

Dati.Culturaltalia: la sezione dedicata ai dati aperti

Sara Di Giorgio Responsabile gruppo tecnico

> Roma 23 maggio 2013 III Convegno SITAR



Vi parlerò di:



- Definizione di Open Data
- Che cosa sono i Linked Open Data (LOD)
- Le licenze per la diffusione dei dati
- Dati.Culturaitalia.it
- La licenza di Culturaltalia in CCO
- I vantaggi dei LOD



Definizione di Open Data



«un contenuto o un dato si **definisce aperto se chiunque è in grado di utilizzarlo, riutilizzarlo e ridistribuirlo**, soggetto, al massimo, alla richiesta di attribuzione e condivisione allo stesso modo» *Open Knowledge Foundation*.

- il dato deve essere accessibile, preferibilmente via Internet e in un formato modificabile;
- deve essere libero da vincoli tecnologici che ne limitino di fatto la più ampia diffusione;
- eventuali vincoli legali non devono pregiudicare la possibilità di distribuzione e riuso.



Open by default



A partire dal 18 marzo 2013, scadenza dei novanta giorni previsti dalla Legge, dati e documenti pubblicati online dalle amministrazioni titolari - senza una esplicita licenza d'uso che ne definisca le possibilità e i limiti di riutilizzo – sono da intendersi come dati aperti, quindi dati che possono essere liberamente acquisiti da chiunque e riutilizzabili anche per fini commerciali.

La norma è il cosiddetto **Decreto Crescita 2.0** (

http://www.altalex.com/index.php?idnot=59517). Decreto Legge 18 ottobre 2012, n. 179 recante "Ulteriori misure urgenti per la crescita del Paese" (c.d. Decreto Crescita 2.0), coordinato con la legge di conversione 17 dicembre 2012, n. 221 e pubblicato in Gazzetta Ufficiale 18 dicembre 2012, n. 294.



I principi dei LOD



La disponibilità e il rilascio di Open Data è un patrimonio prezioso per la società civile e per le imprese, ma affinché si possa valorizzare del tutto l'informazione, **l'apertura da sola non basta**.

Linked Open Data mostrano interessanti opportunità per superare le limitazioni dei modelli Open Data. Se gli Open Data abbattono le barriere culturali, legali ed economiche al riuso, il movimento Linked Data si concentra piuttosto 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.



La classificazione delle 5 stelle del W3C



- 1. i propri dati siano 'aperti' disponibili sul web (in qualsiasi formato jpg, doc,);
- 2. il materiale messo sul web sia disponibile come dato strutturato (per esempio, in excel anziché come scansione dell'immagine di una tabella);
- 3. siano stati scelti formati non proprietari (per esempio, in csv invece che excel);
- 4. siano stati utilizzati URL per identificare gli oggetti, in modo gli utenti possano puntare a questi oggetti;
- 5. i propri dati siano stati collegati a dati prodotti da altri in modo da definire un contesto.





Cosa sono i Linked Open Data



I LOD sono dati pubblicati sul web in modalità leggibile, interpretabile da una macchina il cui significato è esplicitamente definito grazie allo standard RDF Resurce Description Framework



soggetto - predicato - oggetto

Dante ha scritto La Divina Commedia

Dante ha scritto La Divina Commedia

http://dbpedia.org/page/Dante Alighieri

1

http://dbpedia.org/ontology/writer



http://dbpedia.org/page/Divine_Comedy



http://dbpedia.org/ontology/movement

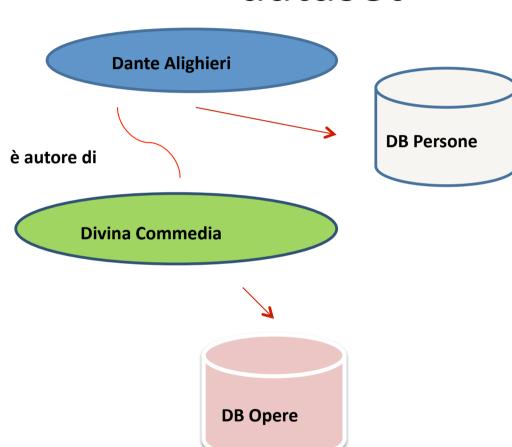
http://dbpedia.org/page/ Dolce Stil Novo





Linked Open Data: i dataset





Ciascuna risorsa che compone la tripla può appartenere a un diverso dataset disponibile sul web. Il meccanismo consente di collegare i dati presenti in una fonte dati (link interni) con dati presenti in una fonte diversa (link esterni).

Web di dati





Linking esterno

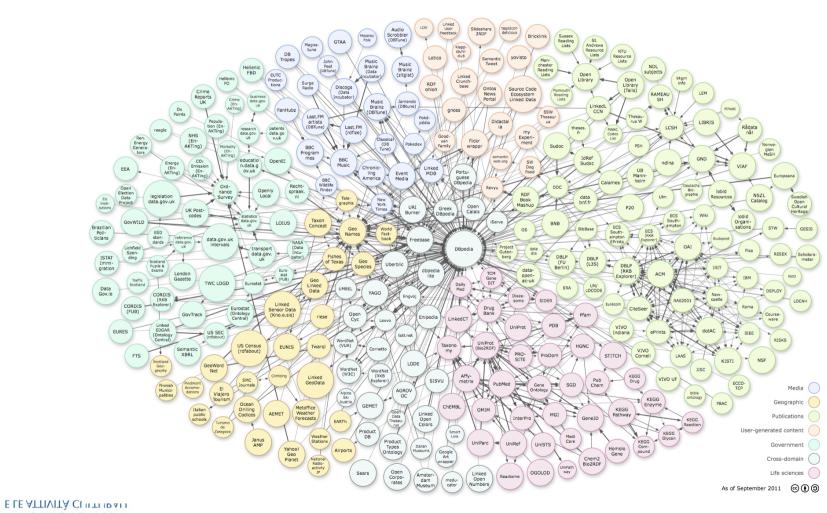


- il contenuto informativo locale viene legato ad altre informazioni, che possono essere altri dataset prodotti dalla stessa amministrazione, oppure insiemi di dati già presenti nel Web dei Dati
- Il linking esterno consiste nella produzione di triple RDF in cui il soggetto e l'oggetto sono entità appartenenti a dataset diversi (interni o esterni all'amministrazione). In pratica, si tratta di allineare entità di diversi dataset.
- Per esempio, nel caso si pubblichino dati che fanno riferimento a entità geografiche, è certamente utile aggiungere un collegamento alla stessa entità geografica presente in GeoNames
- L'operazione di linking, tuttavia, non è sempre banale come può sembrare. Per aiutare l'utente finale in questo processo, si possono utilizzare strumenti specifici che consentono di creare tali collegamenti esterni in maniera agevole.



W3C Linking Data Cloud





Diffusione dei dati: le licenze



L'aspetto relativo alle licenze dei dati è cruciale per quanto riguarda l'uso che gli utenti possono effettivamente fare dei dati, i vincoli di copyright da applicare a lavori derivati da quei dati, il mantenimento della paternità sui dati, ecc.

Le licenze più diffuse **sono quelle basate su Creative Commons** (CC). CC è un'organizzazione non a fini di lucro che nasce con l'intenzione di armonizzare l'articolato mondo del diritto d'autore (in Italia regolato dalla legge n. 633 del 22 aprile 1941) e del copyright. Nel 2002, CC ha pubblicato un primo insieme di licenze che si sono affermate *come standard de facto* a livello internazionale.

Diffusione dei dati: le licenze



LICENZA CREATIVE COMMONS ZERO (CCO)

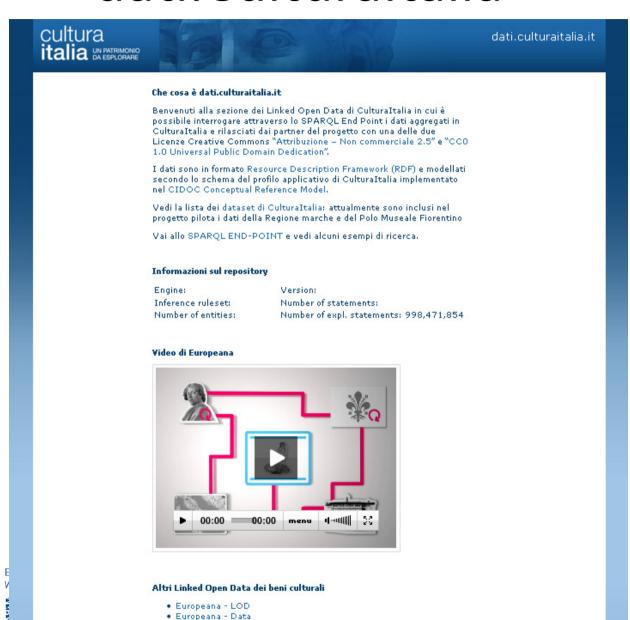
La Creative Commons Zero esprime "la più ampia e libera utilizzazione gratuita, anche per fini commerciali e con finalità di lucro".

LICENZA CREATIVE COMMONS ATTRIBUZIONE (CC-BY)

Un'alternativa alla licenza CCO è la licenza Creative Commons "Attribuzione o equivalente". Questa permette al soggetto utilizzatore di riprodurre, distribuire, comunicare, esporre, rappresentare, nonché di modificare e usare un insieme di dati anche a fini commerciali con il solo obbligo di attribuire la paternità dell'opera. In generale, questa licenza è adottabile per le banche dati che risultano chiaramente tutelate dal diritto d'autore e/o dal diritto sui generis.

La licenza CC-BY può essere estesa mediante alcuni attributi, quali:

- 1. Share Alike (SA): obbliga i lavori derivati a essere licenziati con la stessa licenza del lavoro originale
- 2. Non Commercial (NC): consente la copia, la distribuzione e l'uso del lavoro (o dati) solo per scopi non commerciali
- 3. No Derivative Works (ND): consente la copia, distribuzione e l'uso del lavoro, impedendo la creazione di lavori derivati





Il progetto pilota dati.culturaitalia.it contiene alcuni set di open data aggregati in Culturaltalia e rilasciati dai partner del progetto con una delle due Licenze Creative Commons Share Alike 2.5 e Public Domain CCO.



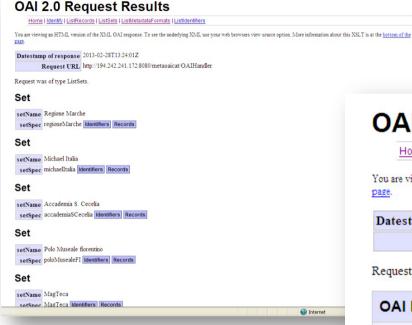
- disponibili i primi dataset di test (Polo museale fiorentino, catalogo opere d'arte della Regione Marche) in CCO successivamente anche altri dataset in CCO e CC-BY.
- SPARQL endpoint che rende interrogabili i dati in formato CIDOC-CRM attraverso:
 - 1) un'interfaccia di interrogazione SPARQL
 - 2) un'interfaccia di interrogazione iSPARQL;
 - 3) una ricerca di testo.
- linking dei dati in formato CIDOC con gli authority file del VIAF (Viartual International Authority file: www.viaf.org) e con i nomi geografici di GeoNames (www.geonames.org/)



- Il nostro OAI Provider rende disponibili metadati in formato XML e RDF strutturati secondo vari schemi:
 - oai-dc (xml): schema OAI-PMH adottato dall'Open Archives Initiative Protocol for Metadata Harvesting
 - pico (xml): PICO Application Profile, il profilo applicativo di Culturaltalia
 - edm (rdf): Europeana Data Model, adottato dal portale Europeana
 - cidoc (rdf): CIDOC Conceptual Reference Model nell'implementazione
 Erlangen CRM / OWL







OCLC ONLINE COMPUTER LIBRARY CENTER

OAICat

An Open Source OAI-PMH v2 Repository Implementation

E LE ATTIVITÀ CULTURALI MINISTERO PER I BENI



OAI 2.0 Request Results

Home | Identify | ListRecords | ListSets | ListMetadataFormats | ListIdentifiers

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view source option. page.

Datestamp of response 2013-02-28T13:21:05Z

Request URL http://194.242.241.172:8080/metaoaicat/OAIHandler

Request was of type ListRecords.

OAI Record: oai:culturaitalia.it:ANSC00002595

OAI Record Header

OAI Identifier oai:culturaitalia.it:ANSC00002595 pico edm cidoc_crm oai_dc formats Datestamp 2012-02-01T18:04:53Z

setSpec accademiaSCecelia Identifiers Records

PICO Metadata (pico)

<pico:record>

<dc:identifier> ANSC.00001.00045.00005</dc:identifier>

<dc:title xml:lang="it" >Melie</dc:title>

<dc:type xsi:type="dcterms:DCMIType" >Sound</dc:type>

<dc:type xml:lang="it" >canto narrativo</dc:type>

<dcterms:isPartOf xml:lang="it" >name=raccolta;value=028, a cura di Sergio Liberovici</dcterms:isPartOf>

Diffusione dei dati: la licenza di Culturaltalia





Ministero per i beni e le attività culturali
ISTITUTO CENTRALE PER IL CATALOGO UNICO DELLE BIBLIO TECHEITA HANE
EPER LEINFORMAZIONI BIBLIOGRAFICHE
00185 Roma - Vale del Casto Peterio 105 - Tel 10649210425 - Fec 064959302
PI 00916901004 - CF .006860381
email: to outobenioutralit

LICENZA CULTURA ITALIA

Premesse

<u>La Licenza è un accordo quadro</u> che serve a disciplinare l'eventuale conferimento di Anteprime e Metadati da parte del Fornitore di contenuti per le iniziative del MIBAC: non comporta pertanto nessun impegno immediato, ma la Licenza produce effetto soltanto nel momento in cui il Fornitore di contenuti decide – autonomamente e senza obblighi o termini, e senza che siano

richieste formalità particolari - di mettere a disposizione determinati contenuti a sua totale discrezione. Il MIBAC avrà la facoltà di accettare o meno, in tutto o in parte, il conferimento in licenza delle Anteprime e dei Metadati, restando di esclusiva competenza e discrezionalità del MIBAC la decisione circa l'uso, totale o parziale, o il non uso dei contenuti conferiti, di cui il MIBAC s'impegna a mantenere un elenco aggiornato, mettendolo a disposizione del Fornitore di contenuti.

Con la Licenza il Fomitore di contenuti concede l'utilizzo di Anteprime e Metadati secondo i termini e le condizioni contenute nel Contratto per lo Scambio di Dati di Europeana (European Data Exchange Agreement), che si allega al presente accordo nella traduzione italiana. In particolare, nell'esecuzione del Contratto per lo Scambio di Dati tra le parti, il MIBAC sarà soggetto agli stessi obblighi e godrà degli stessi diritti di Europeana nei confronti del Fomitore di contenuti. Di conseguenza, per ciò che concerne i Metadati, il Fomitore li concederà secondo i termini e le condizioni della licenza CC0 1.0 Universal Public Domain Dedication i. Per ciò che riguarda le Anteprime, il Fomitore sarà tenuto a specificare i termini e le condizioni secondo cui il MIBAC è autorizzato a permetterne l'uso ai visitatori di Cultura Italia.

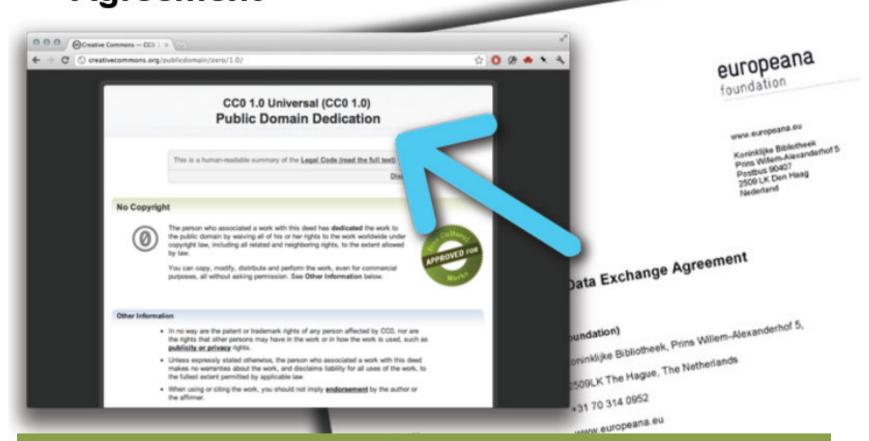
Convenendo che le premesse suddette debbano considerarsi parte integrante della Licenza, il MIBAC e il Fomitore di contenuti (d'ora in avanti "Parti") definiscono il contenuto della propria intesa nei termini seguenti:

MINISTERO PER I BENI E LE ATTIVITÀ CL⁴ITI IP A I I



cultura italia

New Europeana Data Exchange Agreement



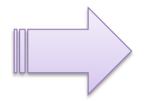
Europeana chiede agli aggregatori di ottenere il permesso dai loro fornitori di dati, compresa una copia del DEA

La DEA distingue



Metadati

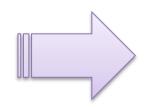
Informazioni testuali (inclusi i collegamenti ipertestuali) utili per individuare, interpretare e/o gestire il Contenuto





Anteprime

Una rappresentazione visiva e/o sonora dei Contenuti in dimensioni ridotte, sotto forma di una o più immagini, file di testo, file audio e/o file di immagini in movimento.

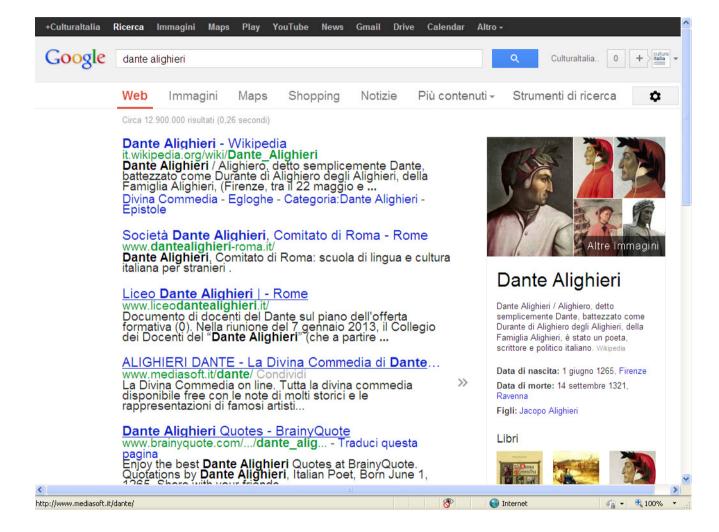








I vantaggi dei LOD





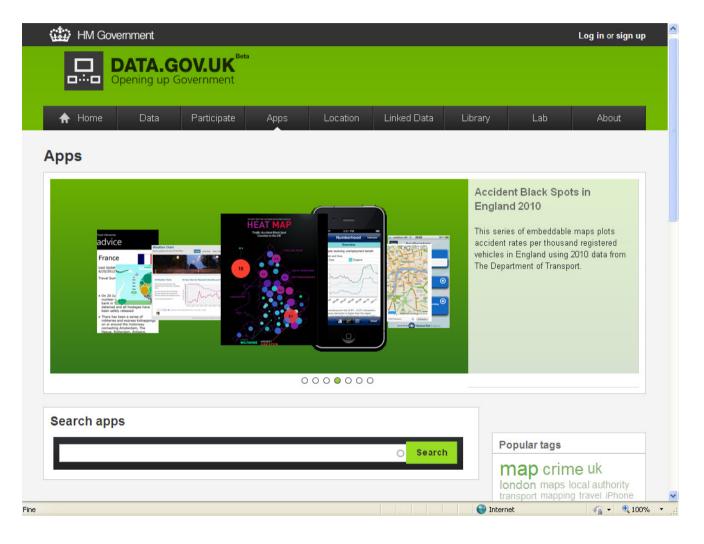


Integrare i dati nei siti web



I vantaggi dei LOD





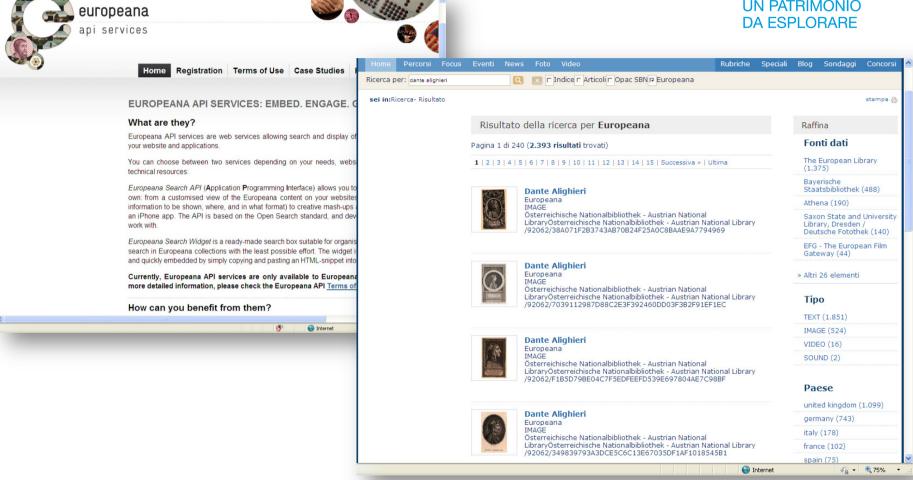






I vantaggi dei LOD











Grazie! sara.digiorgio@beniculturali.it

