

istat working papers

N.3
2011

Tecnologie e servizi per la condivisione dei dati

Daniele Frongia, Vincenzo Patruno

istat working papers

N.3
2011

Tecnologie e servizi per la condivisione dei dati

Daniele Frongia, Vincenzo Patruno

Comitato di redazione

Coordinatore: Giulio Barcaroli

Componenti:

Rossana Balestrino	Francesca Di Palma	Luisa Picozzi
Marco Ballin	Alessandra Ferrara	Mauro Politi
Riccardo Carbini	Angela Ferruzza	Alessandra Righi
Claudio Ceccarelli	Danila Filipponi	Luca Salvati
Giuliana Coccia	Cristina Freguja	Giovanni Seri
Fabio Crescenzi	Aurea Micali	Leonello Tronti
Carla De Angelis	Nadia Mignolli	Sonia Vittozzi

Segreteria:

Lorella Appolloni, Maria Silvia Cardacino, Laura Peci, Gilda Sonetti, Antonio Trobia

Istat Working Papers

Tecnologie e servizi per la condivisione dei dati

N. 3/2011

ISBN 978-88-458-1674-1

Istituto nazionale di statistica
Servizio Editoria
Via Cesare Balbo, 16 – Roma

Tecnologie e servizi per la condivisione dei dati

Daniele Frongia, Vincenzo Patruno

Sommario

Uno dei fattori che ha contribuito maggiormente al successo di YouTube sta nella porzione di codice all'interno dei tag `<object>` e `</object>` che compare a fianco di ogni video. Copiare e incollare questa riga all'interno di una pagina web ha l'effetto di incorporare il filmato desiderato all'interno di qualunque sito o blog. Queste porzioni di codice così come altri tipi di tecnologie come i feed RSS e le API stanno offrendo delle importanti opportunità di condivisione sia per gli utenti che per aziende e istituzioni. Per queste ultime vedremo il modello di condivisione dei dati nell'ambito del Census Hub Project per la diffusione dei risultati del prossimo Censimento 2011. Valuteremo inoltre le esperienze di alcune organizzazioni internazionali come OCSE e Nazioni Unite, che confermano come l'utilizzo di nuovi servizi e tecnologie online possa costituire un'ottima opportunità per condividere dati e per raggiungere un pubblico più diffuso e partecipe.

Parole chiave: Web 2.0, condivisione dati, visualizzazione dati, API, Web Services

Abstract

One of the factors that contributed most to the success of YouTube is in the portion of code within the tags `<object>` and `</object>` closed to each video. Copying and pasting these few lines of code into a web page has the effect to embed the video into any site or blog. This code as well as other technologies such as RSS feeds and APIs do offer the opportunity of sharing data both for users and for companies and institutions. For the latter, we will discuss the sharing model data within the Census Hub Project for the dissemination of results of the forthcoming 2011 Census. Besides, we will consider the experiences of some international organizations such as OECD and the UN, confirming that the use of new technologies and online services can be an excellent opportunity to share information and to reach a more involved audience.

Keywords: Web 2.0, Data Sharing, Data Visualization, API, Web Services.

1. Introduzione¹

C'era una volta l'"hyperlink". Ovviamente esiste ancora (ed è in buona salute), ma in passato un link ad una qualunque risorsa raggiungibile sul Web costituiva il solo modo che si aveva per poter accedere ai contenuti di quella risorsa. E infatti, quando all'interno di un documento HTML si faceva riferimento ad esempio a un dato, ad una informazione, a un prezzo, ad una notizia, la consuetudine era quella di replicarne il contenuto e di "linkare" il documento alla pagina web contenente quel dato, quella informazione, quel prezzo, quella notizia.

Questo fino a quando la rapida e incessante evoluzione delle tecnologie nell'attuale Società dell'Informazione ha contribuito in modo decisivo al cambiamento degli scenari di riferimento per quanto riguarda le modalità di accesso a informazioni e dati. Il tutto in un contesto in cui l'utente Web ha cessato di essere un "lettore" del Web per diventarne invece il protagonista. Un protagonista che a sua volta veicola l'informazione, la contestualizza, la condivide, la discute, la promuove, la valorizza.

E non si tratta di un mero "upgrade" dell'utente, del suo protagonismo, della sua voglia di mostrarsi. I paradigmi che stanno alla base del Web 2.0 sono infatti complessi e articolati; e trovano un solido fondamento nelle centinaia di milioni di utenti che ormai da anni utilizzano servizi online come i social network. I fattori chiave del Web 2.0 sono:

- a) I siti Web 2.0 sono piattaforme che consentono una forte interazione tra utenti
- b) Aumentano centralità e protagonismo degli utenti, che forniscono valore aggiunto con la produzione di contenuti e la condivisione della conoscenza
- c) Gli utenti usufruiscono di servizi innovativi mediante potenti interfacce grafiche
- d) I servizi offerti vengono aggiornati di continuo
- e) Vengono create nuove applicazioni basate su informazioni provenienti da altri servizi

I servizi 2.0 più noti si contraddistinguono per le modalità innovative in cui le informazioni (video, foto, articoli ecc.) vengono presentate.

In questo contributo ci occupiamo di tecnologie e di servizi per la condivisione di dati strutturati (dataset). In particolare, nei primi due paragrafi parleremo delle tecnologie di condivisione mentre nel terzo vedremo come vengono applicate da utenti e istituzioni in Europa e oltreoceano.

2. Condivisione di video e widget

Partiamo da YouTube. Uno dei fattori chiave che ha contribuito al successo del "mega portale" di condivisione video sta probabilmente in quella porzione di codice all'interno dei tag `<object>` e `</object>` che troviamo di fianco ad ogni filmato.

Copiare e incollare queste poche righe all'interno di una pagina web ha l'effetto di "incorporare" il filmato desiderato all'interno di qualunque sito o blog. Dal punto di vista più strettamente tecnico, quelle righe di codice hanno l'effetto di scaricare ed utilizzare un "embedded player" all'interno del quale verrà visualizzato il filmato in questione. A patto di avere installato sul proprio browser Adobe Flash Player (oppure per i più temerari Gnash, la sua alternativa "libera"), il plug-in necessario affinché questo tutto questo possa avvenire correttamente. Questa condivisione di tipo "embed", che consente cioè di incorporare il video all'interno di documenti esterni a YouTube presenta, su tutti, due vantaggi.

Il primo è sicuramente la facilità di condivisione. Un semplice copia/incolla è alla portata di tutti. Il secondo è che non viene effettuata alcuna copia del video. Quando lo stesso video viene incorporato su più pagine web esterne, in realtà la copia del filmato sarà sempre una, resterà sempre sulla piattaforma YouTube e verrà sempre scaricata dai server dei data center di Google. La gestio

¹ Questo contributo è stato realizzato grazie al lavoro congiunto di Daniele Frongia e Vincenzo Patruno. In particolare Daniele Frongia ha curato l'introduzione e i paragrafi 3 e 4; Vincenzo Patruno ha curato l'introduzione e i paragrafi 2 e 3.

ne del video, cioè, sarà sempre centralizzata e localizzata in un unico punto della Rete. In altre parole, se chi ha caricato il video decidesse un giorno di cancellarlo da YouTube, questo sparirebbe immediatamente da tutti i siti e blog su cui era stato condiviso.

Questo la dice lunga anche sul valore che può avere la condivisione in chiave "advertising". E le ultime modifiche al player YouTube per proporre video correlati a quello visionato e pubblicità vanno proprio in questa direzione. Se non ci fosse questo aspetto, non avremmo la disponibilità gratuita delle piattaforme in questione. Ma la condivisione ha un valore anche e soprattutto sociale, che non solo permette la circolazione simultanea di idee, arte, notizie, dati e cultura sul Web, ma consente alle idee, ai progetti e alle intuizioni migliori, all'arte, all'informazione, in altre parole a tutto ciò che suscita interesse in Rete (e visto il mestiere che facciamo, ai dati e alle statistiche di qualità), di emergere e di propagarsi senza intermediari, senza manipolazioni e senza censure.

Le stesse logiche utilizzate per condividere video possono essere utilizzate per condividere altri tipi di oggetti. E infatti possiamo condividere in maniera embed immagini, file audio, applet, animazioni flash e qualunque altro user generated content. Anche i dati statistici ufficiali del nostro paese. Questi oggetti vengono comunemente detti widget. E' possibile realizzarli da soli oppure ci si può avvalere anche in questo caso delle numerose piattaforme a disposizione sul Web.

Su tutte WidgetBox, ClearSpring oppure SpringWidgets. In questo caso buona parte del lavoro è stato già fatto. Si tratterà di "riempire" il widget con il contenuto desiderato. E per propagare il widget in modo efficace, ci si potrà avvalere delle straordinarie opportunità offerte dai social network. I widget vengono infatti in genere predisposti per essere condivisi con pochi clic sulle maggiori piattaforme "social" come Facebook, MySpace, Hi5 e così via. In modo da tentare di sfruttare l'eventuale viralità che si può ottenere propagando il widget su queste piattaforme. E per sottolineare l'importanza cruciale che rivestono i social network sul fronte della condivisione, è significativo il recente rilascio di una nuova funzionalità di YouTube che va sotto il nome di autosha-ring. Si tratta della possibilità di "agganciare" l'utenza YouTube a quella di Facebook (la stessa cosa funziona anche con Twitter), in modo tale da generare automaticamente un link verso YouTube su Facebook ogni volta che su YouTube viene caricato un video. E questo nonostante Google e Facebook siano concorrenti.

3. Tecnologie per la condivisione: API, web services e applicazioni mash-up

L'idea che è alla base dei widget è che posso sistemare degli oggetti all'interno di una pagina web il cui contenuto viene gestito centralmente dal sito web su cui il widget si trova fisicamente. L'idea pertanto è quella che c'è un soggetto "produttore" di dati e di informazioni. E una serie di soggetti "consumatori" di quei dati e di quelle informazioni. Ma consumare quei dati e quelle informazioni attraverso un widget è solo una delle possibilità che possono essere messe a disposizione. Una possibilità per altro piuttosto "passiva" in quanto non sono previsti margini di intervento sul contenuto per coloro che condividono il widget, a parte magari qualche operazione di customizzazione del layout. Una volta incluso il codice in una pagina web, il widget si presenterà sempre allo stesso modo e la sua gestione sarà localizzata sul sito su cui il widget si trova. C'è però la possibilità prelevare dati, informazioni e servizi in modo diverso. Una nuova generazione di sistemi e di siti web sta infatti nascendo utilizzando le tecniche comunemente dette mash-up. Un mash-up è una applicazione che viene costruita assemblando tra loro dati e servizi di terze parti attraverso le cosiddette API, ossia di un insieme di tecniche che consentono ad applicazioni software di poter colloquiare e interagire tra loro. E' proprio tramite l'impiego delle API che ad esempio è possibile utilizzare le funzionalità offerte da Google Maps e costruire, sul proprio sito, applicazioni che visualizzano e interagiscono con mappe. E lo stesso si può dire delle API di Yahoo, Amazon, eBay, Flickr e così via, pensate ognuna per assolvere precise funzionalità su un certo tipo di dati e informazioni. Ci sono poi i web services, di cui parliamo nel paragrafo dedicato al progetto Census Hub di Eurostat.

Figura 1: Widget che mostra i dati Istat sulla popolazione del Comune di Roma

Popolazione Comune di Roma		
	2007	2008
Popolazione	2718768	2724347
Maschi	1281482	1283145
Femmine	1437286	1441202
Nati	24645	27603
Morti	25258	25914
Saldo migratorio	13778	3890

Fonte ISTAT

Embed code: "<script type="text/javascript" src="http://www.vincenzopatrano.org/dir/bilancio.js"></script>"

In figura 1 è mostrato un esempio di widget contenente i dati sulla popolazione residente nel comune di Roma. L'aspetto interessante è che i dati provengono via web services e in tempo reale da demo.istat.it, sito web Istat di diffusione dei dati sulla popolazione. Quando l'Istat renderà pubblici i dati per l'anno 2009, la tabella si aggiornerà automaticamente su tutti i siti web e blog che l'avranno nel frattempo inserita nelle proprie pagine attraverso l'embed code. E sempre parlando di dati Istat, potete vedere come è possibile condividere anche un'intera applicazione Web, anche questa realizzata utilizzando alcune API Istat, andando a copiare il codice

```
<script type="text/javascript" src="http://www.vincenzopatrano.org/dir/net_migration.js"></script>
```

e a incollarlo dove preferite all'interno di una qualunque pagina HTML.

4. Servizi e progetti di condivisione dati

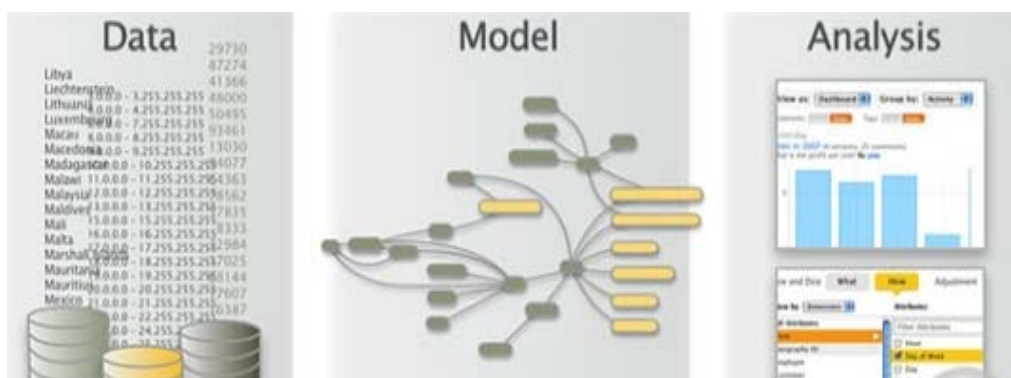
Come detto in precedenza, le nuove tecnologie web costituiscono un'opportunità per utenti e istituzioni per condividere e riusare dati. Vediamo alcuni servizi e esperienze in ambito internazionale.

Good Data (<http://www.gooddata.com/>)

Good Data è un servizio che consente di memorizzare dati online e fornisce strumenti per analizzarli in un ambiente collaborativo².

² <http://segnalazionit.org/2009/11/good-data>

Figura 2: Good Data



La Nuova Zelanda utilizzerà questa piattaforma per condividere i dati del proprio censimento in modo tale che possano essere visualizzati, combinati ed elaborati da soggetti terzi.

Swivel (<http://www.swivel.com>)

Swivel, un altro servizio per la condivisione di dataset, è stato il sito più *alla moda* per le istituzioni: organizzazioni come OECD, UNESCO Institute for Statistics, UNECE, United States Census Bureau e United States Department of Agriculture hanno aderito a Swivel creando un proprio canale ufficiale.

Nonostante un forte impulso iniziale, nel corso degli ultimi due anni questi canali ufficiali sembrano aver perso i loro appeal, complice il fatto di non aver opportunamente investito nella costruzione di una community di utenti (anche per la statistica occorre utilizzare tecniche di marketing!). Alcuni canali sono stati chiusi mentre il servizio si è rifatto il look e prova il rilancio con nuove funzionalità.

Figura 3: Swivel



IBM Many Eyes (<http://manyeyes.alphaworks.ibm.com/manyeyes/>)

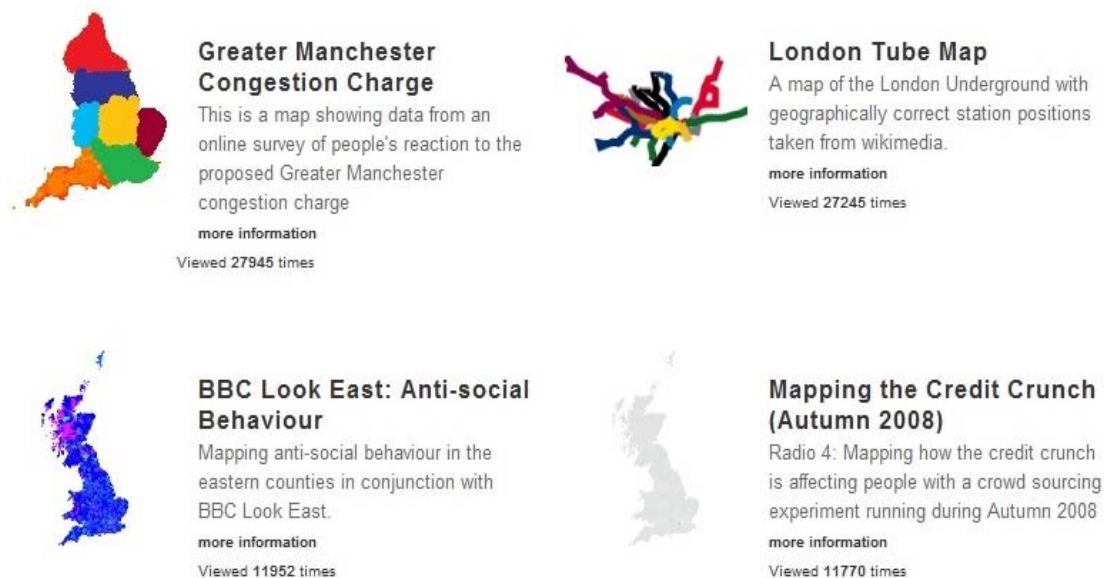
Many Eyes, nato in casa IBM, è un progetto simile a Swivel ma che offre una vasta possibilità di visualizzare i dati in modalità interattiva. Non solo visualizzare: descrivere, commentare e, analogamente a quanto visto con YouTube, riutilizzare le visualizzazioni incorporandola su altri blog o siti.

Many Eyes si descrive così: *una scommessa sulla potenza dell'intelligenza visiva umana per identificare modelli. Il nostro obiettivo è quello di "democratizzare" la visualizzazione e per consentire un nuovo tipo sociale di analisi dei dati.* Anche Many Eyes è utilizzato da diverse organizzazioni internazionali tra cui Eurostat, presente sulla piattaforma con numerosi dataset.

MapTube (<http://www.maptube.org/>)

MapTube, sviluppato dal Centro di analisi spaziale avanzata dell'University College London, permette agli utenti di condividere e di sovrapporre mappe per confrontare visivamente diversi insiemi di dati. Nel "carrello" è infatti possibile aggiungere più mappe e poi sovrapporle. In questo caso, come per Many Eyes, l'accento è posto sui mash-up e sulla visualizzazione interattiva.

Figura 4: MapTube



Google Fusion Tables (<http://tables.googlelabs.com/>)

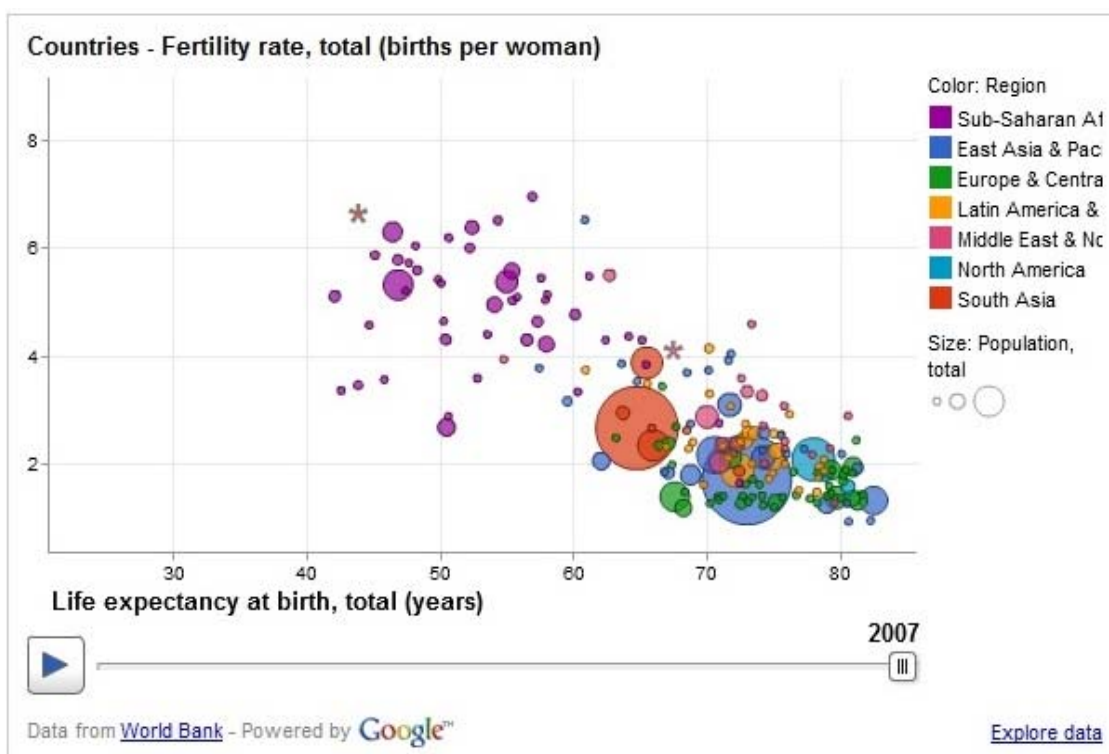
Tra questi servizi non poteva certo mancare Google che, con Fusion Table consente di caricare dati, visualizzarli, creare mash-up con le mappe, limitare (se necessario) l'accesso, combinare più fonti e ovviamente "discutere" i dati con la community.

Google Public Data Explorer (<http://www.google.com/publicdata/home>)

Public Data Explorer è un altro servizio dei laboratori di Google che deriva dall'esperienza, di successo, di Public Data³. Il suo scopo è rendere *grandi serie di dati facili da esplorare, visualizzare e comunicare*. E' uno strumento pensato per studenti, giornalisti, politici, istituzioni e chiunque voglia "giocare" per creare visualizzazioni e incorporarle altrove. Anche in questo caso, come per i video di YouTube, i dati risiedono in un luogo unico su un server, e quindi i grafici che si basano su di essi possono rimanere sempre aggiornati automaticamente intervenendo su un unico dataset.

³ <http://segnalazionit.org/2009/12/google-public-data-istat/>

Figura 5: Google Public Data Explorer



Come abbiamo visto, questi siti rappresentano potenziali partner per le istituzioni, in quanto forniscono un canale alternativo per la diffusione dei dati statistici.

Un'indagine condotta dall'UNECE sull'attuale utilizzo del Web 2.0 ha rivelato, nel primo semestre 2008, una ridottissima partecipazione attiva degli istituti di statistica su tali siti web⁴. Nel 2009/2010 la tendenza sembra essersi invertita e sono numerosi i casi di istituti di statistica che hanno iniziato ad utilizzare metodi e strumenti Web 2.0.

Vediamo ora alcune organizzazioni ed esperienze internazionali che utilizzano questi servizi o ne che ne hanno realizzato dei "propri":

UN Data, a world of information (<http://data.un.org/>)

Le Nazioni Unite hanno pubblicato un nuovo sito web chiamato *UN Data* per la condivisione dei dati provenienti dalle proprie agenzie. Al momento vi sono raccolti oltre 60 milioni di record, esplorabili e visualizzabili in forma tabellare, oltre che scaricabili in formato XML, CSV o testo⁵. Nei primi due anni di vita oltre 6 milioni di utenti hanno effettuato ricerche su PIL, popolazione, povertà, aspettativa di vita e altri fenomeni nei vari paesi. Un buon risultato, tenendo conto che si tratta pur sempre di statistiche ufficiali, e quindi con un "appeal di partenza" piuttosto contenuto.

Data.gov (<http://www.data.gov>)

Anche Data.gov è un sito ufficiale, stavolta degli Stati Uniti, e il suo scopo è quello di *aumentare l'accesso del pubblico* ai dataset di alta rilevanza sociale prodotti dal governo federale. Non è un caso, quindi, che una sezione del sito sia dedicata all'Open Government⁶, su cui torneremo nel prossimo paragrafo. Anche questo servizio consente di votare e scaricare dati, anche in

⁴ <http://segnalazionit.org/2009/07/il-web-2-0-e-gli-istituti-di-statistica-12/>

⁵ <http://www.stat-project.com/commenti.asp?ID=386>

⁶ http://en.wikipedia.org/wiki/Open_government

formato KML/KMZ per la visualizzazione con Google Earth e Maps. Da notare, infine, che a differenza di altri servizi analoghi, Data.gov fornisce anche metadati accurati per ciascun dataset.

Data.gov.uk (<http://data.gov.uk>)

Questo sito è la risposta UK a quello USA e il suo taglio è esplicitamente chiarito sulla home: *Unlocking innovation, working with UK Public Sector information and data*. In questo caso quindi viene posto l'accento sul fattore "innovazione" e, non a caso, nel progetto sono presenti personaggi del calibro di Tim Berners-Lee e Nigel Shadbolt. Data.gov.uk, complessivamente, è un progetto molto ambizioso che prevede la costruzione di una forte interazione con gli utenti, che possono proporre idee e sviluppare applicazioni aggiuntive.

Eurostat Census Hub Project (<http://cswsq.wikidot.com/wiki:eurostat-hub>)

Riprendiamo a parlare di tecnologie di condivisione e ripartiamo dai web services, che sono delle particolari API implementate esclusivamente sul Web. Ed è proprio tramite un modello architetturale basato su web services che Eurostat diffonderà al pubblico i dati dei prossimi censimenti della popolazione previsti nel 2011. La particolarità di questo tipo di approccio è che è basato su un modello di condivisione adottato da tutti gli istituti di statistica europei. In pratica, se interroghiamo il sistema chiedendo i dati di un singolo paese, questi verranno prelevati in tempo reale direttamente dal server dell'istituto del paese per cui i dati vengono richiesti. Se il sistema viene interrogato chiedendo una tavola in cui dati di vari paesi europei vengono messi a confronto, i dati verranno richiesti a ciascun server di ciascun paese e "assemblati" in modo tale da poter essere presentati ad esempio sotto forma di tavola HTML all'utente che ne ha fatto richiesta. Il sistema Eurostat pertanto non produce dati propri, ma si limita a veicolare le richieste dati, a prelevare i dati direttamente dal server dell'istituto che li produce e ad assemblare il tutto per permettere al sistema di dare una risposta all'utente. Con il vantaggio di poter "confinare" all'interno di ogni singolo istituto la manutenzione dei dati censuari del singolo paese che quell'Istituto rappresenta. Per questa sua particolarità, il progetto va sotto il nome di Census HUB e rappresenta un interessante caso di applicazione mash-up che si avvale di un'intera architettura costruita seguendo specifiche ben precise e standard condivisi per la trasmissione dati. In particolare, viene utilizzato il framework SDMX, un insieme di standard messi a punto proprio per la trasmissione sul Web di dati e metadati di tipo statistico. Ovviamente la questione è più complessa e articolata, e rappresenta non solo la dimostrazione che si può lavorare proficuamente in modo cooperativo, ma anche che le tecnologie attuali, grazie anche ad investimenti orientati in questa direzione, stanno evolvendo naturalmente verso modelli più sofisticati e raffinati di data sharing.

5. L'apertura delle istituzioni: un percorso obbligato

E' ormai chiaro come aziende e istituzioni (soprattutto le prime) abbiano compreso le potenzialità contenute nei comportamenti degli utenti nella Rete.

Da una proficua interazione tra produttori di dati e di utenti, le istituzioni possono⁷:

- a) Raggiungere un pubblico più vasto e ridurre la distanza con i cittadini
- b) Comunicare i dati con più efficacia e coinvolgere maggiormente gli utenti
- c) Migliorare la propria reputazione e il posizionamento online⁸
- d) Offrire la possibilità ai cittadini di utilizzare e riutilizzare dati, generando così nuova conoscenza.

⁷ <https://docs.google.com/present/edit?id=0AXEFdBzQXflnZDliejJ6a18yMjFmeDM3czR6dw&hl=it>

⁸ <http://www.sis-statistica.it/magazine/spip.php?article169>

Concludiamo riprendendo il già citato progetto di Open Government negli Stati Uniti. Scrive Sebastiano Amato⁹:

L'espressione *Open Data* è utilizzata da tempo in informatica e indica la **divulgazione di dati** senza alcuna censura, copyright o brevetto: insomma, senza alcuna restrizione.

In sé, quindi, questa filosofia non rappresenta una novità; ciò che invece potrebbe essere innovativo è l'uso che dell'Open Data si potrebbe fare **nella pubblica amministrazione** per la realizzazione dell'*Open Government*. Sono già molteplici gli esempi esistenti dell'applicazione di questo concetto che, in breve, impegna l'amministrazione pubblica a realizzare una piattaforma di *data storage* e a **concederne l'accesso in maniera libera**, delegando ai privati l'onere, il dovere e la responsabilità di rappresentazione ed elaborazione (...). Così facendo si instaurano **meccanismi virtuosi di trasparenza** e di collaborazione con il cittadino nell'impegno al miglioramento della collettività, cui tutti possiamo e dobbiamo partecipare.

Inoltre la sensazione di appartenenza ad uno stato pubblico si amplifica in quanto il cittadino vi **entra come parte in causa** e non solamente come beneficiario ultimo.

Nell'era del Web 2.0, dunque, il processo di apertura delle istituzioni e della condivisione dei dati è diventato un percorso obbligato. I dati ottenuti con soldi pubblici devono essere pubblici. Lo stiamo sentendo da tempo e sembra che il concetto sia forse cominciato ad essere tiepidamente considerato dalla politica, se è vero che in una recente intervista¹⁰ al direttore di Wired Italia, il ministro Brunetta ha annunciato che entro l'anno anche l'Italia avrà il suo Data.gov, un portale da cui scaricare in modo da poter riutilizzare i dati della Pubblica Amministrazione. La "partecipazione" dei cittadini alla cosa pubblica passa, anche, attraverso la condivisione dei dati: una questione che ha quindi a che fare, oltre che con la tecnologia, con la democrazia.

⁹ <http://www.zeusnews.it/index.php3?ar=stampa&cod=12130&numero=999>

¹⁰ <http://www.wired.it/news/archivio/2010-06/03/liveblogging-con-brunetta.aspx>