.

Altre **difficoltà specifiche** sono state riscontrate rispetto ai **dati istituzionali regionali** per via della frammentazione dei relativi siti Web istituzionali, causata dalla circostanza che ogni regione possiede una sua banca dati. Non si tratta di problemi di opacità vera e propria, ma piuttosto difficoltà inerenti alla frammentarietà del contesto.

Nel reperire l'informazione necessaria alle proprie funzioni, il Servizio non procede richiedendo un'autorizzazione per il suo impiego: più semplicemente, **se l'informazione è pubblica viene utilizzata, in caso contrario non viene impiegata**. Essa viene reperita per via parlamentare oppure amministrativa: per via parlamentare si può ricorrere alle informazioni desumibili dallo strumento delle interpellanze e interrogazioni; in via amministrativa gli Uffici reperiscono l'informazione laddove essa sia pubblica.

Per via del rapporto impari tra capacità d'organico ed ambiti tematici trattati, nella maggior parte dei casi il Servizio non è in grado di effettuare attività specifiche di **incrocio ed elaborazione di dati**. È ragionevole quindi ritenere che un aumento della disponibilità di informazioni pubbliche ed

una loro strutturazione semantico-ontologica adeguata agevolerebbero non solo la fruizione e la conoscenza da parte delle communities di sviluppatori-riutilizzatori e dei cittadini (laddove – ovviamente – i dati fossero aperti al pubblico), ma in primo luogo il funzionamento del Servizio Studi stesso e, verosimilmente, di altri uffici tecnici interni al Parlamento.

Cap. 4. A way forward

Questa sezione conclusiva mira, da un lato, a evidenziare i principali spunti emersi nei precedenti capitoli e, dall'altro, a mettere in evidenza azioni e politiche che possano favorire lo sviluppo delle competenze e delle buone pratiche relative ai dati aperti in Italia – con particolare riferimento a quelli di origine parlamentare – nella prospettiva di massimizzare le ricadute positive per l'economia digitale e la società nel suo complesso.

L'open data si rivela una eccellente occasione per rendere la **gestione dei sistemi informativi più efficiente, trasparente e utile a terzi**. Più efficiente poiché facilita l'interscambio di dati tra enti o uffici diversi. Più trasparente poiché permette a chiunque lo desideri di fruire di dati a loro volta elaborabili, riducendo l'asimmetria informativa tra cittadini e pubblica amministrazione. Più utile a terzi poiché sono sempre più numerosi i soggetti che fanno dei flussi di dati una delle componenti chiave dei servizi da essi erogati, e la disponibilità crescente di dati pubblici aperti non può che rappresentare uno stimolo ulteriore.

Open Data per l'efficienza dei sistemi informativi.

Tuttavia, tali opportunità possono essere colte appieno solo in presenza di un complesso insieme di fattori. Occorre in primo luogo che **l'apertura dei dati sia un processo incorporato ab origine nel ciclo di vita del dato pubblico**, quindi anche nelle fasi di acquisizione e trasformazione dei dati, le quali vanno concepite in modo tale da abilitare – fatti salvi i limiti di legge – una tempestiva messa a disposizione dei dati al pubblico. Non solo: la stessa agenda di pubblicazione di dati aperti deve tenere conto della domanda, anche implicita, di potenziali riutilizzatori, eventualmente mediata da soggetti che svolgano, a valle, il ruolo di “broker” di informazione. Inoltre, è necessario che l'accuratezza intrinseca dei dati, la loro interpretabilità e fruibilità vengano mantenute su elevati standard di qualità, in modo tale da superare, internamente, eventuali incongruità nella rappresentazione informativa, e così da facilitare, verso l'esterno, il riuso dei dati.

Open Data by design (and not by accident).

Alla luce delle precedenti considerazioni, una buona strategia open data determina, non marginalmente, un **impatto positivo sull'efficienza interna delle organizzazioni**. Per converso, una delle caratteristiche di un sistema informativo ben concepito è ad oggi quella di considerare soggetti esterni come utenti dello stesso, seppure dotati di privilegi di accesso puntuale meno avanzati rispetto a quelli interni.

Il modello di produzione e rilascio di open data da parte degli organi parlamentari italiani va espressamente nella direzione qui descritta. Ciò si può affermare per tutti gli ambiti informativi presi in considerazione, sia quelli attinenti alla struttura dell'istituzione e alle sue attività, sia quelli concernenti i Servizi strumentali, gestionali e di documentazione.

Il presente lavoro ha permesso per altro di meglio comprendere come l'attività di raccolta, elaborazione e diffusione di informazione da parte degli organi parlamentari italiani abbia un'origine ben più radicata nel tempo rispetto alla relativamente recente – per quanto già foriera di risultati d'eccellenza – evoluzione verso i linked open data. Ciò è probabilmente vero per la maggior parte delle tipologie di istituzioni che hanno, in questi ultimi anni, intrapreso la strada degli open data. Tuttavia, questo aspetto assume un particolare rilievo nel caso di organismi mediante i quali si realizza il mandato democratico, con la relativa attività parlamentare e produzione legislativa, la quale a sua volta implica, in entrata e in uscita, flussi di informazione di rilevanza pubblica, con la connessa esigenza di trasparenza e interpretabilità da parte dei cittadini, delle imprese e delle istituzioni.

Inoltre, il presente lavoro consente di saggiare come **il contributo alla produzione di linked open data da parte del Parlamento italiano sia ad oggi considerevole** in termini di volume di informazione (oltre 500 milioni di triple pubblicate dalla Camera dei deputati e oltre 30

Una mole considerevole di Linked Data, ed un caso d'uso concreto tra i più rilevanti in Italia.

milioni di triple pubblicate dal Senato della Repubblica) e di copertura tematica, collocando le istituzioni italiane in una posizione ragguardevole anche nel contesto internazionale.

Tale aspetto assume un valore ulteriore se si considera la tecnologia utilizzata, propria del cosiddetto Web dei dati, la quale consente un

arricchimento informativo distribuito e un facile collegamento tra dati esposti e documentati da soggetti diversi, con il raggiungimento della cosiddetta “quinta stella” sulle cinque della scala di riferimento coniata da Tim Berners-Lee. Questo percorso viene d'altra parte abilitato al massimo livello dalla **scelta effettuata dagli organismi parlamentari italiani di adottare ontologie formali** le quali importano ontologie esistenti e ampiamente documentate. Si tratta, inoltre, di ontologie estensibili, il che facilita l'accrescimento e la diversificazione dell'offerta di dati, com'è negli obiettivi delle istituzioni coinvolte.

L'apertura dei dati all'esterno è spesso utile alle istituzioni stesse.

Tali scelte in termini di rappresentazione informativa hanno avuto – e si prevede avranno ulteriormente – un **impatto positivo sull'organizzazione stessa delle istituzioni**, in particolare per quanto concerne l'interscambio di informazione e l'individuazione di modelli

comuni, ma anche in conseguenza del disaccoppiamento tra i dati e gli applicativi software che li generano, a tutto vantaggio delle stesse organizzazioni, e dei fruitori dei dati, siano essi personale interno, soggetti in carica elettiva, cittadini, imprese, associazioni, e riutilizzatori in genere.

Naturalmente, ogni decisione in merito alle modalità di pubblicazione di dati deve far fronte a un “trade-off”. L'esposizione di dati grezzi consente di produrre un'immagine oggettiva, ma non sempre interpretabile in assenza di competenze e strumenti presso i soggetti interessati a elaborare i dati. D'altro canto, la pubblicazione di elaborazioni più immediatamente comprensibili comporta, come svantaggio, il rischio di influenzare l'interpretazione a monte. Questo concetto vale per i dati di ambito parlamentare – si pensi, ad esempio, a quelli relativi all'attività di deputati e senatori – ma si applica in numerosi altri domini informativi oggetto di pubblicazione di open data. Ciò considerato, continua a rimanere auspicabile che gli organismi pubblici mettano a disposizione **evidenze in forma il più possibile oggettiva e neutrale**. Tuttavia, specie in ambiti informativi particolarmente complessi (non solo attinenti all'ambito parlamentare: si pensi, ad esempio, ai dati di bilancio delle pubbliche amministrazioni), è altrettanto necessario che venga fornita **documentazione esaustiva e accessibile ai fini di abilitare elaborazioni consapevoli da parte di terzi**. Non a caso, nel caso in esame, è possibile navigare liberamente il grafo di linked data mediante query SPARQL, ma vengono anche fornite e documentate al pubblico interrogazioni di esempio, utili a guidare l'utente a formulare le proprie. Inoltre, nella medesima ottica, pare fondamentale il ruolo di intermediari indipendenti, posti a valle del settore pubblico, in grado di aggregare più fonti informative nel loro portafoglio di offerta di servizi agli utenti, fungendo da ulteriore porta d'accesso ai dati.

Alla luce di quanto presentato, si individuano **tre linee di attività** (tra di loro collegate) finalizzate a conferire ulteriore impulso alla già soddisfacente attività di rilascio di dati aperti, collegati ed elaborabili da parte degli organismi parlamentari italiani: 1) il prosieguo nella collaborazione inter-istituzionale per ciò che concerne la produzione del dato parlamentare; 2) un ulteriore coinvolgimento di soggetti esterni alle istituzioni; 3) l'ampliamento degli ambiti di informazione rilasciata, anche a partire dalle richieste di soggetti riutilizzatori. Queste tre linee di attività vengono di seguito sinteticamente descritte.

1) **Rafforzare la collaborazione inter-istituzionale nei processi di produzione del dato parlamentare.**

Ciò al fine di ridurre la pur fisiologica frammentazione tra uffici e organismi differenti, con l'effetto di migliorare l'efficienza ed efficacia delle proprie ed altrui attività, anche a livello inter-istituzionale. Il proseguimento del **dialogo tra i due rami del Parlamento** è di particolare importanza in tal

sensu. Nello specifico, è auspicabile una sempre più stretta interazione tra i Servizi che gestiscono l'infrastruttura tecnologica e i Servizi Studi da un lato, e altri organismi che producono documentazione a uso dei parlamentari e di altri soggetti dall'altro. Nell'ambito della strategia open data, tale dialogo può favorire scelte condivise di tipo tecnico e giuridico, anche con l'obiettivo di un ulteriore allineamento semantico tra le ontologie utilizzate. Nella medesima ottica, le istituzioni coinvolte potrebbero considerare di individuare un soggetto stabilmente responsabile della gestione e promozione di tutto ciò che comporti "fare open data" all'interno di un organo parlamentare, e che coordini un gruppo di lavoro dedicato.

Gli standard internazionali agevolano anche l'interoperabilità tra i rami del Parlamento.

2) **Abilitare un circolo virtuoso tra gli organismi parlamentari in quanto produttori del dato e soggetti esterni in quanto riutilizzatori del dato.**

Alcune possibili azioni concrete in questo senso potrebbero comprendere:

Supportare l'intero ciclo di vita del dato aperto, compreso il riutilizzo.

- l'ulteriore messa a disposizione di manualistica, tutorial, infografiche e simili materiali informativi che illustrino l'importanza del principio dell'Open Government e il ruolo precipuo – ad esso funzionale – ricoperto dai dati aperti, anche attingendo al vasto insieme di riferimenti già esistenti;
- la riproposizione di spazi ad hoc per migliorare l'interazione con i riutilizzatori e la fruizione di questi:
 - in forma consultiva, ad esempio mediante sezioni dei portali Web istituzionali adibite a ricevere commenti e suggerimenti, anche in materia di agenda di pubblicazione dei dati; e, più in generale, nell'ambito di incontri aperti al pubblico;
 - nel contesto di attività di sviluppo collaborativo, ad esempio di hackathon (sulla stessa linea di "Code4Italy", tenutosi a Montecitorio nel 2014), anche a partire dai nuovi dati messi a disposizione e dalle applicazioni nel frattempo sviluppate da soggetti terzi.

3) **Ampliare gli ambiti di informazione rilasciata, anche a partire dalle richieste di soggetti riutilizzatori.**

Pare utile condividere l'evoluzione della strategia di pubblicazione di open data parlamentari con i soggetti interessati al riutilizzo dei dati, a partire dalle *community* di associazioni, imprese, sviluppatori e ricercatori già attivi, incluse quelle interpellate nel corso della stesura del presente lavoro. In particolare, i desiderata espressi dalle stesse – dall'ampliamento delle basi informative da condividere, ad esempio quelle concernenti le risorse utilizzate dalle amministrazioni, sino all'ulteriore miglioramento dell'interoperabilità semantica – possono essere considerati come punti di partenza, del resto già coerenti con il piano delle istituzioni parlamentari italiane in materia di open data.

La domanda dei riutilizzatori può guidare l'apertura dei dati, o addirittura la raccolta di nuovi dati strutturati.

Si consideri peraltro che il presente lavoro fornisce spunti di interesse anche per ciò che concerne forme di riutilizzo inter-istituzionale dei dati parlamentari. Così, ad esempio, la pubblicazione in formato aperto ed elaborabile automaticamente degli atti parlamentari apre la strada a numerose applicazioni (si pensi, tra le tante, all'analisi automatica dei testi dei disegni di legge per il calcolo di statistiche sulle frequenze delle parole ovvero dei temi trattati). Tale prospettiva inoltre – se condivisa a livello inter-istituzionale - potrebbe portare all'alimentazione automatica delle banche dati normative e dei lavori preparatori quale ad esempio il portale della normativa vigente Normattiva, contribuendo anche – dal punto di vista del supporto tecnico - alle attività di semplificazione normativa.

Annex

Glossario

Concetto	Definizione
Banche di dati	Si intendono raccolte di opere, dati o altri elementi indipendenti sistematicamente o metodicamente disposti ed individualmente accessibili mediante mezzi elettronici o in altro modo (art. 2, n. 9, L. 633/1941).
Costo marginale	Si intende la variazione nei costi totali di messa a disposizione del dato che si verifica quando viene riprodotta e distribuita ("scaricata") un'unità in più di prodotto. In particolare, considerando l'economia basata sui dati, i costi marginali comprendono i costi complessivi di raccolta, produzione, riproduzione e diffusione di documenti. L'attività di produzione comprende la creazione e l'assemblamento, e la diffusione può comprendere anche l'assistenza agli utilizzatori (fonte: Considerando 14, Direttiva 2003/98/CE).
CKAN	Il Comprehensive Knowledge Archive Network (CKAN) è una sistema open source e basato sul Web per l'immagazzinamento, la catalogazione e la distribuzione di dati, quali ad esempio fogli di calcolo o contenuti di database.
Dataset	Una collezione di dati, generalmente riguardanti una stessa organizzazione, che vengono erogati e gestiti congiuntamente; insieme di dati strutturati in forma relazionale.
Dato	Un dato è una rappresentazione reinterpretabile delle informazioni in un formato convenzionale idoneo alla comunicazione, all'interpretazione o all'elaborazione (norma ISO/IEC 2382-1). I dati possono essere creati da persone o generati da macchine/sensori, spesso sotto forma di "sottoprodotto". Alcuni esempi sono i dati geospaziali, statistici, meteorologici, della ricerca, ecc.
Dato anonimo	Il dato che in origine, o a seguito di trattamento, non può essere associato ad un interessato identificato o identificabile (art. 4, c. 1, lett. n), Codice Privacy).
Dato personale	Qualunque informazione relativa a persona fisica, identificata o identificabile, anche indirettamente, mediante riferimento a qualsiasi altra informazione, ivi compreso un numero di identificazione personale (art. 4, c. 1, l. b, Codice Privacy). Si tratta di una definizione molto ampia che necessita di essere contestualizzata per lo specifico caso concreto di applicazione. Un particolare insieme di dati personali è costituito dai dati identificativi, cioè i dati personali che permettono l'identificazione diretta dell'interessato (art. 4, c. 1, lett. b), Codice Privacy).
Dato pubblico	Si tratta del dato conoscibile da chiunque (art. 1, c. 1, lett. n), CAD e art. 2, c. 1, lett. d), D. Lgs. 36/2006).

Concetto	Definizione
Dati aperti/Open Data	La "definizione standard" oggi generalmente riconosciuta, che si è largamente affermata presso le communities promotrici del modello Open Data, è quella contenuta nell' Open Definition , aggiornata a Settembre 2014, secondo la quale sono aperti quei dati "a cui è liberamente possibile accedere, e che possono essere liberamente utilizzati, modificati e ridistribuiti da chiunque, per qualunque scopo", rispettando, al più, requisiti che preservino l'origine del dato e la sua apertura. La precedente versione della Open Definition faceva riferimento, al più, al requisito di attribuzione e condivisione allo stesso modo". Da una prospettiva giuridica, sono quei dati che presentano le seguenti caratteristiche: 1) sono disponibili secondo i termini di una licenza che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato; 2) sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera a) (art. 68, c. 3, CAD), sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi metadati; 3) sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione (art. 68, c. 3, lett. b), CAD).
Diffusione di dati personali	Il dare conoscenza dei dati personali a soggetti indeterminati, in qualunque forma, anche mediante la loro messa a disposizione o consultazione (art. 4, c. 1, lett. m), Codice Privacy).
Disambiguazione	La disambiguazione (in inglese Word Sense Disambiguation o, abbreviato, WSD) è l'operazione con la quale si precisa il significato di una parola o di un insieme di parole (frase), che denotano significati diversi a seconda dei contesti e che quindi sono ambigue .
Documento	La rappresentazione di atti, fatti e dati a prescindere dal supporto nella disponibilità della pubblica amministrazione o dell'organismo di diritto pubblico. La definizione di documento non comprende i programmi informatici (art. 2, c. 1, lett. c), D. Lgs. 36/2006).
Formati aperti	Secondo la legge italiana, si definisce aperto il "formato di dati reso pubblico, documentato esaustivamente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi" (art. 68 comma 3 lett. a) del Codice dell'Amministrazione Digitale).
Licenza standard	Il contratto, o altro strumento negoziale, redatto ove possibile in forma elettronica, nel quale sono definite le modalità di riutilizzo dei documenti delle pubbliche amministrazioni o degli organismi di diritto pubblico (art. 2, c. 1, lett. h), D. Lgs. 36/2006). Si tratta in generale di licenze condivisa, elaborata e approvata da comunità di giuristi nel Web che definisce le condizioni e le modalità di riutilizzo da parte di chiunque anche per scopi commerciali (conformemente all'art. 8 del D. Lgs. 36/2006, all'art. 9 della L. P. 16/2012 e all'art. 68 CAD).
Piattaforma	In generale una piattaforma è una tecnologia di base, sulla quale vengono poi sviluppate altre tecnologie. Può essere di tipo hardware (ovvero i circuiti sui quali vengono eseguiti i vari programmi) o software (ovvero un ambiente che permette e velocizza l'esecuzione di altri programmi).
Repository	Archivio o sito Web nel quale sono raccolti e conservati dati e informazioni corredati da descrizioni (metadati) in formato digitale, e direttamente accessibile dagli utenti. I repository rappresentano l'equivalente elettronico di una biblioteca.
Riutilizzo	La Direttiva 2013/37/UE definisce il riutilizzo come <i>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.</i>

Concetto	Definizione
Semantica	Si intende quella parte della linguistica che studia il significato delle parole. Accordarsi sulla definizione dei significati è un'attività complessa, ma permette di costruire categorie e reti di concetti utili a gestire grandi moli di dati e informazioni. Da ciò deriva la semantica computazionale, che è lo studio dell'automatizzazione del processo di ragionamento con l'ausilio di rappresentazioni del significato di espressioni di una lingua naturale.
Società civile	È un concetto molto ampio, complesso e in continua ridefinizione che comprende i seguenti attori sociali: sindacati, datori di lavoro e organizzazioni di produttori, organizzazioni non governative, organizzazioni che rappresentano interessi generali quali l'ambiente, i diritti umani, il benessere sociale, la salute e la cultura, le associazioni professionali e le organizzazioni di base. Include inoltre tutte le organizzazioni di volontariato e no profit che svolgono un ruolo importante nel dare voce alle preoccupazioni dei cittadini e nella fornitura di servizi in grado di soddisfare i bisogni della gente.
Thesaurus	Il thesaurus è un elenco strutturato in cui le parole sono raggruppate per somiglianza semantica (contenente sinonimi e qualche volta antonimi); si differenzia dal dizionario che contiene definizioni e pronuncia. Secondo la definizione ISO il thesaurus è "un vocabolario di un linguaggio di indicizzazione controllato in maniera formalizzata in modo che le relazioni a priori tra i concetti sono rese esplicite" (ISO 2788-1986). Le relazioni tesaurali sono utilizzate per facilitare la navigazione nel thesaurus agli utenti. Il thesauro inoltre si differenzia dall'ontologia, in quanto quest'ultima può contenere relazioni più complesse e una logica inferenziale inerente al modello .
Titolare del dato	La pubblica amministrazione o l'organismo di diritto pubblico che ha originariamente formato per uso proprio o commissionato ad altro soggetto pubblico o privato il documento che rappresenta il dato o che ne ha la disponibilità (art. 2, c. 1, lett. i), D. Lgs. 36/2006).
Trasparenza	Giuridicamente, si intende l'accessibilità totale delle informazioni concernenti l'organizzazione e l'attività delle pubbliche amministrazioni, allo scopo di favorire forme diffuse di controllo sul perseguimento delle funzioni istituzionali e sull'utilizzo delle risorse pubbliche (art. 1, c. 1, D. Lgs. 33/2013).
Uniform Resource Identifier	La locuzione Uniform Resource Identifier (in acronimo URI) in informatica, si riferisce a una stringa che identifica univocamente una risorsa generica, come ad esempio un indirizzo Web, un documento, un'immagine, un file, un servizio, un indirizzo di posta elettronica.
Uniform Resource Name	Uno Uniform Resource Name o URN è un URI che identifica una risorsa all'interno di un namespace, ma, a differenza di un URL , non permette l'identificazione della locazione della risorsa stessa. Un esempio di URN è il codice ISBN : questi identifica univocamente un libro, ma non ci dà alcuna informazione sulla locazione dello stesso.