
I metadati per il catalogo nazionale Open Data DATI.GOV.IT

3^a

Conferenza



OpenGeoData Italia

dati geografici:

***l'Open Data
consapevole***

19 febbraio
2015

ROMA

Dati.gov.it

Il catalogo del portale nazionale degli Open Data è disponibile all'indirizzo <http://www.dati.gov.it/catalog>, viene aggiornato periodicamente prelevando le informazioni dai cataloghi messi a disposizione dalle Pubbliche Amministrazioni

La procedura di aggiornamento prevede un automatismo denominato "DGIBot", che verifica con cadenza settimanale i contenuti pubblicati

Lo schema di riferimento adottato per la metadattazione è quello definito dall'Agenzia per l'Italia Digitale nelle linee guida per la valorizzazione del patrimonio informativo pubblico.

Gli standard ed i formati supportati fanno riferimento agli standard definiti dal Programma ISA4 della Commissione europea, in particolare DCAT e DCAT-AP, che definiscono rispettivamente i vocabolari per i cataloghi dati e il set minimo di informazioni che devono essere predisposte nei metadati per rendere interoperabili i cataloghi delle Pubbliche Amministrazioni della UE.





dati geografici:

**l'Open Data
consapevole**

 19 febbraio
2015
ROMA

Formati supportati dati.gov.it

Standard/ Formato	Specifiche	Descrizione
DCAT / DCAT-AP	RDF/XML, JSON-LD	Il vocabolario DCAT e il relativo Application Profile, sono entrambi standard di riferimento per il recupero delle informazioni sui dataset.
RSS	RSS 2.0, ATOM	RSS e ATOM sono formati basati su XML per la sottoscrizione di contenuti web.
CKAN AP	CKAN	Sono API native della piattaforma CKAN utilizzata da dati.gov.it.
SODA API	Socrata Open Data API	Sono API native della piattaforma SOCRATA.
Tabular	CSV, TSV, XLS e XLSX	Sono supportati i formati Comma Separated o Tab Separated Values e Microsoft Excel (XLS e XLSX). Viene fornito un template di esempio da riutilizzare.

Ontologia DCAT per la descrizione di dataset

Per descrivere i cataloghi open data e i relativi dataset è stato definito dal W3C un vocabolario studiato appositamente per tali scopi .

DCAT [<http://www.w3.org/TR/vocab-dcat/>]: è un vocabolario RDF/OWL per descrivere cataloghi e Dataset Open Data sul web a prescindere dai formati di distribuzione. Attualmente è un Working Draft del W3C.

DCAT definisce il concetto di Dataset ed è general-purpose e particolarmente adatto a descrivere qualsiasi Open Data.

Riutilizza altri vocabolari standard del Semantic Web come **dc-terms**, **foaf**, **dc-terms**, **rdf**, **rdfs**, **skos**, **vcard**.

DCAT ed il relativo Vocabolario sono riutilizzabili oltre che per dati.gov.it anche per altri hub di aggregazione di informazioni sui dataset



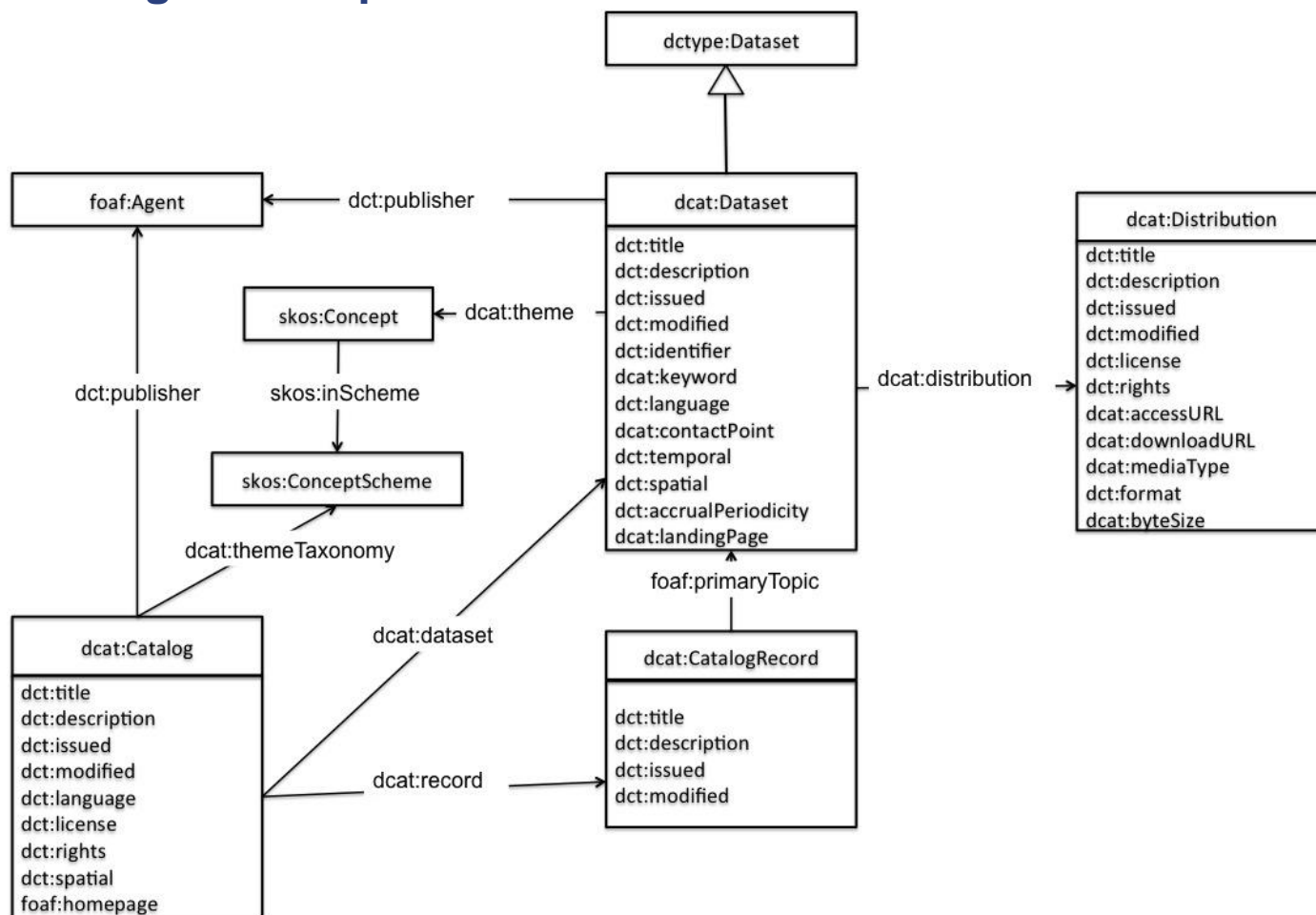


dati geografici:

***l'Open Data
consapevole***

19 febbraio
2015
ROMA

Ontologia DCAT per la descrizione di dataset



Fonte: <http://www.w3.org/TR/vocab-dcat/>



dati geografici:

**l'Open Data
consapevole**

19 febbraio
2015
ROMA

Nome	Vocabolario	Descrizione e note
name	DC:identifier	Identificativo univoco del dataset
title	DC:title	Titolo del dataset.
Extras {“theme”}	DCAT:theme	Rappresenta l’area tematica principale, alla quale il dataset viene associato. L’area tematica principale consiste in una descrizione conforme al vocabolario dei domini principali (<i>top-level domains</i>) EuroVoc. Ulteriori aree tematiche possono essere specificate attraverso l’appartenenza ai groups .
Extras {“micro theme”}	DCAT:theme	Il microtheme corrisponde ad un “concetto” o termine nella gerarchia del tesauro EuroVoc. Ulteriori aree tematiche che impiegano vocabolari diversi possono essere specificate attraverso l’appartenenza ai groups .
notes	DC:description	Descrizione estesa del dataset.



Nome	Vocabolario	Descrizione e note
<i>author</i>	DC:creator , FOAF:name	Nominativo della persona, dell'organizzazione o del servizio che produce i dati del dataset.
<i>Author_email</i>	FOAF:mbox	Indirizzo email associato alla persona, organizzazione o servizio responsabile della produzione dei dati del dataset.
<i>Maintainer</i>	DC:publisher , FOAF:name	Nominativo della persona, organizzazione o servizio che pubblica e rende disponibile e accessibile il dataset.
<i>Maintainer_email</i>	FOAF:mbox	Indirizzo email associato alla persona, organizzazione o servizio responsabile della pubblicazione del dataset.
<i>Metadata_created</i>	DC:issued , DC:created	Data in cui il dataset è stato creato (created) o rilasciato (issued) in modalità aperta per la prima volta



Nome	Vocabolario	Descrizione e note
Metadata_modified	DC:modified	Data in cui il dataset è stato modificato e quindi aggiornato.
license_id	DC:license	Il formato della licenza con la quale i dati sono forniti. Il campo consiste in un codice conforme agli identificativi di licenze disponibili su DATI.GOV.IT.
url	DCAT:landingPage	Indica la URL dove il dataset è descritto.
tags[]	DCAT:keyword	Indica le parole chiave, separate da virgole che descrivono il dataset.
groups[]	DCAT:theme	I groups sono degli insiemi che contengono tutti i dataset ai quali è associato il nome del gruppo stesso. Ogni gruppo rappresenta una area tematica e classifica i dataset secondo un vocabolario o un tesaurus.



dati geografici:

**l'Open Data
consapevole**

19 febbraio
2015
ROMA

Nome	Vocabolario	Descrizione e note
Extras {“spatial”}	DC:spatial	<p>Quando i dati del dataset hanno significato solo all'interno di una determinata copertura spaziale questo campo diviene obbligatorio. Ai fini dell'interoperabilità la copertura territoriale deve essere specificata in un formato conforme allo standard GeoJSON</p> <p>Le geometrie GeoJSON supportate sono solo Punti (<i>Point</i>), Linee (<i>LineString</i>) e Poligoni (<i>Polygon</i>). Sono inoltre supportati i Multi Poligoni (<i>Multipolygon</i>) ma solo nel caso sono disgiunti ovvero i poligoni riportati non si intersecano o sono contenuti l'uno nell'altro:</p>
resources []	DCAT:distribution	<p>Rappresenta un array di risorse ciascuna delle quali specifica una forma attraverso la quale il dataset è reso disponibile. Ogni risorsa viene impiegata per distribuire formati o contenuti diversi del dataset.</p>



dati geografici:

 l'Open Data
 consapevole

 19 febbraio
 2015
 ROMA

Nome	Vocabolario	Descrizione e note
Resources <i>[{name}]</i>	DC:title	Il nome della risorsa. Tipicamente è il nome del file, ma può essere anche un nome per una API o che identifica uno SPARQL Endpoint
Resources <i>[{format}]</i>	DC:format	Il formato della risorsa o del file.
Resources <i>[{url}]</i>	DCAT:access URL , DCAT:downloadURL	Il link per effettuare il download del file o per accedere ad una API o un endpoint SPARQL
Resources <i>[{created}]</i>	DC:created	Data di creazione o prima disponibilità
Resources <i>[{last_modified}]</i>	DC:modified	Data di aggiornamento o modifica,



Osservazioni

- Nei campi **Extras{"theme"}** , **extras{"microtheme"}** e **groups[]**, le descrizioni di Eurovoc sono di alto livello. Si può trovare difficoltà a definire puntualmente il tema ed il sottotema, con il rischio di fornire una classificazione inutile per l'utente. Si può definire un vocabolario skos/rdf condiviso per i nostri scopi in aggiunta a Eurovoc.
- Anche per il caso del campo **tags[]**, si potrebbe adottare un vocabolario controllato che possa essere di guida. In questo caso però non è possibile correlarlo come vocabolario skos/rdf..
- **Extras {"spatial"}**, il formato geoJSON supporta anche Point, LineString, Polygon, MultiPoint, MultiLineString, and MultiPolygon. Va prevista una specifica linea guida per questo campo in particolare per rappresentazioni complesse ed articolate.

Riferimenti utili

Guida sintetica per lo scambio di dati con il catalogo nazionale Open Data :

<http://goo.gl/ZEVxDF>

DATI:GOV.IT: <http://www.dati.gov.it/>

DCAT: <http://www.w3.org/TR/vocab-dcat/>

DCTerms: <http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms#>

FOAF: <http://www.foaf-project.org/>

OWL: <http://www.w3.org/TR/owl-features/>

RDF: <http://www.w3.org/RDF/>

SPARQL: <http://www.w3.org/TR/rdf-sparql-query/>

Virtuoso: <http://virtuoso.openlinksw.com/>

3^a
Conferenza
OpenGeoData Italia



dati geografici:
***l'Open Data
consapevole***

19 febbraio
2015
ROMA