

# DIAMO VALORE AI DATI DELLA PUBBLICA AMMINISTRAZIONE

L'AGENZIA PER L'ITALIA DIGITALE È IMPEGNATA NELLA DEFINIZIONE DI AZIONI PER GARANTIRE UN APPROCCIO METODOLOGICO COMUNE NELL'APERTURA DEI DATI E PER VALORIZZARE IL PATRIMONIO INFORMATIVO PUBBLICO. L'OPEN DATA DEVE DIVENTARE PARTE INTEGRANTE DELLA NORMALE OPERATIVITÀ DELLE AMMINISTRAZIONI.



Oggi l'open data è ampiamente considerato un efficace strumento per garantire trasparenza e responsabilizzazione della Pa, nonché incentivare il coinvolgimento civico dei cittadini. Esso consiste nel mettere a disposizione di chiunque dati che finora "vivevano" unicamente all'interno di applicazioni e di archivi chiusi, e di farlo abbattendo le restrizioni tecnologiche all'accesso e fornendo i vincoli legali minimi per il riuso dei dati. Il riuso dei dati viene altresì considerato come impulso alla creazione e allo sviluppo di servizi innovativi e applicazioni utili ai cittadini. Sebbene si riscontrino ancora forti inerzie da parte delle Pa nell'aprire dati in loro possesso, è doveroso sottolineare la nascita di numerose iniziative in tal senso: sono stati infatti creati tanti portali open data e un numero sempre maggiore di dataset viene reso disponibile. Tuttavia la proliferazione di portali, la disomogeneità e la mole dei dati rende difficoltosa, per gli utenti, la fruizione delle informazioni. In primo luogo i dati pubblicati risultano di non facile lettura e comprensione; in secondo luogo è elevata la complessità di sviluppo e mantenimento di applicazioni che possano gestire efficientemente

questi dati e allo stesso tempo produrre un servizio utile agli utilizzatori finali. È innegabile che, al momento, le prerogative dell'open data risultano disattese.

## I principi per l'apertura dei dati

Il problema di garantire contemporaneamente accesso e flessibilità di rappresentazione dei dati è stato affrontato dagli esperti del settore nel corso dell'ultima decade. Queste esperienze e best practices sono poi confluite in alcuni standard W3C e sono spesso raggruppate sotto la sfera del *web semantico*. Il *web semantico* è un unico e globale spazio web informativo di facile accesso e comprensione, sia da parte di umani che di macchine. È una visione analoga a quella che 20 anni fa fu del web come archivio globale di documenti. In questo spazio troviamo generiche risorse, definite semanticamente usando il framework concettuale Rdf (*Resource Description Framework*). Rdf consente di identificare univocamente una risorsa con un Uri, descriverla con dati e metadati e di metterla in relazione ad altre risorse. In

Rdf il costrutto informativo di base è la tripla <oggetto> <predicato> <risorsa>, in cui un soggetto è una risorsa, un oggetto è una risorsa o un valore, un predicato è una relazione tipizzata tra due risorse o una proprietà di una risorsa. Rdf può essere implementato con diverse forme sintattiche e i dati risultanti vengono chiamati *linked data*, proprio per la loro possibilità di riferirsi (cioè "collegarsi") tra loro. Questi dati, grazie all'identificazione tramite Uri, sono accessibili puntualmente attraverso l'infrastruttura Internet esistente, sono interrogabili mediante il protocollo e linguaggio di interrogazione Sparql e sono schematizzabili con i linguaggi Rdfs/Owl. Tipicamente si usano schemi (cioè ontologie) che sono condivisi globalmente, agevolando di fatto la comprensione. Il beneficio più evidente nell'utilizzare i linked data è il poter sfruttare la semantica incorporata internamente ai dati e le infrastrutture web per abilitare la scoperta e l'inferenza di nuove informazioni.

Recentemente Tim Berners-Lee, considerato il padre fondatore del web, ha proposto una serie di principi da seguire

nell'apertura di dati, definendo di fatto un modello, noto come modello delle 5 stelle, qualitativo per i dati aperti. Questo modello individua, non a caso, come livello massimo quello dei *linked open data*.

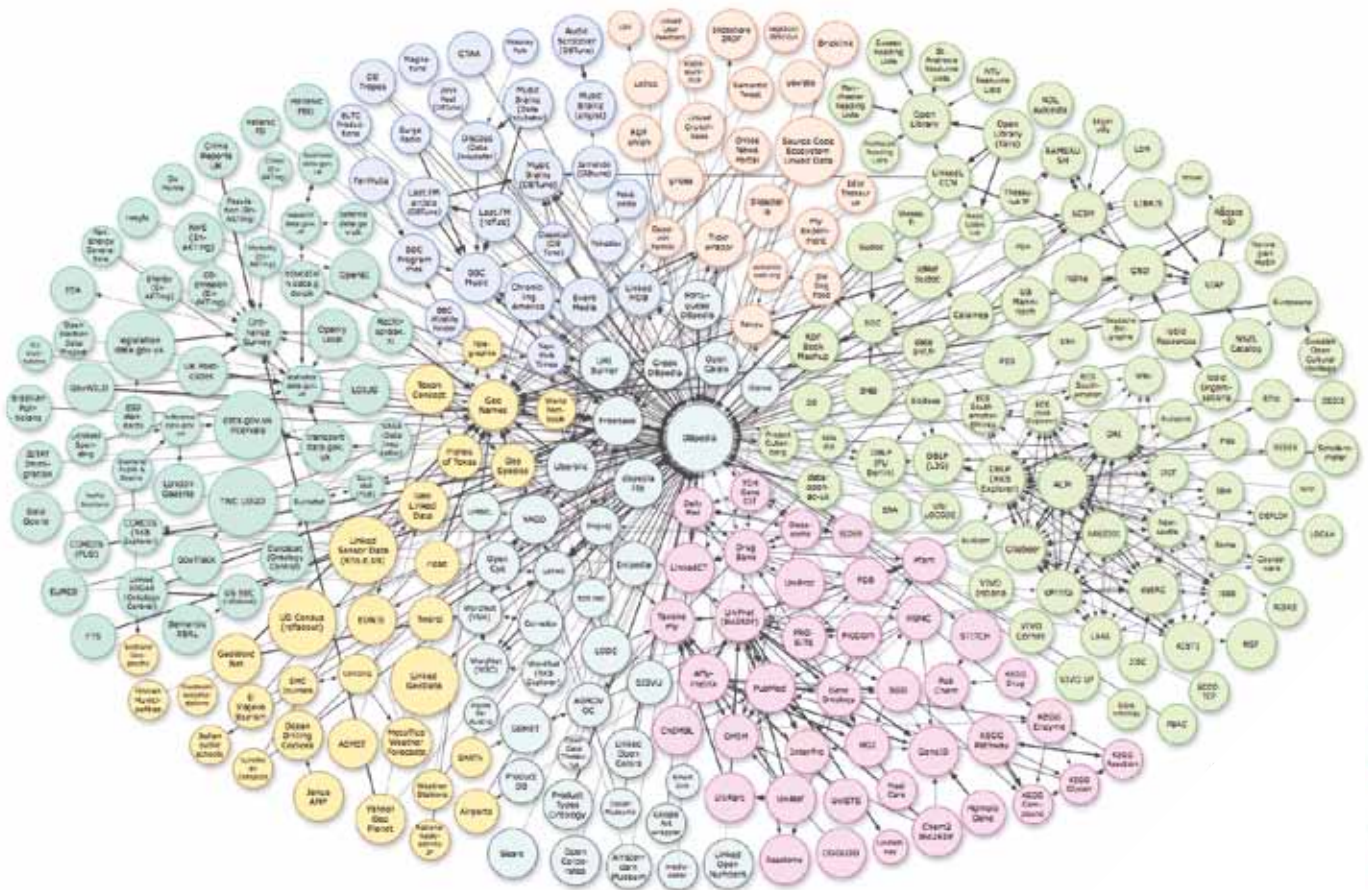
## Le linee guida italiane per la valorizzazione degli open data

L'Agenzia per l'Italia Digitale è ormai impegnata da più di un anno in questo scenario nell'ambito delle attività riguardanti l'interoperabilità semantica all'interno del *Sistema pubblico di connettività e cooperazione* (SpC). SpC è il framework italiano di interoperabilità istituito allo scopo di consentire alle Pa di cooperare efficacemente tra loro. I *linked open data* sono stati individuati come uno degli strumenti di abilitazione del livello di interoperabilità semantica all'interno della Pa e tra le Pa dei paesi degli stati membri. Le attività

dell'Agenzia hanno riguardato prima l'apertura, in formato *linked*, di dati in proprio possesso (disponibili attraverso il portale [spcdata.digitpa.gov.it](http://spcdata.digitpa.gov.it)), e poi con la definizione delle "Linee guida per l'interoperabilità semantica attraverso i *linked open data*". Esse propongono un approccio metodologico comune per le Pa produttrici di dati, indicando, per ogni fase, i passi operativi da compiere, gli strumenti da utilizzare e alcune opportune raccomandazioni da tenere in considerazione. Inoltre vengono trattati anche aspetti legali, come le licenze d'uso, ed economici, come quelli riguardanti la catena del valore degli open data. Il valore e l'unicità di questo lavoro è stato riconosciuto a livello di Comunità europea e ha consentito all'Agenzia di essere invitata a partecipare in gruppi di lavoro, come quello sul vocabolario comune dei servizi pubblici (Cpsv), e chiamata in causa per la raccolta di best practices a livello comunitario, come quelle sulla definizione di Uri persistenti. Considerato che l'Agenzia ha il

dovere di creare opportunità di sviluppo economico, le sue azioni sono orientate all'abilitazione dei possibili modelli di business dell'open data e alla valorizzazione del patrimonio informativo pubblico. Verranno considerate diverse categorie di modelli di business: la creazione diretta di servizi finali (es. premium, freemium ecc.), lo sviluppo di servizi da terze parti (es. concorsi per app, *hackaton*, ecc.), i servizi di gestione e generazione dell'informazione (es. *business intelligence*) e infine, la sponsorizzazione e la promozione di marchi. In un'era in cui l'importanza del dato è quanto mai centrale, a prescindere dallo specifico modello di business, il vero valore aggiunto è l'informazione che si riesce a generare dai dati. Questa motivazione rafforza la volontà dell'Agenzia di promuovere i *linked open data*, soprattutto negli ambiti cosiddetti data-centrici.

Da decreto, l'Agenzia ha il compito, tra gli altri, di definire annualmente l'agenda



**Fig. 1. Rilascio di open data**  
Basi di dati pubblicate nella Linked Open Data Cloud. I nodi rappresentano i singoli dataset pubblicati e gli archi indicano la presenza di connessioni semantiche tra essi.

Tipologia di dataset

- Media
- Geografici
- Pubblicazioni
- Contenuto generato dagli utenti
- Pubbliche amministrazioni
- Cross-domain
- Scienze naturali

Fonte: Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net>. Licenza CC-BY-SA.

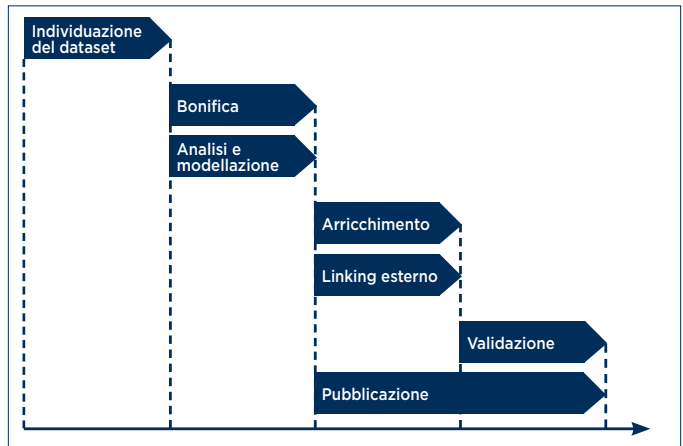


nazionale e le linee guida nazionali per la valorizzazione del patrimonio dei dati pubblici. L'agenda deve contenere gli obiettivi delle politiche di valorizzazione del patrimonio informativo pubblico. Le linee guida sono un documento di indirizzo con indicazioni concrete sui passi da compiere per l'apertura dei dati. Per la loro definizione, l'Agenzia ha istituito un gruppo di lavoro con Pa centrali e locali. Inoltre entro febbraio di ogni anno, l'Agenzia deve fornire un rapporto annuale sullo stato di attuazione di tale processo di valorizzazione. In questo scenario, l'Agenzia proporrà una serie di azioni col fine di agevolare le Pa nell'apertura dei loro dati e dare una visione organica e sinergica tra le diverse iniziative. Si ritiene particolarmente importante far leva sulla totale consapevolizzazione delle Pa sul valore dell'open data. Al momento le iniziative sono spesso viste solo come degli adempimenti per l'area dei sistemi informativi e comunque sconnesse dalle funzioni dell'amministrazione. L'Agenzia cercherà di invertire questa rotta, promuovendo il concetto per il quale l'open data deve essere parte integrante del processo operativo e amministrativo di gestione del flusso informativo.

**Antonio Maccioni**  
 Agenzia per l'Italia Digitale

**FIG. 2**  
 RILASCIO  
 DI OPEN DATA

Le fasi dell'approccio metodologico all'interoperabilità semantica attraverso linked open data in un possibile piano di rilascio.



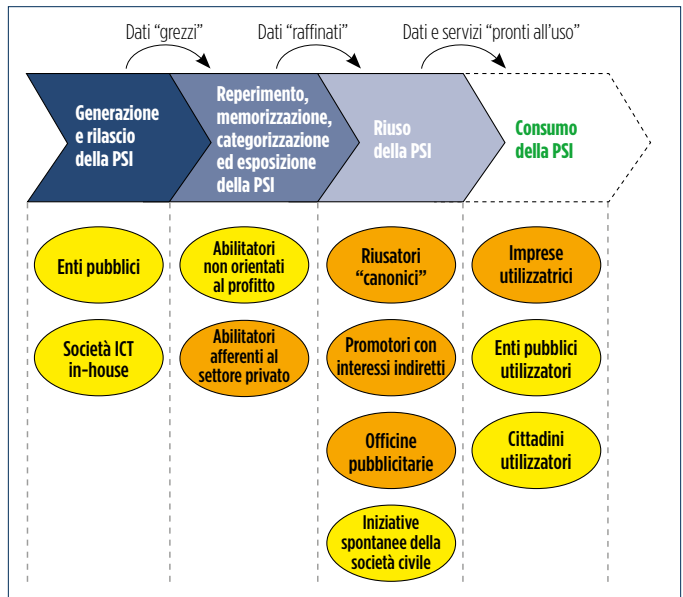
**FIG. 3**  
 CATENA DI VALORE

Catena del valore legata alla PSI (Public Sector Information, dati rilasciati dalla pubblica amministrazione).

**LEGENDA**

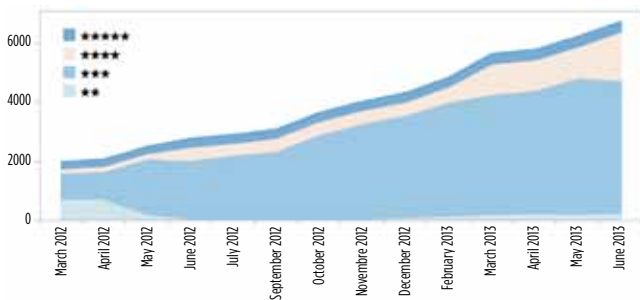
- Attori non orientati al profitto
- Attori orientati al profitto

Fonte: Commissione di coordinamento Spc, "Linee guida per l'interoperabilità semantica attraverso i linked open data", 2012.

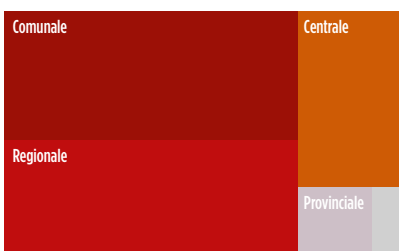


**IL MONITORAGGIO DELL'OPEN DATA IN ITALIA**

Il portale dati.gov.it pubblica periodicamente gli aggiornamenti delle rilevazioni sullo stato dell'open data in Italia. I dati sono presentati attraverso delle infografiche dinamiche. Al 23 giugno 2013 sono disponibili 6726 dataset rilasciati in formato aperto.



Ripartizione dei dataset per livello di riusabilità (scala di Tim Berners-Lee).



Ripartizione del numero di dataset rilasciati rispetto al livello amministrativo degli enti.



Distribuzione geografica delle oltre 70 amministrazioni che rilasciano open data. L'area di ognuna delle bolle è direttamente proporzionale al numero di dataset rilasciati.