



Interreg Mediterranean



odeon

Open Data for European Open iNnovation

Utilizzare gli Open Data nei propri prodotti: elementi tecnici e fattori di criticità

12 Novembre 2020

prof. Andrea Albarelli - Università Ca' Foscari Venezia

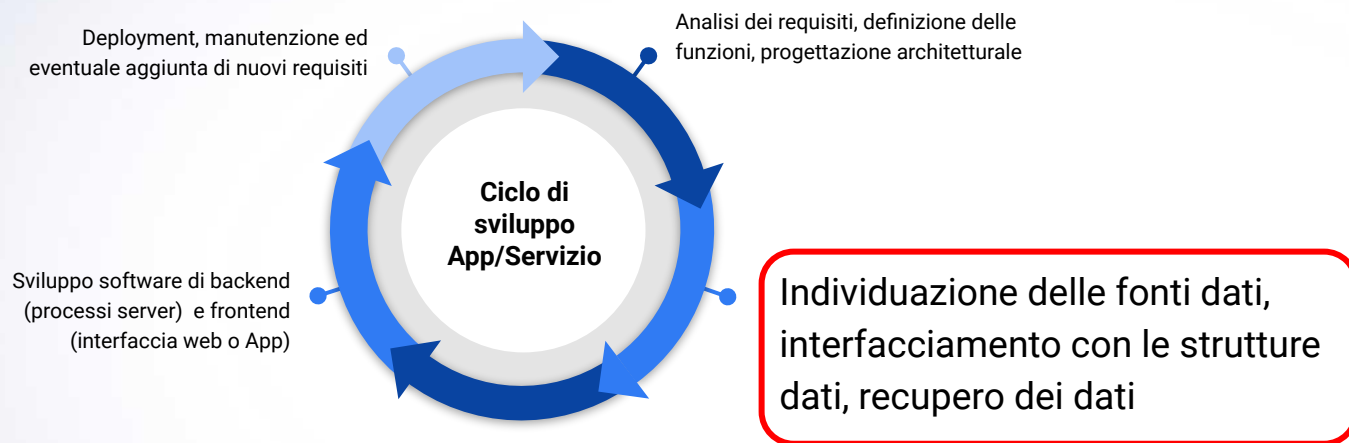
Sviluppare una App/Servizio con Open Data

Sviluppare un'applicazione o un servizio che utilizza gli open non è sostanzialmente diverso da qualsiasi altra applicazione



Sviluppare una App/Servizio con Open Data

Dando per scontati gli altri aspetti, focalizziamoci sulle problematiche relative alle fonti dati, alcune open, altre no...



- Come trovo le fonti dati che mi servono ?
- Come le rendo compatibili con il mio contesto ?
- Come accedo e come le aggiornano ?

Iniziamo a studiare il processo da un esempio

Supponiamo di voler creare un'applicazione che permetta di navigare tutti i dati demografici italiani, offrendo un'interfaccia comune e permettendo di rispondere a domande del tipo:

- Quanti sono i residenti a Treviso con età maggiore di 18 anni e sposati ?
- Quanti sono i laureati in matematica in Veneto non ancora in pensione ?
- Quanti residenti stranieri erano registrati in Piemonte nell'anno 2014 ?

Come troviamo i dati ?
Come li gestiamo ?
Come li raccogliamo ?



Trovare gli Open Data: che strumenti ?

Per "trovare" i dataset da utilizzare si possono usare portali specializzati che "catalogano" fonti diverse di Open Data

<http://www.opendataitalia.it/>

<http://dati.istat.it/>

<https://dati.veneto.it/>

<https://opendatahubs.eu/>

Ma cosa dobbiamo cercare ? Come ?

Siamo sicuri di aver trovato tutto ?

Cosa succede se delle fonti dati spariscono ?

Come gestisco delle nuove fonti dati



Criticità con la ricerca delle fonti dati

Le fonti dati sono molte (bene) e quasi tutte sono dotate di metadati (ottimo), ma:

- Esistono cataloghi ampi, ma nessuno globale, non esiste cioè un "Google" degli Open Data
- In realtà non esiste neppure un formato comune per i metadati (anche se esistono ontologie di dominio, ad esempio CIDOC-CRM per il cultural heritage, DCAT-AP_IT per la PA italiana)
- Non sono diffusi i protocolli di interrogazione dei repository per una ricerca "machine 2 machine"



Gestire gli Open Data: che formati ?

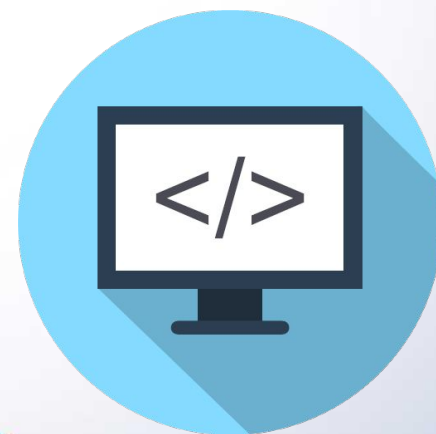
Gli Open Data sono in formati "standard", come CSV, XLS, XML, JSON, ma cosa significa effettivamente essere standard ?

esempio:

https://dati.veneto.it/opendata/Matrimoni_suddivisi_per_Et_delle_Spose_Anno_2018

In pratica lo "standard" si limita alle informazioni atomiche (e forse nemmeno quelle). Non è poco, ma non basta per automatizzare l'integrazione.

In teoria RDF / SPARQL potrebbero aiutare, ma non sono poi così diffusi e omogenei



Criticità nella gestione dei formati

Il problema è che per attuare l'integrazione in modo automatico servono semantiche strutturate che non siano mediate dall'uomo (che le deve interpretare)

- Ontologie chiare ed universali (ma è un obiettivo possibile ?)
- Maggiore utilizzo di strumenti descrittivi avanzati (RDF, ma non solo, sempre abbinato ad ontologie chiare)
- Probabilmente anche una serie di datatype compositi



Raccogliere e aggiornare gli Open Data

Come fare per scaricare e tenere aggiornati i dati da parte del nostro sistema di backend ?

Per questo ci sono alcune soluzioni, per esempio Regione Veneto utilizza una URL permanente (ODATA)

esempio: https://dati.veneto.it//opendata/elenco_strutture_ricettive_del_veneto

Ma non tutti lo fanno... e allora ?



Criticità nel recupero dei dati

Anche il recupero dei dati non è sempre automatizzabile

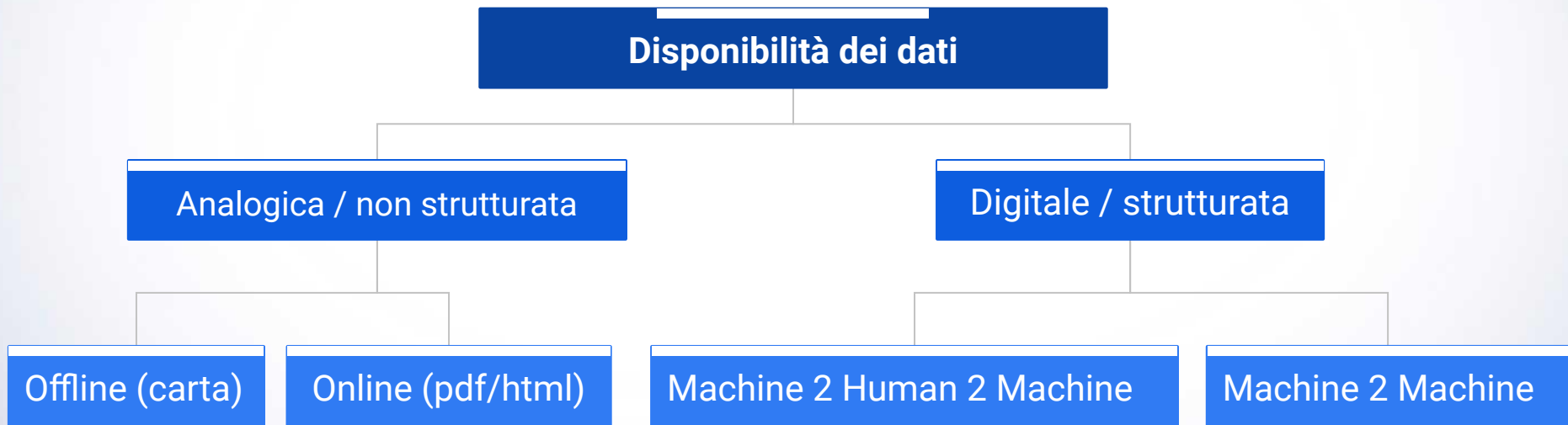
- In molti casi i dati sono disponibili da una pagina web come download e non da una URL permanente
- L'aggiornamento spesso non è garantito: i dati sono forniti così come sono, non come un oggetto vivo, a volte sono addirittura da esportare
- Anche quando machine-2-machine i protocolli sono abbastanza vari (es. ISTAT)



Open Data: fruibilità formale o di fatto ?

Gli Open Data per massimizzare la loro utilità devono essere fruibili cioè di facile accesso, comprensibili ed utilizzabili

Ma lo sono allo stato attuale ? Ed in che termini ?



Quali sviluppi per favorire la fruibilità M2M ?

- Allargare di molto i tipi base, evitando scelte individuali da parte dei singoli fornitori (stelle per le strutture recettive, fasce di età, etc.)
- Trovare metodi per la definizione di meta-ontologie, possibilmente universali e trasversali ed aggiornabili nel tempo
- Definire dei sistemi di ricerca delle fonti dati universali, preferibilmente riferendosi alle ontologie generali e alle funzionalità specifiche dei dati, anche definendo assi standard (spaziali, temporali, etc.)
- Definire protocolli di accesso e trasferimento specifici (un http per i dati ?)

Grazie per l'attenzione...

Domande o curiosità ?

albarelli@unive.it



Interreg
Mediterranean

