

LINKED OPEN DATA: LA NUOVA FRONTIERA DEGLI OPEN DATA

di Stefano De Francisci (defranci@istat.it),
Monica Scannapieco (scannapi@istat.it)

■ Il crescente livello di informatizzazione della Pubblica Amministrazione da una parte e lo sviluppo delle tecnologie digitali dall'altra conferiscono ai dati del settore pubblico un ruolo sempre più importante ai fini della trasparenza amministrativa, della partecipazione pubblica e dell'accrescimento economico e sociale. Per questo, pubblicare e scambiare dati secondo modalità che ne permettano il riuso in contesti nuovi e anche diversi rispetto ai fini per i quali i dati vengono raccolti, diventa un obiettivo importante cui mirare.

In Italia, grazie alle strategie espresse dall'Agenda Digitale e all'adesione alle iniziative internazionali (Revisione della Direttiva 2003/98/EC sul riuso dei dati pubblici <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:IT:PDF>), un numero crescente di amministrazioni si sta facendo carico del lavoro necessario per pubblicare i propri dati in formato "aperto".

Tuttavia, per rendere ancor più efficace il percorso verso una consapevole usabilità dei dati occorre andare oltre il generico stato "open" e perseguire un obiettivo più ambizioso ed evoluto, costituito da una duplice azione: rendere comprensibili a livello semantico i dati e favorire il collegamento tra dataset differenti. Il movimento Linked Data (<http://linkeddata.org/>) si concentra sulla messa a punto di strumenti che permettono di dare ai dati (aperti o non) un'identità e di renderli collegati tra loro e interoperabili.

Se i Linked Data sono aperti sono indicati come Linked Open Data (LOD).

UN PARADIGMA DI NUOVA GENERAZIONE

Con il paradigma LOD, i dati sono sempre corredati da metadati; in particolare, sono rappresentati da linguaggi e modelli standardizzati dal W3C (World Wide Web Consortium). L'utilizzo di tali linguaggi (ad esempio RDF - Resource Description Framework) permette che all'interno del Web i dati divengano risorse, univocamente identificabili, collegabili tra di loro e accessibili tramite HTTP. E' facile, dunque, sviluppare applicazioni (come web services o applicazioni per la visualizzazione) che consentano di accedere direttamente ai dati. Inoltre, i dati sono interrogabili tramite un linguaggio che permette all'utente di formulare in modo libero le proprie richieste. Esplorando i link è inoltre possibile "navigare" tra i dati esattamente come si navigano le pagine HTML sul Web, realizzando il cosiddetto "Web of Data". Il solo paradigma Open, invece, garantisce l'accessibilità ai dati per elaborazioni successive, ma non assicura la possibilità di interpretarli sulla base della loro semantica e soprattutto non solleva l'utente dalla necessità di disporre di applicazioni che consentano di utilizzarli.

LINKED OPEN DATA IN ISTAT

L'Istat è il principale fornitore di dati open tra tutte le pubbliche amministrazioni italiane: come testimonia un'analisi fatta dal Foromez (S. Marras, Open data: Lo scenario di riferimento, XI Conferenza Italiana di Statistica, 2013) fornisce più di 600 dataset, la maggior parte dei

quali pubblicati sul Web warehouse I.stat da cui è possibile effettuare il download dei dati nei formati CSV, Excel e SDMX.

Nonostante ciò, rispondendo alle indicazioni dell'Agenzia per l'Italia Digitale e considerata la rilevanza del paradigma LOD, a partire dal 2012 sono state avviate sperimentazioni mirate a testarne il possibile utilizzo per la diffusione dei dati.

Una prima sperimentazione ha sviluppato un sistema software per la traduzione automatica da dati espressi secondo lo standard ISO per la condivisione dei dati statistici SDMX, a dati in RDF Data Cube Vocabulary, promosso nell'ambito del Web semantico dal W3C per la rappresentazione di dati multidimensionali. Nel corso del 2014 è prevista un'integrazione del traduttore con il Single Exit Point (SEP) al fine di arricchirlo con un'uscita in formato RDF. Una seconda sperimentazione ha testato la possibilità di collegare i dati del Censimento della Popolazione al livello di sezione di censimento con l'archivio dei numeri civici e con basi di dati territoriali mediante LOD. Nel corso del 2014, è prevista la pubblicazione effettiva in formato LOD dei dati del Censimento delle Popolazioni al livello di sezione di censimento, collegati a basi territoriali. Il percorso verso l'assunzione del nuovo paradigma è quindi intrapreso e a breve l'Istituto offrirà agli utenti modalità di fruizione del proprio patrimonio informativo ancora più ricche ed efficaci.

