



L'esperienza di linked heritage nell'ambito del linking heritage information

Gordon McKenna

Introduzione

Questo lavoro passerà in rassegna l'esperienza del progetto Linked Heritage, finanziato dall'Unione Europea, nell'ambito dei linked data. Verranno discussi:

- Il progetto nel suo contesto;
- Il Work package 2 - Linking Cultural Heritage Information;
- I risultati della ricerca nell'ambito dell'uso dei linked data nel settore dei beni culturali;
- Le prospettive future del progetto.

Il progetto Linked Heritage

Il progetto Linked Heritage è co-finanziato dal ICT Policy Support Programme come parte del Competitiveness and Innovation Framework Programme, ha avuto inizio a Aprile 2011 e dura trenta mesi. I suoi obiettivi sono tre:

- Contribuire a Europeana con grandi quantità di nuovi contenuti sia da settori pubblici che privati;
- Dimostrare il miglioramento della qualità del contenuto in termini di ricchezza dei metadati, potenziale di riuso e unicità;
- Dimostrare la possibilità di potenziamento della ricerca, del recupero e dell'uso dei contenuti di Europeana.

L'obiettivo di Linked Heritage è di facilitare e distribuire su ampia scala gli aggiornamenti nel lungo periodo di Europeana e i suoi servizi. Si concentra su questioni relative a:

- Terminologie descrittive non-standard;
- La mancanza di contenuti del settore privato e del 20° secolo;
- La conservazione di modelli di metadati complessi nell'ambito dello schema di metadati di Europeana.

Tra i partner del progetto ci sono tutti i più importanti gruppi provenienti da venti paesi membri dell'Unione Europea, insieme con Israele e Russia. Ci sono ministeri e agenzie governative, fornitori di contenuti, aggregatori, centri di ricerca di punta, editori e SME.¹ Gli obiettivi del progetto sono:

¹Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche (IT); Università Degli Studi di Padova (IT), Consiglio Nazionale delle Ricerche (IT), Ministère de la Culture et de la Communication (FR), Eesti Vabariigi Kultuuriministeerium (ER), Hellenic Ministry of Culture (GR), National Technical University of Athens (GR), University of Patras (GR), Collections Trust LBG (UK), An Chomhairle Leabharlanna Ireland (IE), Pintail Ltd (IE), Fundacio Privada I2CAT, Internet i Innovacio Digital A Catalunya (SP), Philipps Universitaet Marburg (GW), Stiftung Preussischer Kulturbesitz (GW), Central Library of the Bulgarian Academy of Sciences (BU), Javni Zavod Republike Slovenije za Varstvo Kulturne Dediscine (Slovenia), The Cyprus Research and Educational Foundation (CY), Stowarzyszenie Miedzynarodowe Centrum Zarzadzania Informacja (PL), Riksarkivet (SW), MEDRA S.R.L. (IT), Gottfried Wilhelm Leibniz Universitaet Hannover (GW), Editeur Limi-

- Contribuire a Europeana con nuovi contenuti provenienti sia da settori pubblici che privati;
- Predisporre il miglioramento della qualità dei contenuti di Europeana nuovi e già esistenti in termini di ricchezza dei suoi metadati, del suo potenziale di riuso e della sua unicità;
- Dimostrare la possibilità di potenziamento della ricerca, del recupero e dell'uso dei contenuti di Europeana sia nell'ambito del portale di Europeana che per mezzo di terzi tramite Europeana API.

E questo nel seguente modo:

- Mettendo insieme gruppi di partecipanti rappresentativi (fornitori di contenuti, aggregatori, ministeri e enti con poteri decisionali, addetti ai lavori, società private e associazioni);
- Consultando, creando consensi, facendo rete e condividendo prospettive e priorità;
- Individuando e promuovendo le buone pratiche (standard appropriati, specifiche, raccomandazioni utili per contribuire, acquisire e arricchire l'ingestione e l'arricchimento dei contenuti di Europeana;

ted (UK), MVB (Marketing und Verlagsservice des Buchhandels) (GW), Országos Széchenyi Könyvtár (HU), Koninklijke Musea voor Kunst en Geschiedenis (BE), Istituto Umeni - Divadelního Ustavu (Czech Republic), Instituto Superior Técnico (PO), Valsts Agentūra Kultūras Informācijas Sistēmas (Latvia), PACKED (Platform voor de Archivering en Conservering van Audiovisuele Kunsten) (BE), CORDIA (Slovakia), Università Degli Studi di Roma La Sapienza (IT), C.T.F.R. SRL (IT), Departament de Cultura i Mitjans de Comunicació (SP), Promoter di Masi Pietro & C S.N.C. (IT), Université de Savoie (FR), Association Dedale (FR), UMA Information Technology (AU), Digital Heritage LBG (UK).

- Completando su larga scala (compresa la necessaria integrazione della tecnologia nel rispetto delle norme europee) e validando gli standard e le specifiche di buone pratiche individuati. Ciò consentirà di contribuire a Europeana con tre milioni di nuovo oggetti;
- Preparando un demo su come le specifiche perfezionate devono essere applicate e su come attuare l'arricchimento dei contenuti di Europeana.
- Facendo formazione e disseminazione per costruire potenzialità e consapevolezza nel settore dei beni culturali, in particolare nell'uso di produzioni tecniche di Linked Heritage, ma anche nell'ambito delle tecnologie di Europeana.

Il progetto si divide in sette pacchetti di lavoro:

WP 1 – Project management and coordination Riguarda la gestione di base del progetto Linked Heritage, vale a dire il monitoraggio dei progressi e la gestione delle relazioni con la Commissione. Gestisce anche la creazione e la coordinazione dei gruppi di lavoro nazionali e tematici.

WP 2 – Linking cultural heritage information Esamina l'uso potenziale dei linked data nel settore culturale (vedi sezione seguente).

WP 3 – Terminology Si concentra sul possibile uso delle terminologie nell'ambito del progetto e in un contesto culturale più ampio.

WP 4 – Public Private Partnership Esplora gli standard in uso nei settori culturali commerciali non relativi ai beni culturali e le possibilità di integrare quegli standard con questi settori, in particolare modo con Europeana.

WP 5 – Technical Integration Predisporre gli strumenti e i requisiti tecnici del progetto.

WP 6 – Coordination of Content Gestisce l'accesso a Europeana da parte dei partner di Linked Heritage.

WP 7 – Dissemination & Training Diffonde la conoscenza del lavoro svolto nell'ambito del progetto in tutta la comunità e fornisce gli strumenti di apprendimento per far sì che la comunità possa utilizzarne i risultati.

Work Package 2 - Linking Cultural Heritage Information

La presente ricerca è parte dei risultati di questo pacchetto di lavoro, i cui obiettivi sono:

- Esplorare lo stato dell'arte dei linked data, le applicazioni e le potenzialità;
- Individuare i modelli, i processi e le tecnologie più adeguati per l'utilizzo di archivi di informazioni sui beni culturali come i linked data;
- Considerare come le pratiche relative ai linked data possano essere applicate agli archivi di informazioni sui beni culturali, arricchirli e far sì che questi vengano allineati ad altri depositi ed applicazioni di linked data;
- Esplorare lo stato dell'arte sugli identificatori persistenti (sia per quanto riguarda gli standard che i tool di gestione);

- Individuare l'approccio più adeguato per l'identificazione persistente, vale a dire un unico standard oppure un set di standard diversi;
- Progettare un modello di fattibilità e realizzare un demo di infrastruttura flessibile, ponderabile, sicura e affidabile per una rete di archivi di informazioni sui beni culturali "potenziati con linked data";
- Esplorare lo stato dell'arte dei modelli di metadati culturali ed in particolare la loro interoperabilità tra biblioteche, musei, archivi, editori, produttori di contenuti e i modelli di Europea-na ESE (European Semantic Element) e EDM (European Data Model);
- Fare una sintesi dei benefici potenziali che i metadati arricchiti dei beni culturali possono apportare a Europea-na e ai servizi che ne fanno uso.

I linked data nel settore dei beni culturali

I partner

Tra gli obiettivi del pacchetto di lavoro c'è anche un'indagine sui partner e sui fornitori di Linked Heritage. Tra l'altro, sono state esplorate la loro conoscenza dei linked data e la loro esperienza nell'uso dei linked data.

Informazioni su coloro che hanno partecipato all'indagine²

Le cifre rappresentate in tabella 1 mostrano come il contenuto veicolato in Europea tramite il progetto Linked Heritage copra tutti i domini in ambito culturale compresi gli aggregatori. C'è anche comunque un numero di risposte significative provenienti da organizzazioni che non contribuiscono in termini di contenuto e che quindi non comparirebbero nella sezione della ricerca relativa ai metadati.

Tipologia	Risposte	%
Museum	4	10.3
Library	5	12.8
Archive	4	10.3
Sound archive	1	2.6
Publisher	0	0
Aggregator	10	25.6
Other	23	59.0

Tabella 1: Tipologia delle organizzazioni partecipanti.

Circa il 60% delle risposte non rientrano nella tipologia "standard". Quindi è sembrato utile elencare quali sono state le risposte alla domanda: "Se avete contrassegnato 'Altro' indicate il tipo di organizzazione":

- Intermediario tra provider e progetto Linked Heritage;
- Gruppo di musei;
- Organizzazione governativa per la tutela dei beni culturali fissi e dei beni culturali mobili e viventi ad essi associati;

²Le cifre che seguono si riferiscono a quelle organizzazioni che hanno risposto all'indagine.

- Libri nazionali a stampa;
- Partner tecnico;
- Università;
- Agenzia di registrazione DOI [Digital Object Identifier];
- Centro di ricerca e innovazione;
- Ministero della Cultura;
- Società nel campo del beni culturali;
- Istituto per la ricerca scientifica con collezioni museali;
- Società di servizi di gestione e qualità;
- Punto di contatto nazionale;
- SME – consulenze;
- Emittente pubblica e archivio di media (video, audio e fotografico);
- Ente normatore;
- Documentazione teatrale (fotografie);
- Organizzazione pubblica;
- Amministrazione pubblica regionale responsabile per i sistemi informativi dei beni culturali;
- Fornitore di tecnologia;
- Associazione e centro d'informazione;
- Agenzia culturale.

Paesi

Country	Risposte	Country	Risposte
Austria	1	Ireland	2
Belgium	4	Israel	1
Bulgaria	1	Italy	6
Cyprus	1	Poland	1
Czech Republic	1	Russian Federation	1
Estonia	1	Slovakia	1
France	3	Spain	2
Germany	4	Sweden	1
Greece	3	United Kingdom	2
Hungary	1		

Tabella 2: Numero di risposte per paese partecipante.

Ovviamente i numeri in tabella 2 riflettono il numero di partner coinvolti nel progetto, ma c'è una copertura in tutta Europa, con un paio di partner al di fuori dell'Unione Europea. Nel complesso, l'informazione ottenuta induce gli autori alla conclusione che il campione è abbastanza rappresentativo del settore.

Linked data

Consapevolezza

Il no rappresentato in tabella 3 nella pagina seguente ha destato la sorpresa degli autori, ma dimostra che c'è un 'mercato' per l'informazione e gli strumenti relativi ai linked data!

Response	Risposte	%
Yes	30	75.0
No	10	25.0

Tabella 3: "Lei o la sua organizzazione ha familiarità con il concetto di linked data?"

Uso

A coloro che hanno risposto "Sì" alla domanda in tabella 4 è stato chiesto di fornire dettagli su la/le fonte/fonti di linked data usata/usate e perché. Le fonti usate sono: DBpedia (4); GeoNames (3); Freebase (1); IPTC (1); Thesauri in SKOS (1). Solo due risposte contenevano informazioni sul perché una certa fonte era stata usata: DBpedia (fonte interessante d'informazione), GeoNames (per la disambiguazione dei nomi di luogo)

Response	Risposte	%
Yes	7	17.5
No	33	82.5

Tabella 4: "Lei o la sua organizzazione ha avuto esperienze nell'uso di linked data relativi alle collezioni?"

Pubblicazione

A coloro che hanno risposto "Sì" alla domanda in tabella 5 a fronte è stato chiesto di fornire dettagli. Li hanno forniti in tre: <http://data.kunstkamera.ru/sparql> and <http://data.kunstkamera.ru/>; Record bibliografici completi da OPAC e Digital Library (OSZKDK) in DC. Autorità dei nomi in FOAF; Tesoro in SKOS. I dettagli e le

Response	Risposte	%
Yes	4	10.0
No	36	90.0

Tabella 5: Alla domanda: " Lei o la sua organizzazione ha avuto esperienze nella pubblicazione di linked data relativi alle vostre collezioni?"

URL sono: http://nektar.oszk.hu/wiki/Semantic_web. A sostegno di RDF nella Digital Library (OSZKDK). Il Dipartimento francese per gli Archivi ha pubblicato il proprio tesoro in SKOS in una prospettiva di riuso con i linked data. Un progetto nazionale in corso metterà insieme tutti i vocabolari attualmente in uso nel ministero con l'intento di creare una rete di concetti da collegare ad altre iniziative come ad esempio RAMEAU in SKOS.

Progetti e iniziative sui linked data

Response	Risposte	%
Yes	15	37.5
No	25	62.5

Tabella 6: Alla domanda "Lei o la sua organizzazione è a conoscenza di progetti o iniziative nel suo paese nel campo dei beni culturali?"

A coloro che hanno risposto "Sì" è stato chiesto di fornire dettagli. La tabella 7 nella pagina successiva riporta le risposte in ordine alfabetico di paese.³

³Dettagli delle risposte in tabella 7 nella pagina seguente. RAMEAU: <http://www.cs.vu.nl/STITCH/rameau/index-fr.html>, ISIDORE: <http://rechercheisidore.fr>,

Country	Progetto o iniziativa
France	RAMEAU ISIDORE Pactols BABEL COLLECTIONS PALISSY EROS PATRIMOINE LOT WIKIMEDIA COMMONS FRANCE:
Germany	Linked data service of the German National Library "Several initiatives throughout the country"
Israel	Vocabularies of the Israel Museum Jerusalem (SKOS)
Italy	Linked Open Data Italia SPAR ontologies Datagov.it LinkedOpenCamera Spaghetti Open Data
Russia	Open Kunstammer
Sweden	LIBRIS
Spain	Open Data Gencat Euskadi Patmapa Cantabria's Cultural Heritage Ontology
United Kingdom	Various government data sets

Tabella 7

Europeana Open Data Agreement

Response	Risposte	%
Yes	11	29.7
Not sure	20	54.1
No	6	16.2

Tabella 8: A coloro che hanno risposto abbiamo domandato: “La nuova licenza di Europeana prevede che i fornitori debbano convenire che i metadati da loro forniti siano pubblicati come Linked Open Data, vale a dire che ne viene consentito qualsiasi uso, anche di tipo commerciale. La sua organizzazione è d’accordo?”

A coloro che hanno risposto è stato anche chiesto di spiegare le loro

Pactols: <http://www.frantiq.fr/thesaurus-pactols>, BABEL: <http://babel.alienor.org>, COLLECTIONS: http://www.culture.fr/fr/sections/collections/moteur_collections, PALISSY: http://www.culture.gouv.fr/public/mistral/dapapal_fr?ACTION=NOUVEAU&USRNAME=nobody&USRPWD=4%24%2534P, EROS: http://www.c2rmf.fr/pages/page_id18479_u112.htm, PATRIMOINE LOT: <http://www.patrimoine-lot.com>, WIKIMEDIA COMMONS FRANCE: <http://commons.wikimedia.org/wiki/Accueil>, Linked data service of the German National Library: http://www.d-nb.de/eng/hilfe/service/linked_data_service.htm, “Several initiatives throughout the country”, Vocabularies of the Israel Museum Jerusalem that have been migrated to SKOS: <http://www.imj.org.il/imagine/thesaurus/allobject.htm> and <http://www.imj.org.il/imagine/thesaurus/objects/objectTOC.htm>, ItalyLinked Open Data Italia: <http://www.linkedopendata.it/en-home>, SPAR ontologies: <http://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies>, Data.gov.it. Associazione italiana per l’Open Government: <http://www.datagov.it>, <http://www.linkedopencamera.it>, <http://www.spaghettiopendata.org>, RussiaOpen Kunstammer: <http://www.kunstkamera.ru>, Sweden LIBRIS (joint catalogue of the Swedish academic and research libraries): <http://www.kb.libris.se>, SpainOpen Data Gencat: <http://dadesobertes.gencat.cat/en/index.html>, Euskadi: <http://opendata.euskadi.net/w79-home/es/>, Patmapa: <http://patmapa.gencat.cat/>, Cantabria’s Cultural Heritage Ontology: <http://hdl.handle.net/10760/13938>, United Kingdom, Various government data sets: <http://data.gov.uk>.

risposte.

Chi ha risposto "Sì" ha dichiarato (a fianco ad ognuna compaiono i numeri di risposte):

- 1 – Pubblicare sul Web è sinonimo di Open Data;
- 1 – Abbiamo partecipato al progetto ATHENA;
- 1 – I metadati forniti a Europeana sono stati scelti in modo specifico per i linked data ad accesso aperto.

Chi ha risposto "Non so" ha dichiarato:

- 4 – I metadati non sono nostri (è stata una decisione del provider);
- 4 – Ne stiamo discutendo;
- 2 – Ne stiamo discutendo (potrebbero esserci ostacoli di tipo legale);
- 2 – Non è stata una nostra decisione (è stata presa dai vertici);
- 1 – Forniremo dati minimi;
- 1 – Siamo contro un uso di tipo commerciale.

Chi ha risposto "No" ha dichiarato:

- 3 – Siamo contro l'uso di tipo commerciale da parte di terzi;
- 1 – Le politiche nazionali non consentono un uso commerciale
- 1 – Non contribuiamo a Europeana.

The Linking Open Data Cloud

Linking Open Data (LOD) Cloud⁴ (The Cloud) è la rappresentazione più conosciuta dei linked data. Mostra 'pacchetti' di linked data e

⁴<http://linkeddata.org>.

i collegamenti tra i pacchetti. Nel maggio 2007 si presentava come mostrato in figura 1 (con 12 pacchetti).

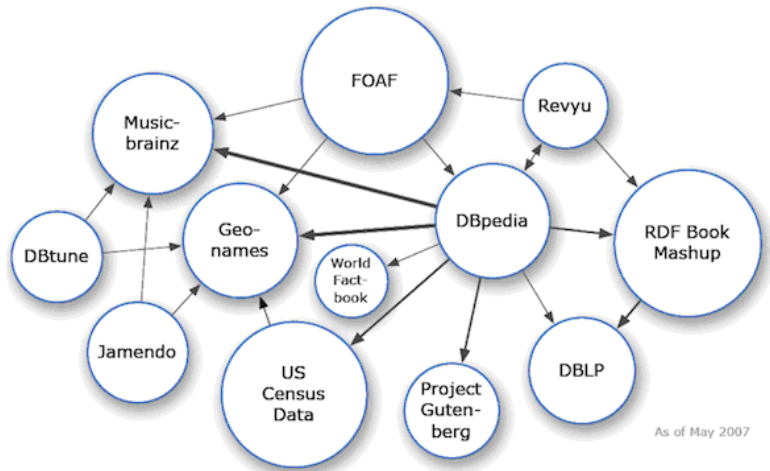


Figura 1: The Cloud in Maggio 2007

La versione a colori che rappresenta il dominio del pacchetto a settembre 2011 si presentava come in figura 2 nella pagina successiva (con 311 pacchetti).

Come si può vedere, “The Cloud” cresce molto velocemente e, nella sua ultima versione, è più difficile distinguere di cosa è composta. Per fortuna “The Cloud” viene gestita usando un wiki tenuto sul sito The Data Hub.⁵ Questo fa parte del progetto Linking Open Data⁶ che è a sua volta parte del Semantic Web Education and Outreach Interest Group (SweoIG)⁷ del W3C e quindi rappresenta una parte

⁵<http://thedatahub.org>, <http://thedatahub.org/group/lodcloud>.

⁶<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>.

⁷<http://www.w3.org/wiki/SweoIG>.

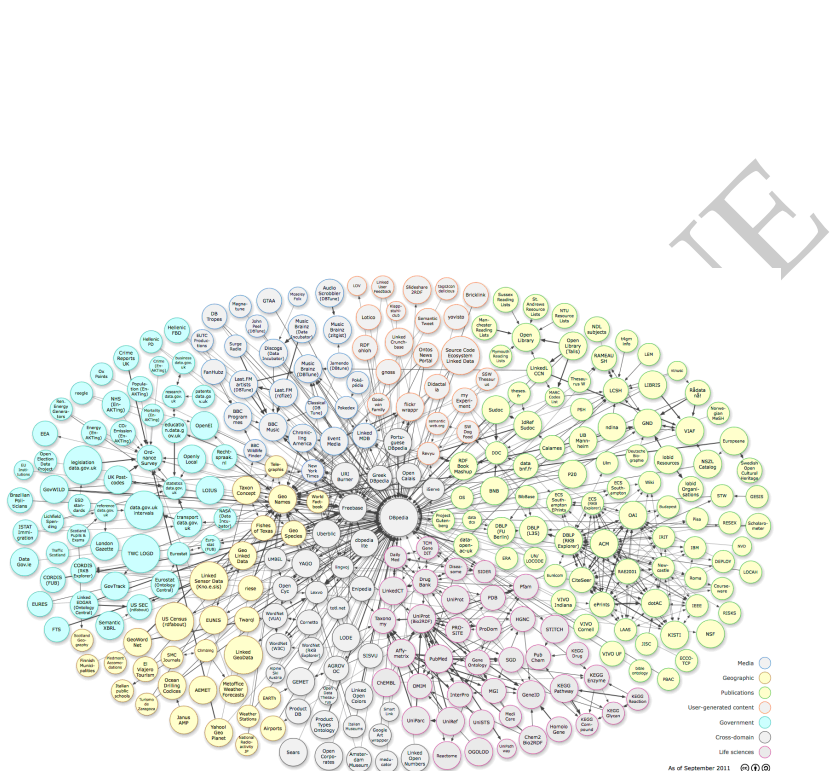


Figura 2: The Cloud in Settembre 2011

As of September 2011

significativa dei linked data disponibili. The Data Hub è un archivio di conoscenza aperta (e non) con informazioni su pacchetti e progetti (compreso il gruppo LOD Cloud). Una volta scelto il gruppo LOD Cloud, l'utente si trova di fronte al primo di una serie di schermate (attualmente sette):

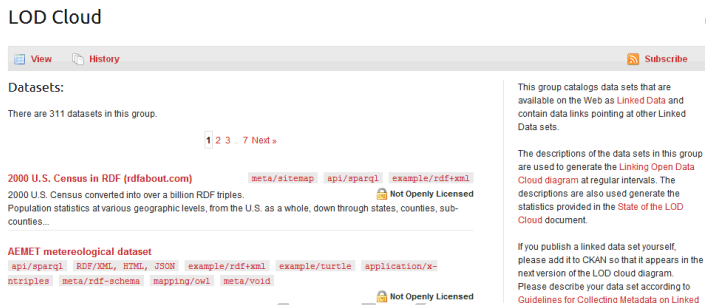


Figura 3: La schermata di ricerca di Data Hub con i risultati

Per ciascun pacchetto la schermata fornisce informazioni su:

- Nome del pacchetto (come link al record completo);
- Descrizione del pacchetto;
- Link alle risorse (compresi gli esempi) disponibili per il pacchetto;
- Lo status IPR del pacchetto.

Per ogni pacchetto è disponibile il record completo, come mostra la figura 3.

Tale schermata contiene ulteriori informazioni per ciascun pacchetto:

- I collegamenti tra i pacchetti (compreso il numero di link);
- Il numero di triple nel pacchetto (un ordine di grandezza)

Amsterdam Museum as Linked Open Data in the Europeana Data Model

View
Edit
History
Subscribe

The Amsterdam Museum dataset describes more than 70.000 cultural heritage objects related to the city of Amsterdam described by the museum.

The metadata was retrieved from an XML Web API of the museum's Adlib collection database and converted to RDF compliant with the Europeana Data Model (EDM). This makes the Amsterdam Museum data the first of its kind to be officially converted and made available in this format.

Resources

Description	Format
SPARQL endpoint	api/sparql
Public Git repository with RDF (browser version)	HTML
Public Git repository (use this path for git clone)	api/git
example: Local view for object "Transom"	example?idf+xml
example: Local view for object "Commemoration plate"	example?idf+xml
SPARQL endpoint UI (in HTML form)	HTML
Download	-

Additional Information

Field	Value
links:dbpedia	43
links:geonames-semantic-web	658
namespace	http://purl.org/collections/nl/am/
shortname	Amsterdam Museum
triples	5000000
vocab-mappings	skos:exactMatch

First time at the Data Hub?

the Data Hub is a catalogue for data.
[Click here to find out more ...](#)

Source
<http://semanticweb.cs.vu.nl/todam/>

Author
 Victor de Boer, Jan Wielemaker, Jacco van Ossenbruggen, Antoine Isaac, Gaus Schreiber

Maintainer
 Victor de Boer

Version
 1.0

Tags

- amsterdam
- country:netherlands
- crossdomain
- cultural
- culturalheritage
- datagovuk
- deref-vocab
- edm
- europeana
- lod
- museum
- no-license-metadata
- no-provenance-metadata
- publications
- published-by-third-party
- rdf

Figura 4: Il record del pacchetto Data Hub

- Ulteriori dettagli sulla situazione IPR del pacchetto (non visibili nella schermata);
- Nei tag:
 - Informazioni sul soggetto;
 - I 'formati' utilizzati.

The Cloud è 'Aperta'?

Potrebbe sembrare una domanda strana, tuttavia, nell'esaminare per la prima volta le informazioni sul sito di The Data Hub è apparso chiaro che una parte significativa di The Cloud non è aperta. Infatti, in The Cloud: **Aperto = in grado di essere riutilizzato commercialmente.**

Se esaminiamo i dati mostrati:

In termini di pacchetti (311)	
IPR Status	%
Open	42.6
Not open	57.4
In termini di triple (ca 38 miliardi)	
IPR Status	%
Open	30.9
Not open	69.1

Tabella 9

Il risultato è alquanto sorprendente in quanto mostra che la maggior parte di The Cloud non è aperta. Una delle ragioni di questa anomalia potrebbe essere che somiglia ad un paesaggio storico con in evidenza i diversi periodi in superficie. In tal caso il presupposto è

che stiamo vedendo diversi pacchetti che sono i primi componenti di The Cloud, nei tempi in cui IPR e possedere una licenza non erano considerate cose importanti. Detto ciò, gli ultimi aggiornamenti riportano ancora pacchetti 'non aperti'. Se si esaminano le licenze in uso nel dettaglio si avranno ulteriori particolari.

Quali licenze IPR vengono usate?

Licenze aperte

Dei 132 pacchetti (ca 11.9 miliardi di triple) con licenze aperte:

Licence type	% by Package	% by Triples
Creative Commons Attribution (CC BY)	28.8	45.8
Creative Commons Attribution Share Alike (CC BY-SA)	18.2	10.2
Open Data Commons Public Domain Dedication and Licence (ODC PDDL)	10.6	0.2
Creative Commons CC Zero (CC0)	9.1	2.9
UK Crown Copyright with data.gov.uk rights	7.6	27.4
Other (Public Domain)	6.8	7.0
Other (Open)	5.3	5.0
Other (Attribution)	3.0	0.4
UK Open Government Licence (OGL)	3.0	0.1
GNU Free Documentation Licence (GNU FDL)	3.0	0.0
Open Database Licence (ODbL)	2.3	0.9
GNU General Public Licence (GNU GPL)	0.8	<0.1
New BSD license and Simplified BSD licence	0.8	<0.1

Tabella 10

Ci si deve aspettare l'uso prevalente di CC BY per una licenza aperta. È una scelta ovvia, come pure CC BY-SA, ODC PDDL e CC0. Quest'ultima è un'opzione relativamente recente ed è la scelta operata da Europeana, e in seconda battuta dai suoi provider, per la pubblicazione dei linked data aperti. È la più permissiva tra le licenze aperte, l'attribuzione non è obbligatoria, è una 'raccomanda-

zione'. Un'iniziativa nazionale che merita di essere citata è quella del Regno Unito. Il governo inglese pubblica una gran quantità di dati usando le proprie licenze open data. Attualmente queste costituiscono oltre il 10% di The Cloud. La Open Government Licence inglese è interoperabile con CC BY.

Licenze non aperte

Dei 178 pacchetti (ca 26.7 miliardi di triple) con licenze non aperte, o privi di informazioni sulla licenza:

Licence type	% by Package	% by Triples
not given	69.1	89.4
None	14.6	0.3
Creative Commons Attribution Non-commercial (CC BY-NC)	7.3	5.8
Other (Not Open)	6.7	<0.1
Creative Commons Attribution (CC BY)	1.1	0.6
Other (Non-Commercial)	0.6	3.9
Creative Commons Attribution Share alike (CC BY-SA)	0.6	<0.1

Tabella 11

La tabella 19 a pagina 32⁸ mostra come per oltre l'80% dei pacchetti e circa il 90% delle triple della parte 'non aperta' di The Cloud non c'è informazione su IPR. È interessante osservare come questa situazione non sembra avere ripercussioni sull'uso di The Cloud e che alcuni tra i pacchetti più recenti non hanno licenze. Per coloro che pubblicano i loro dati in The Cloud con una licenza, ma che non vogliono renderli aperti, ci sono le seguenti due opzioni:

- CC BY-NC;
- La loro licenza 'non-standard' presumibilmente con requisiti speciali.

⁸Si noti che CC BY e CC BY-SA sono aperti ma nei dati sono descritti come non aperti, cosa che abbiamo mantenuto nella tabella.

Quanto è grande The Cloud?

Come già accennato, ci sono circa 38 miliardi di triple in The Cloud. C'è una grande varietà nelle dimensioni. 9 pacchetti (2.89%) hanno oltre un miliardo di triple. Circa un quarto dei pacchetti sono relativamente piccoli con meno di 100.000 triple. Il più piccolo ne ha solo 368. Da ciò si evince che c'è un elemento 'test' di linked data in The Cloud, confermato dal fatto che alcuni pacchetti vengono descritti come 'test'. Il numero medio di triple in un pacchetto è di circa 124 milioni. I dieci più grandi pacchetti con licenza aperta sono:

Package	Number of triples
LinkedGeoData	3.00 billion
UK Legislation	1.90 billion
Linked Sensor Data (Kno.e.sis)	1.73 billion
data.gov.uk Time Intervals	1.00 billion
DBpedia	1.00 billion
Open Library data mirror in the Talis Platform	0.54 billion
The Open Library	0.40 billion
Freebase	0.34 billion
transport.data.gov.uk	0.33 billion
Data Incubator: MusicBrainz	0.18 billion

Tabella 12

LinkedGeoData (licenza CC BY) è un knowledge base di tipo spaziale ricavato dal progetto OpenStreetMap.⁹ Il suo obiettivo è fornire un elemento semantico al Web semantico. Tre pacchetti – UK Legislation, data.gov.uk Time Intervals e transport.data.gov.uk – fanno parte dell'iniziativa del governo inglese tesa a pubblicare i loro dati pubblici ad accesso aperto. Questi sono tutti pubblicati

⁹<http://www.openstreetmap.org>.

con la speciale licenza aperta inglese “UK Crown Copyright with data.gov.uk rights”. Linked Sensor Data (Kno.e.sis) (licenza CC BY) contiene dati con informazioni relative alle stazioni meteorologiche e a osservazioni provenienti da un centro con base in una università statunitense. DBpedia, Open Library data mirror in the Talis Platform, The Open Library e Freebase sono fonti ben note di informazioni enciclopediche su una vasta gamma di argomenti; hanno anche una serie di licenze aperte: CC BY-SA, Other (aperta), Other (dominio pubblico) e CC BY. Data Incubator: MusicBrainz (altra licenza (dominio pubblico)) contiene informazioni relative a musica, e, precisamente a album, artisti, incisioni, etichette e relazioni tra loro.

I dieci maggiori pacchetti senza licenza aperta sono:

Package	Number of triples
TWC: Linking Open Government Data	9.80 billion
Data.gov	6.40 billion
Source Code Ecosystem Linked Data	1.50 billion
2000 U.S. Census in RDF (rdfabout.com)	1.00 billion
PubMed	0.80 billion
DBTune.org MySpace RDF Service	0.66 billion
UniParc	0.63 billion
DBTune.org AudioScrobbler RDF Service	0.60 billion
Linking Italian University Statistics Project	0.59 billion
UniProt UniRef	0.49 billion

Tabella 13

TWC: Linking Open Government Data è il pacchetto più grande di The Cloud ed è un’aggregazione di dati del governo statunitense. Contiene dati pubblicati nel pacchetto Data.gov package. The Data Hub non ha alcuna informazione relativa alla licenza di tali dati. An-

che 2000 U.S. Census in RDF (rdfabout.com) sono dati del governo U.S.A. sulle statistiche relative alla popolazione ed ha una licenza CC BY-NC. Su The Data Hub non sono riportate informazioni sulle licenze in merito ai seguenti pacchetti:

- Source Code Ecosystem Linked Data contiene codici sorgente strutturati di progetti open source. Viene curato da un'università canadese.
- PubMed è una fonte Americana di pubblicazioni mediche.
- DBTune.org MySpace RDF Service e DBTune.org AudioScrobbler RDF Service sono parte di una mini-cloud di nove pacchetti di argomento musicale.
- UniParc e UniProt UniRef fanno parte di archivi di scienze della vita di istituzioni accademiche americane.
- Linking Italian University Statistics Project è la pubblicazione dei dati del governo italiano sugli studenti universitari.

Quali sono gli argomenti nei dati?

Nel wiki di The Data Hub all'interno delle descrizioni per ogni pacchetto ci sono una serie di 'tag' differenti. Alcuni di questi sono chiaramente basati su soggetti e offrono un'indicazione sui contenuti dei pacchetti. Sembrerebbe non esserci una terminologia controllata in uso, quindi lo stesso soggetto potrebbe essere rappresentato da un tag diverso in pacchetti diversi. Nella nostra analisi abbiamo individuato un certo numero di tag che sembrano rappresentare lo stesso soggetto. Si noti anche che i pacchetti possono avere più di un soggetto. I dieci soggetti più comuni in The Cloud sono i seguenti:

Subject tag	Number of packages with tag	% of packages with tag
publications	94	30.23
government	54	17.36
life sciences	46	14.79
geographic	40	12.86
media	32	10.29
library	22	7.07
United Kingdom	22	7.07
education	20	6.43
user generated content	19	6.11
bibliographic	15	4.82

Tabella 14

Questi risultati rispecchiano generalmente le categorie illustrate nella versione a colori del diagramma che rappresenta The Cloud. Rappresentano anche un'istantanea dello stato attuale del contenuto. In queste aree di The Cloud i dati sono predominanti. Al confronto, i dati sui beni culturali sono molto pochi, e questo probabilmente perché fino all'avvento di Europeana in questa comunità non c'era interesse per i linked data. La presenza di un tag 'Regno Unito' dimostra ampiamente l'influenza della politica del governo inglese nel pubblicare linked data. Il ruolo degli USA non è evidente, ma questo perché i pacchetti non hanno un tag 'Stati Uniti' anche quando potenzialmente dovrebbero.

Quali formati sono usati per codificare i dati?

Vengono usati vari formati per codificare i dati per The Cloud. La letteratura sui linked data parla prevalentemente di 'vocabolario'. Qui seguiamo ad usare il termine 'formato' per evitare la confusione con l'uso di vocabolario nell'ambito dei beni culturali come insieme di termini descrittivi utilizzati al posto degli elementi di metadati. Si noti anche che alcuni dei formati sono chiamati 'ontologie'. I più comunemente usati sono elencati in tabella 15.

Format	Number of packages using the format	% of packages using the format
Resource Description Framework (rdf)	261	83.92
Dublin Core (dc)	97	31.19
Friend of a Friend (foaf)	84	27.01
Simple Knowledge Organization System (skos)	57	18.33
RDF Schema (rdfs)	42	13.50
Web Ontology Language (owl)	34	10.93
Basic Geo (geo)	25	8.04
Advanced Knowledge Technologies Reference Ontology (akt)	22	7.07
eXtensible HyperText Markup Language (xhtml)	19	6.11
Bibliographic Ontology (bibo)	14	4.50
none given	13	4.18
Music Ontology (mo)	13	4.18
DBpedia Ontology (dbpedia)	12	3.86
vCard (vcard)	11	3.54
Semantically-Interlinked Online Communities (sioc)	10	3.22
Creative Commons (cc)	8	2.57
Functional Requirements for Bibliographic Records (frbr)	6	1.93
GeoNames Ontology (geonames)	6	1.93
XML Schema (xsd)	6	1.93
Event Ontology (event)	5	1.61

Tabella 15: L'abbreviazione tra parentesi dopo il nome indica l'etichetta data al formato.

I formati sono di tre tipologie:

Base – Quelli che generalmente organizzano le entità di The Cloud, compresi i legami tra le entità. Come prevedibile, si trovano in quasi tutti i suoi pacchetti. Quindi è probabile che qualsiasi pacchetto di beni culturali ne faccia uso. Essi sono: Resource Description Framework, RDF Schema, Web Ontology Language, e XML Schema.

Descrittivi – Quelli i cui elementi hanno dati descrittivi sulle entità che vengono usate in molti pacchetti. In genere vengono sviluppati da un gruppo di partner interessati che vogliono pubblicare le loro informazioni come linked data. Molto spesso nascono da specifici progetti o iniziative. Questi sono: Dublin Core (risorse web), Friend of a Friend (persone), Simple Knowledge Organization System (terminologie), Basic Geo (geografia), Bibliographic Ontology, Music Ontology, vCard (business card), Semantically-Interlinked Online Communities (reti sociali), Creative Commons (IPR), Functional Requirements for Bibliographic Records e Event Ontology.

Specifici per pacchetto – Quelli i cui elementi rappresentano i dati specifici contenuti in un particolare pacchetto. Sono stati sviluppati nel contesto della pubblicazione di un singolo pacchetto come linked data, tuttavia possono essere usati nella pubblicazione di altri pacchetti che conducono ad essi divenendo standard de facto. Questi sono: Advanced Knowledge Technologies Reference Ontology, DBpedia Ontology e GeoNames Ontology.

Il fatto che ci siano alcuni formati di questo tipo usati da più di un pacchetto è significativo, sta ad indicare che questi 'pacchetti parenti' giocano un ruolo fondamentale in The Cloud. Ne sono esempi ovvi

DBpedia e GeoNames, e vedremo uno schema simile nella sezione seguente nell'ambito dei collegamenti in The Cloud. Nel momento in cui Berners-Lee suggerisce di usare un formato 'standard' risulta poi sorprendente che 75 formati siano usati da due pacchetti o anche meno. Usando un'analogia presa in prestito dalla biologia, ciò a cui stiamo assistendo è forse un'esplosione evolutivistica della 'specie' in un nuovo ambiente. Nell'interesse dell'interoperabilità, si spera che cominci a funzionare 'la sopravvivenza dei più idonei'. Sembra, in realtà, che i linked data siano ancora in una fase sperimentale.

Com'è collegata The Cloud?

La cosa più importante di The Cloud è come sono collegati i pacchetti tra di loro. Il sito The Data Hub ci permette di vedere i link nel dettaglio. In termini di numero di collegamenti di pacchetti, generalmente i dieci più collegati sono:

Package being linked to	Number of packages linking	Number of links
DBpedia	158	31,531,365
GeoNames Semantic Web	42	9,353,935
(none)	34	0
DBLP Computer Science Bibliography (RKBExplorer)	27	1,338,927
Association for Computing Machinery (ACM) (RKBExplorer)	26	1,487,410
ePrints3 Institutional Archive Collection (RKBExplorer)	26	281,385
Freebase	25	10,452,728
CiteSeer (Research Index) (RKBExplorer)	24	805,921
School of Electronics and Computer Science, University of Southampton (RKBExplorer)	24	37,996
ReSIST Project Wiki (RKBExplorer)	24	408

Tabella 16

I 'vincitori' sono nettamente DBpedia, GeoNames Semantic Web e Freebase. Questi sono collegati al 50.8%, 13.5% e 8.0% rispetto agli

altri pacchetti di The Cloud. Si presuppone che questo successo sia dovuto al fatto che sono molto conosciuti. I sei pacchetti nell'elenco che terminano per '(RKBExplorer)' fanno parte di una mini-cloud di circa 50 pacchetti. RKBExplorer¹⁰ è un sistema per la pubblicazione di linked data sviluppato nel corso del progetto ReSIST¹¹ finanziato dall'Unione Europea. È dotato di un browser che consente agli utenti di esplorare i set di dati collegati tra loro. È interessante osservare, e forse ad una prima occhiata sorprendente, che oltre il 10% dei pacchetti di The Cloud non si collegano ad altri pacchetti. Essi sono generalmente collegati, o sono stati pubblicati per essere collegati. In questo gruppo sono compresi anche i pacchetti più grandi, vale a dire Data.gov, 2000 U.S. Census in RDF (rdfabout.com), data.gov.uk Time Intervals, UniParc, The Open Library e GeneID.

I dieci pacchetti generalmente più collegati, in termini di link, sono:

Package being linked to	Number of packages linking	Number of links
UniProtKB Taxonomy	6	46,630,898
MARC Codes List	3	42,409,958
QDOS	1	40,000,000
UniProtKB	10	33,447,122
DBpedia	158	31,531,365
Ordnance Survey Linked Data	16	29,717,902
UniParc	1	27,534,215
IdRef: Sudoc authority data	3	20,040,000
Sudoc bibliographic data	1	20,000,000
flickr TM wrapppr	4	16,358,998

Tabella 17

DBpedia è l'unico pacchetto che compare sia in questa che nella tabella precedente, il che rafforza la sua 'popolarità'. flickrTMwrapppr

¹⁰<http://www.rkbexplorer.com>.

¹¹<http://www.resist-noe.org>.

viene collegato ampiamente da DBpedia per fornire immagini per i suoi concetti. Pacchetti con 'UniProt' all'inizio del nome e il pacchetto UniParc sono parte di una minicloud del dominio delle proteine. Sudoc è il catalogo collettivo accademico francese e qui i link sono tra pacchetti collegati ad esso. Ordnance Survey Linked Data sono i dati geografici del Regno Unito, collegati da pacchetti di quel paese, soprattutto i pacchetti di dati governativi. QDOS è collegato ad un pacchetto che riguarda la musica popolare. Questa analisi prova che il collegamento tra pacchetti non è, per lo meno non ancora, qualcosa che sta crescendo in modo 'organico'. Ci sono iniziative responsabili della creazione di ampie parti di The Cloud. Ciò implica che anche per il settore dei beni culturali è necessaria un'iniziativa del genere. Europea sta assumendo un ruolo fondamentale in questo contesto.¹²

Dati sui beni culturali in 'The Cloud'

Ci sono 18 pacchetti in The Cloud che potrebbero essere individuati come aventi 'beni culturali' come loro soggetto o ad esso collegato:

¹²<http://version1.europeana.eu/web/lod>.

Package IPR Number of triples

VIAF: The Virtual International Authority File	(not given)	200,000,000
Europeana Linked Open Data	(not given) ¹³	185,000,000
British National Bibliography (BNB)	CC0	80,249,538
Hungarian National Library (NSZL) catalog	(not given)	19,300,000
Amsterdam Museum as Linked Open Data in the Europeana Data Model	CC BY-SA	5,000,000
Library of Congress Subject Headings	(not given)	4,151,586
Swedish Open Cultural Heritage	Other (Open)	3,400,000
Calames	[not given]	2,000,000
RAMEAU subject headings (STITCH)	[not given]	1,619,918
data.bnf.fr - Bibliothèque nationale de France	(not given)	1,400,000
National Diet Library of Japan subject headings	(not given)	1,294,669
Gemeenschappelijke Thesaurus Audiovisuele Archieven – Common Thesaurus Audiovisual Archives	ODbL	992,797
Gemeinsame Normdatei (GND)	Other (non-commercial)	629,582
Archives Hub Linked Data	CC0	431,088
Thesaurus for Graphic Materials (t4gm.info)	CC BY-SA	103,000
Italian Museums (LinkedOpenData.it)	CC BY-SA	49,897
Thesaurus W for Local Archives	(not given)	11,000
MARC Codes List Open Data	Other (Public Domain)	8,816

Tabella 18

Due dei pacchetti, Amsterdam Museum e Europeana stessa, sono direttamente collegati a Europeana. Evidenti sono le iniziative francesi per i linked data, specialmente nell'ambito delle terminologie: Calames, RAMEAU subject headings (STITCH), data.bnf.fr - Bibliothèque nationale de France e Thesaurus W for Local Archives. Si è notato pure nell'indagine sui partner di Linked Heritage. Anche la Svezia sta facendo qualcosa di simile con Swedish Open Cultural Heritage. L'Italia sta cominciando a seguire lo stesso percorso. C'è un ulteriore contributo relativo alla terminologia e ai file di autorità con: VIAF Virtual International Authority File, British National Bibliography (BNB), Library of Congress Subject Headings, National Diet Library of Japan subject headings, Gemeinsame Normdatei (GND), Thesaurus for Graphic Materials (t4gm.info) e MARC Codes List Open Data. Infine, c'è un contributo dal mondo delle biblioteche (Hungarian National Library (NSZL) catalog), degli archivi

¹³Sarà pubblicato come CC0.

(Archives Hub Linked Data) e degli archivi audio-visivi (Gemeenschappelijke Thesaurus Audiovisuele Archieven – Common Thesaurus Audiovisual Archives). La parte di The Cloud relativa ai beni culturali è ancora molto piccola (circa 500m triple o <1.5%), tuttavia gli sviluppi di Europeana incrementeranno in modo significativo la sua dimensione. Linked Heritage ne sarà una componente notevole. Approfondiamo ulteriormente nei dettagli la mini cloud dei beni culturali. I pacchetti dei beni culturali utilizzano i formati indicati in tabella 19 nella pagina successiva. Il quadro generale nel complesso è simile a The Cloud, eccetto l'uso di SKOS, molto più significativo, che sta ad indicare l'importanza delle risorse terminologiche e degli authority file nel settore. Di rilievo è la mancanza di un formato specifico per le informazioni relative ai musei. Inoltre, in The Data Hub non è menzionato Europeana Data Model, ma si evince da altre fonti che è stato usato dall' Amsterdam Museum e, probabilmente, dai pacchetti di Europeana.

In The Cloud i pacchetti dei beni culturali sono collegati come mostrato in tabella 20 nella pagina seguente. Come prevedibile, DBpedia è il pacchetto più diffuso a cui collegarsi. Un altro pacchetto 'generale' è GeoNames Semantic Web. Entrambi sono stati messi in evidenza nell'indagine Linked Heritage e rappresentano fonti ben note di informazioni trasversali e geografiche a cui collegarsi. A parte questo, il resto dei pacchetti collegati sono prevalentemente altri pacchetti di beni culturali, e in particolare terminologie standard e authority file. Se consideriamo l'uso della serializzazione rappresentato dalla tabella 21, RDF/XML è usato da tutti i pacchetti tranne due: Europeana Linked Open Data cita solo N-Triple e il pacchetto Calames non cita alcuna serializzazione. N-Triple è abitualmente pubblicato insieme a RDF/XML. L'unica occorrenza di Turtle è insieme a RDF/XML.

Ciò fa pensare che i linked data sui beni culturali dovrebbero essere

Format	Number of packages using the format
Resource Description Framework	13
Simple Knowledge Organization System	11
Dublin Core	7
eXtensible HyperText Markup Language	4
Friend of a Friend	3
Basic Geo	1
Bibliographic Ontology	1
DBpedia	1
Music Ontology	1
Object Reuse and Exchange	1
RDF Schema	1
vCard	1
Web Ontology Language	1
XML Schema	1

Tabella 19

Package being linked to	Number of packages linking	Number of links
DBpedia	5	82,308
Library of Congress Subject Headings	4	108,135
VIAF: The Virtual International Authority File	2	1,820,684
GeoNames Semantic Web	2	510,658
Dewey Decimal Classification (DDC)	2	200,543
RAMEAU subject headings (STITCH)	2	83,530
Swedish Open Cultural Heritage	1	100,489
Gemeinsame Normdatei (GND)	1	20,000
IdRef: Sudoc authority data	1	10,000
(DCMI Type Vocabulary – not in The Cloud)	1	10,000
UK Postcodes	1	5,000
AGROVOC	1	700
Hungarian National Library (NSZL) catalog	1	136
(none)	1	0

Tabella 20

Serialisation	Number of packages using (%)
RDF/XML	16 (88.9%)
N-Triples	5 (27.8%)
Turtle	1 (5.5%)
(none given)	1 (5.5%)

Tabella 21

pubblicati per lo meno come RDF/XML e possibilmente come N-Triples per essere compatibili con i dati esistenti. Tuttavia non c'è motivo per cui non possano essere usate tutte le serializzazioni.

Raccomandazioni sulle buone pratiche

La pubblicazione dei linked data si trova ancora in una fase sperimentale e le buone pratiche sono ancora in una fase emergente. Quindi le raccomandazioni date in questa sezione si basano su:

- La pratica comune adottata nella comunità generale dei linked data, come rappresentata da The Cloud;
- La pratica delle organizzazioni dei beni culturali che hanno pubblicato linked data;
- La pratica generale del settore dei beni culturali.

Alcune delle raccomandazioni offrono una gamma di opzioni senza indicare la scelta 'giusta'. Quella operata dall'organizzazione è la scelta che si basa su circostanze individuali e può essere determinata da considerazioni di ordine etico e legale. Le raccomandazioni possono essere distinte in tre 'aree di scelta'.

Quali informazioni pubblicare come linked data

Se consideriamo quale genere di informazioni vengono pubblicate come linked data in The Cloud e specialmente quella parte relativamente piccola che riguarda i beni culturali, dobbiamo tenere presente due tipi principali di informazioni:

Informazioni sulle collezioni

Queste sono il grosso delle informazioni pubblicate dalle organizzazioni di beni culturali. Tuttavia, queste dovrebbero considerare anche l'opportunità di pubblicare informazioni su:

- Surrogati – i risultati della digitalizzazione;
- Materiali di supporto – inclusi i cataloghi di mostre, libri, file storici e unità di studio;
- Contenuti generati dall'utente – reazioni alle collezioni (dopo aver ottenuto il permesso di pubblicare).

Informazioni terminologiche

Se esaminiamo The Cloud vediamo che una gran parte di essa è costituita da risorse terminologiche utilizzate dalle organizzazioni di beni culturali. Ciò potrebbe essere la conseguenza di iniziative istituzionali internazionali, nazionali e tematiche. C'è molta enfasi in questo senso nei settori delle biblioteche e degli archivi. Ed è inclusa la pubblicazione di autorità di nomi. Anche questo lavoro offre l'opportunità per pubblicazioni in cooperazione, possibilmente di livello internazionale e multilingue e nel contesto di progetti finanziati dall'Unione Europea. Tra gli argomenti delle pubblicazioni terminologiche ricordiamo: tipi di oggetti, metodi di eventi (vale a dire metodo di creazione), luoghi, organizzazioni, eventi, materiali,

iconografia e molti altri. Il consiglio principale nello scegliere che tipologia di dati pubblicare come linked data è:

- Pubblicare informazioni su tutti gli aspetti delle collezioni e dei materiali ad esse relativi;
- Pubblicare informazioni terminologiche e cercare partner con cui collaborare per evitare duplicazioni.

Quale licenza per i linked data

Questa sezione riguarda gli accordi sulle licenze associate con la pubblicazione dei linked data. Le scelte operate in questo ambito sono influenzate da considerazioni di carattere generale su quanto controllo chi pubblica i linked data vuole avere sui propri dati, ma anche da che tipo di dati vengono pubblicati. Come visto, analizzando The Cloud, sembrerebbe che una gran parte dei linked data pubblicati non abbia una licenza per l'uso. Il risultato è che non è chiaro cosa si può fare con questi dati. In tempi di conflittualità come i nostri gli utenti sono particolarmente attenti a non fare cose che li esponano ad una possibile perdita di reputazione istituzionale se non addirittura ad una causa legale. Il consiglio principale per quanto riguarda le licenze è:

- Qualsiasi pubblicazione di linked data deve essere accompagnata da una licenza che chiarisca l'uso che può essere fatto dei dati;
- La licenza può essere standard, vale a dire fornita da Creative Commons, oppure creata specificamente dall'editore.

Generalmente, le licenze sono di due tipologie:

Licenza aperta – Consente qualsiasi uso dei dati, anche di tipo commerciale, a volte con delle restrizioni sull'attribuzione e sull'uso improprio.

Licenza non aperta – L'uso viene ristretto solo a quello non commerciale, con requisiti simili per l'attribuzione e l'uso improprio.

Ci sono una gamma di licenze per entrambe le tipologie, e cioè quelle di Creative Commons e GNU e l'opzione per una licenza istituzionale specifica.

Come pubblicare linked data

In questo ambito un potenziale editore di linked data ha tre scelte da compiere:

Quali standard di formato usare

Non è concepibile che non si usino standard di base quali RDF, RDFS e OWL. Tuttavia, per i formati 'descrittivi' sarebbe consigliabile:

- Non creare un formato proprietario concepito per essere utilizzato esclusivamente per il proprio pacchetto;
- Usare formati standard adeguati per i tipi di dati da pubblicare. Se consideriamo ciò che viene usato ora, alcuni formati sembrano essere dei buoni suggerimenti:
 - Risorse web: Dublin Core;
 - Persone: Friend of a Friend;

- Risorse terminologiche: Simple Knowledge Organization System;
- Risorse bibliografiche: Bibliographic Ontology;
- Musica: Music Ontology.

Queste raccomandazioni si basano sui formati attualmente in uso. Tuttavia c'è 'un vuoto nel mercato' per il formato di linked data di beni culturali.

Usare¹⁴ un formato specifico per i linked data dei beni culturali. I formati da prendere in considerazione sono quelli basati su EDM, CIDOC CRM e LIDO.

Serializzazioni RDF per pubblicare

Sulla base della pratica comune è consigliabile pubblicare i linked data nelle serializzazioni RDF/XML e N-Triple.

Come collegare il pacchetto in The Cloud

Una questione emersa nell'ambito delle discussioni del WP 2 Working Group è stata: quali sono i 'pacchetti affidabili' in The Cloud? Un criterio di affidabilità è se si conosce l'editore del pacchetto. Questo tipo di collegamento sembra essere molto comune in tutte le parti di The Cloud e porta alla formazione di mini cloud di pacchetti collegati tra loro. Sembra esserci una formazione di mini cloud di beni culturali. Una ragione potrebbe essere l'iniziativa di Europea. Altre questioni molto importanti sono:

- L'individuazione di risorse. Se gli identificatori usati sono compatibili con gli identificatori usati nel potenziale pacchetto di collegamento;

¹⁴Il progetto Linked Heritage offre alla comunità l'opportunità di considerare queste possibilità, in particolare quella di utilizzare LIDO. Vedi sezione seguente.

- La compatibilità delle semantiche dei pacchetti. Ad esempio, se si vogliono individuare le 'persone' (entità pubbliche), è la stessa cosa di FOAF, che sostiene di identificare le persone.
- Un pacchetto deve essere accessibile alle richieste.

Quindi, si consiglia di:

- Collegarsi a pacchetti di carattere generale, che siano spesso collegati a DBpedia, GeoNames Semantic Web, fonti nazionali di terminologia (ad esempio UK Postcodes);
- Collegarsi a pacchetti conosciuti nell'ambito dei beni culturali, ad esempio Library of Congress Subject Headings, VIAF: The Virtual International Authority File e Dewey Decimal Classification);
- Fornire una risultante SPARQL al pacchetto.

Naturalmente l'atto finale è registrare il pacchetto nell'archivio The Data Hub!

Il lavoro futuro sui Linked Data

Nella fase successiva del progetto il WP 2 si concentrerà su due obiettivi che dimostreranno il potenziale dei linked data:

Task 2.3 – Technical specifications

This will specify how cultural heritage information can be enriched by, and can enrich, the 'Cloud'. We will identify: models, processes and technologies which offer the best potential. Selection criteria will include:

- In questa fase verrà specificato come le informazioni sui beni culturali possano essere arricchite da, e arricchire, 'The Cloud'. Individueremo modelli, processi e tecnologie che offrano le potenzialità migliori. I criteri di selezione includeranno:
- L'uso attuale dei linked data nei beni culturali e in ambito umanistico;
- L'uso degli standard;
- Capacità di interoperabilità con altri archivi di linked data. Tra questi ci saranno i 'protagonisti principali' già individuati, quali DBpedia e GeoNames;
- Capacità di integrazione con le tecnologie scelte in altri pacchetti di lavoro tematici di Linked Heritage (vale a dire partner pubblici e privati e terminologie);
- Maturità e qualità nell'implementazione tecnica, nella documentazione e nelle azioni di supporto.

Task 2.4 – Enabling linked cultural heritage data

Verrà dimostrato come potenziare le attuali procedure per consentire ai fornitori di contenuti di pubblicarli come linked data, oltre che di pubblicarli in Europeana. Nello specifico:

- I fornitori potranno contribuire con i loro contenuti agli archivi di linked data e mantenere le loro informazioni già esistenti sui linked data;
- Potenziare i processi di ingestione con strumenti che consentano di:
 - Esplorare l'archivio di linked data e le sue connessioni con fonti esterne;

- Creare e curare i collegamenti tra entità;
- Ampliare le possibilità di recupero per includere fonti preferite per i link e informazioni testuali.

Ai fini di una corretta indicizzazione, si invitano i lettori a citare esclusivamente il testo in lingua inglese; l'unico, infatti, che presenta l'indicazione del numero di pagina, l'abstract, le keywords e le date del processo redazionale.

McKenna, G. "Linked heritage experience in Linking Heritage Information". *JLIS.it*. Vol. 4, n. 1 (Gennaio/January 2013): Art: #6304. DOI: [10.4403/jlis.it-6304](https://doi.org/10.4403/jlis.it-6304). Web.



TRADUZIONI