



L'Abate



Opening Soil Data through Semantics and Linked Data Vocabularies



# Opening Soil Data through Semantics and Linked Data Vocabularies

Giovanni L'Abate\*,  
Caterina Caracciolo\*\*, Ferdinando Villa\*\*\*, Edoardo A.C. Costantini\*

\*Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA)

\*\*Food and Agriculture Organization of the UN (FAO)

\*\*\*Basque Centre for Climate Change (BC3); IKERBASQUE, Basque foundation for science

# This presentation

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

1. Soil classification and soil data
2. The problem
3. Previous work
4. Current work
5. Next step

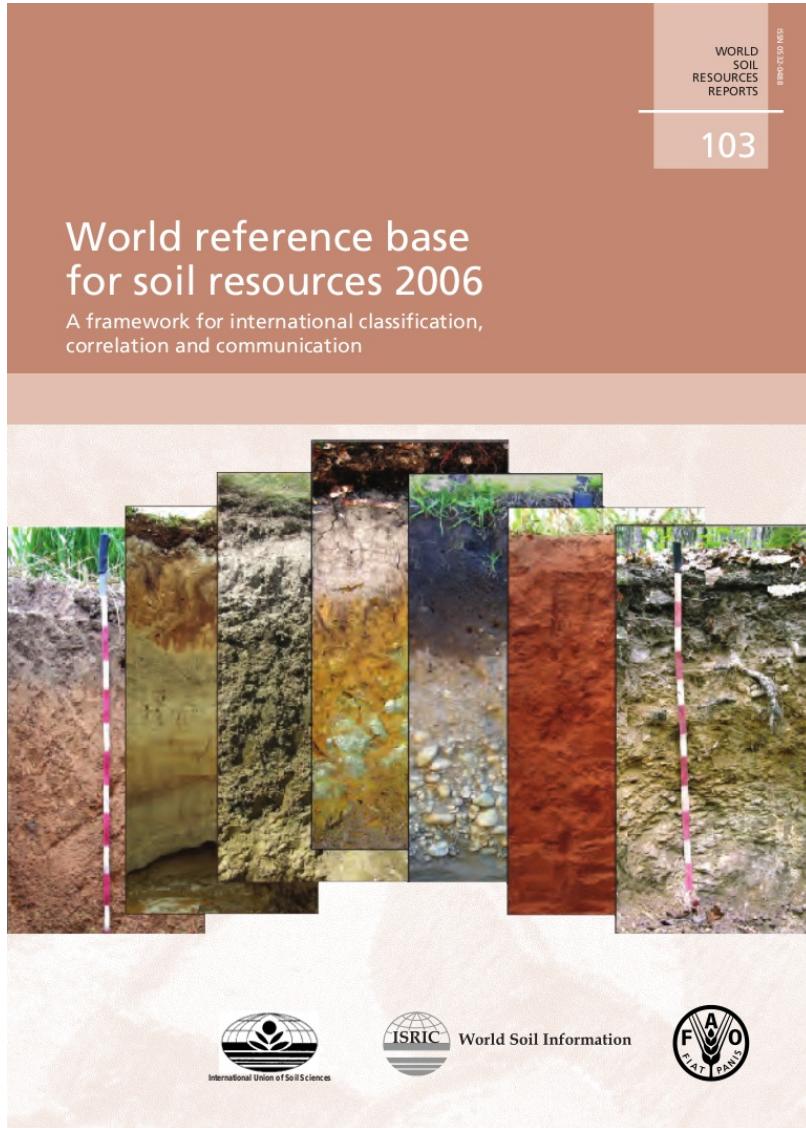
25/06/13

CONFERENZA GARR 2016 30 NOVEMBRE - 2 DICEMBRE 2016, FIRENZE

# The World Reference Base for Soil Resources

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies



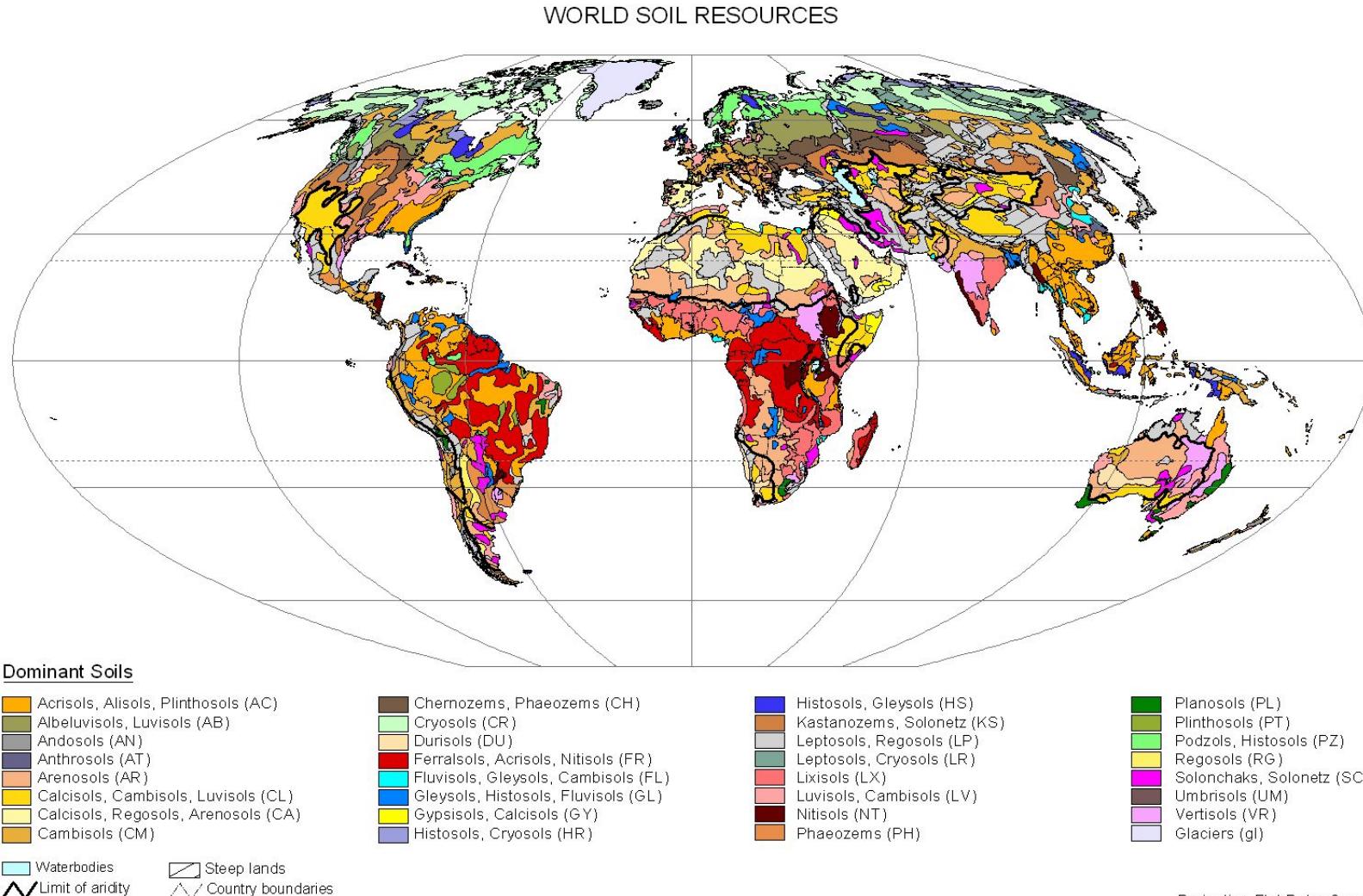
The World Reference Base for Soil Resources (WRB) is the internationally recognized classification system for soils.

It is commonly used to classify soil profiles and generate maps of soil distribution.

# The World Resource Base Map of World Soil Resources

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

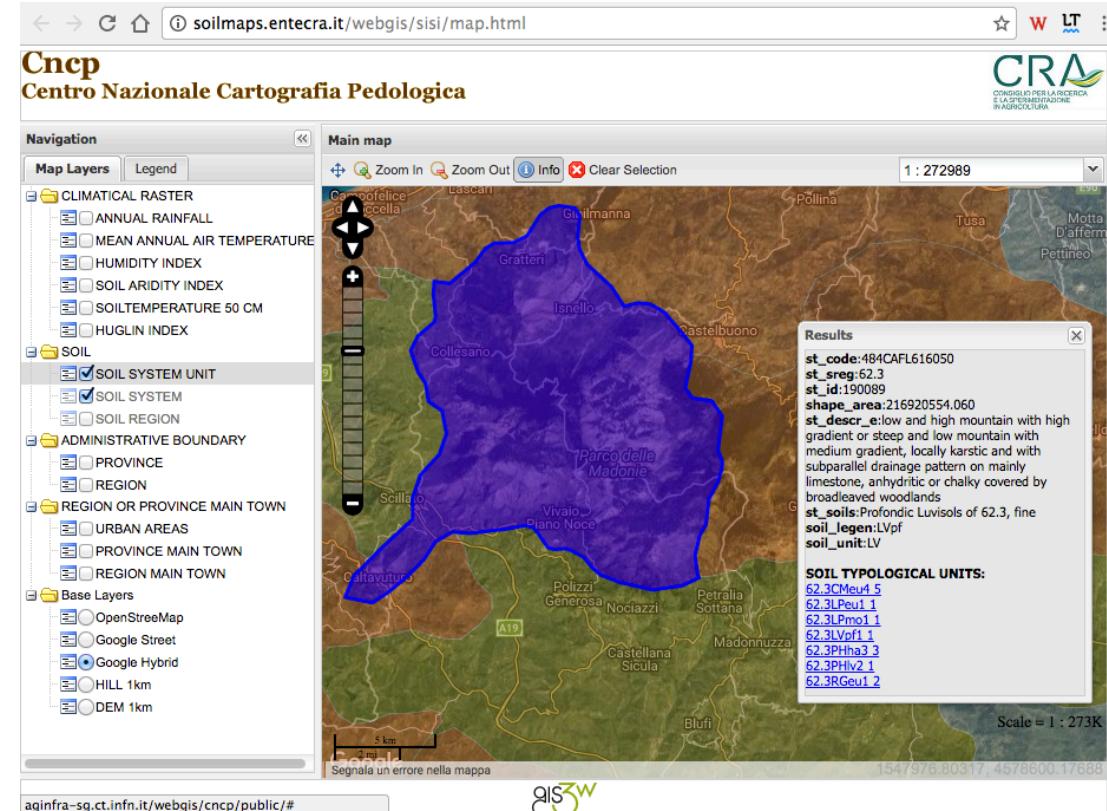
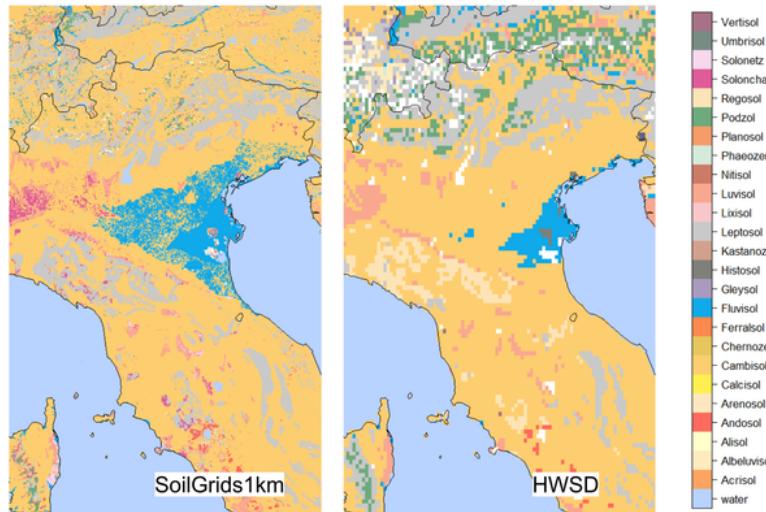


# Example of classified soil maps (on line data)

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

**WRB (2006)**  
Haplic Vertisols (Eutric, Endoskeletal)  
Haplic Cambisols (Eutric)  
Endoleptic Regosols



<https://soilgrids.org>

<http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/>

<http://soilmaps.entecra.it/webgis/sisi/map.html>

25/06/13

CONFERENZA GARR 2016 30 NOVEMBRE - 2 DICEMBRE 2016, FIRENZE

# The WRB classification structure

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

At the first level, the Reference Soil Groups (RSGs), classes are differentiated according to characteristic soil features produced by primary processes.

At the second level (RSGs with qualifiers), soil features resulting from secondary soil-forming processes are used to further specify the primary characteristics.

The classification system allows for the combined use of a set of principal and supplementary qualifiers (qualifiers, either with prefixes or suffixes). This mechanism allows for a very precise characterization and classification of individual soil profiles.

# Implementing soil classification into DBs

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

The screenshot displays a software interface for managing soil data. At the top left, there's a search bar with fields for 'Survey code Type Number' (set to BFJS P 1), 'sec. 1 Original code' (empty), and dropdowns for 'USDA' (Typic Endoaquepts very-fine) and 'WRB' (Calcari Fluvic Gleysols Sulfidic (Protosalic, Protovertic, Clayic)). A 'FILTER' sidebar on the right lists 'STU', 'STS', and 'LC of SS'. Below these are tabs for 'GENERAL CHARACTERISTICS', 'SURFACE CHARACTERISTICS', 'GEOLOGY', 'QUALITIES', and 'CLASSIFICATION'. The 'CLASSIFICATION' tab is active, showing a detailed breakdown of the USDA and WRB classifications. The USDA classification is listed as 'Typic Endoaquepts very-fine, mixed, nonacid, mesic, semiactive'. The WRB classification is listed as 'Calcari Fluvic Gleysols Sulfidic (Protosalic, Protovertic, Clayic)'. Below these are dropdown menus for edit., ord., s. ord., gr. grup., s. grup., text. f., miner., react., temp., activ., and others. Further down, there's a 'Notes' section with radio buttons for 'Soil description form', 'Analysis form', 'Experimental plot', and 'Data quality control'. At the bottom, there's a 'Pedoclimate' section with 'Udometric regime: Xeric' and 'Termometric regime: Thermic'. At the very bottom, there are timestamp fields for 'insert date' (4/27/2015 2:44:20 PM) and 'modify date' (2/16/2016 6:39:58 PM).

Classification: String or Multiple codes?

25/06/13

CONFERENZA GARR 2016 30 NOVEMBRE - 2 DICEMBRE 2016, FIRENZE

# The problem

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

WRB Soil Classification is widely adopted but still not published in open standards and machine readable format

- > data artifacts using WRB show a wide variety of structures and approaches to versioning
- > limiting proper exchange and integration
- > missing Soil Datasets for WRB as LOD

# Missing Soil Datasets for WRB as LOD

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

The screenshot shows the datahub website interface. At the top, there is a navigation bar with links for Datasets, Organizations, About, Blog, Help, and a Search bar. The main content area has a sidebar on the left containing sections for Organizations, Tags, and Formats. The Organizations section is expanded, showing "Italian National In... (2)". The Tags section lists various terms like rdf-format, rdf, lod, format-rdf, country-italy, soil-consumption, hydrogeological ins..., consumo-suolo, and others. The Formats section lists example/rdf+xml and api/sparql. The main content area shows a search bar with the term "soil" and a search icon. Below the search bar, it says "2 datasets found for 'soil'" and "Order by: Relevance". A list of datasets is shown, starting with "Italian National Institute for Environmental Protection and Research, ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale) x". Below this, there are two more dataset entries: "ISPRA - Soil Consumption Indicators in Italy" and "ISPRA - Mitigation measures in Italy for national soil protection". Each entry includes a brief description and links for api/sparql, n-triple, and example/rdf+xml.

**datahub**  
The easy way to get, use and share data

Datasets    Organizations    About    Blog    Help    Search

Home / Datasets

▼ Organizations

Italian National In... (2) x

▼ Tags

rdf-format (2)

rdf (2)

lod (2)

format-rdf (2)

country-italy (2)

soil-consumption (1)

hydrogeological ins... (1)

consumo-suolo (1)

▼ Formats

example/rdf+xml (2)

api/sparql (2)

**Add Dataset**

soil x

2 datasets found for "soil"

Order by: Relevance

Organizations:

Italian National Institute for Environmental Protection and Research, ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale) x

**ISPRA - Soil Consumption Indicators in Italy**  
dati.isprambiente.it - ISPRA Linked Open Data - Soil Consumption Indicators in Italy - a view in the administrative divisions of the country (regions/provinces/municipalities)....

[api/sparql](#) [n-triple](#) [example/rdf+xml](#)

**ISPRA - Mitigation measures in Italy for national soil protection**  
dati.isprambiente.it - ISPRA Linked Open Data - Mitigation measures in Italy for national soil protection

[n-triples](#) [api/sparql](#) [example/rdf+xml](#)

25/06/13

CONFERENZA GARR 2016 30 NOVEMBRE - 2 DICEMBRE 2016, FIRENZE

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

**National Soil Database representative Soil Systems geography**

Posted on : 16/12/2014

Owner:

Consiglio per la Ricerca e la sperimentazione in  
Agricoltura (CRA) - Centro di ricerca per  
l'Agrobiologia e la Pedologia (ABP)

Domain: Agriculture - General/All

Type of data: Soil profiles Soil maps Soil data

Ag-Events Application Profile

[Click to access dataset](#)You need a client for the protocol: [FTP](#) /  
[download](#)**Soil and climate GeoDataBase of Italy**

Posted on : 16/12/2014

Owner:

Consiglio per la Ricerca e la sperimentazione in  
Agricoltura (CRA) - Centro di ricerca per  
l'Agrobiologia e la Pedologia (ABP)

Domain: Agriculture - General/All

Type of data:

Soil profiles Soil climate regimes Soil  
maps Soil data Soil research data[Click to access dataset](#)You need a client for the protocol: [FTP](#) /  
[download](#)**Soils as cultural heritage in Italy**

Posted on : 16/12/2014

Owner:

Consiglio per la Ricerca e la sperimentazione in  
Agricoltura (CRA) - Centro di ricerca per  
l'Agrobiologia e la Pedologia (ABP)

Network(s): AgINFRA

Domain: Economics and Policy

Type of data: Soil profiles Soil data

Geographic Information - Metadata XML

[Click to access dataset](#)You need a client for the protocol: [FTP](#) /  
[download](#)**Soil Information System of Italy**

Posted on : 16/12/2014

Owner:

Consiglio per la Ricerca e la sperimentazione in  
Agricoltura (CRA) - Centro di ricerca per  
l'Agrobiologia e la Pedologia (ABP)

Network(s): AgINFRA

Domain:

Geographical and Regional Information

Type of data: Soil profiles Soil maps Soil data

Geographic Information - Metadata XML

Simple Knowledge Organization System

[Click to access dataset](#)You need a client for the protocol: [REST](#) (any  
type)**Browse datasets**

The RING considers **datasets** only those services that expose data in a machine-processable form (see [here](#)). Services can expose data in various forms: as downloadable files, as feeds, as protocols endpoints etc. For each service, details about datasets are displayed on the right.

Displaying 1 - 6 of 6

starting from the most recently registered

You can do a simple free-text search or you can use the filters  
on the left and on the right to search using advanced criteria.[RSS](#)

25/06/13

CONFERENZA GARR 2016 30 NOVEMBRE - 2 DICEMBRE 2016, FIRENZE

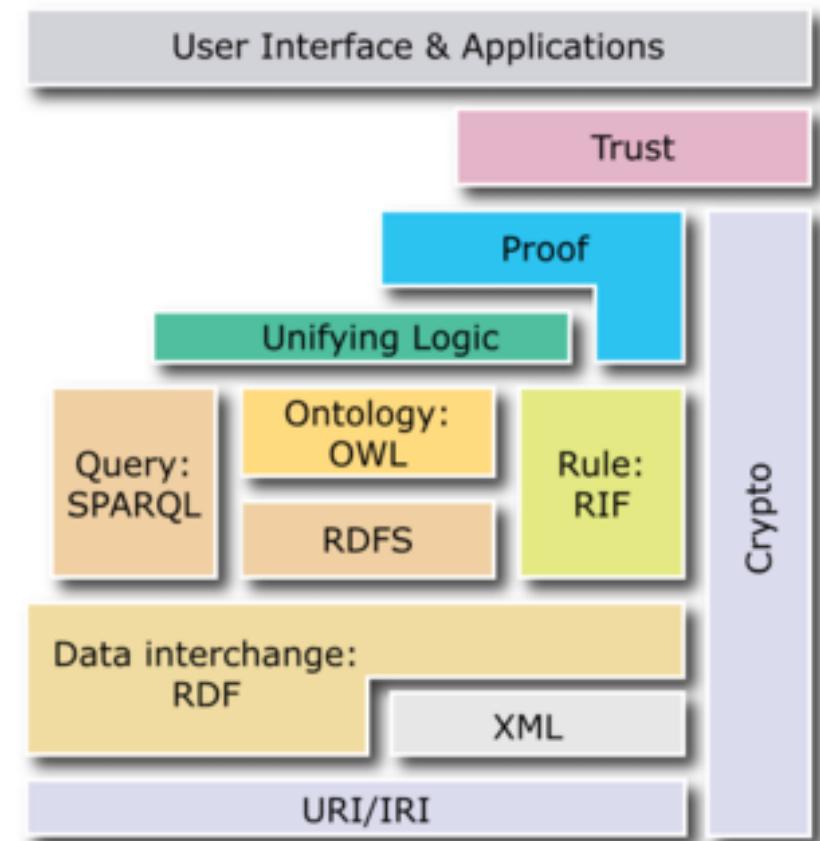
# Goal of this work

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

Propose a way to publish WRB so that it can be reused by information systems in open formats

RDF and SPARQL to describe the resources and query them.



# Resource Description Framework (RDF)

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

RDF Model and Syntax: exposes the structure of the RDF model, and describes a possible syntax.

RDF Schema: exposes the syntax for defining schemas and vocabularies for metadata. RDF triples are composed of subject, predicate and object.

Through RDF triples they represent the resources, their classes and properties and their values.



25/06/13

CONFERENZA GARR 2016 30 NOVEMBRE - 2 DICEMBRE 2016, FIRENZE

# Previous work and limitations

L'Abate      Opening Soil Data through Semantics and Linked Data Vocabularies

artemide.art.uniroma2.it/vocbench2/#Concepts

Signed in as aginfra ( Publisher ) : agINFRA\_Soil\_Terms Global data management Administration About VocBench English RSS feed Preferences Help Sign out

VocBench VERSION 2.4.2 [Build 20160530] (DEVELOPMENT)

Exact word Go Advanced search

Concepts Properties Schemes SPARQL ICV Validation History Concept navigation history Content language

Concepts Show URI Show non-preferred

WRB Reference Soil Group (RSG) (en); UNITA' DEL WORD REFERENCE BASE (it)

- Acrisols (en)
- Albeluvisols (en)
- Alisols (en)
- Andosols (en)
- Anthrosols (en)
- Arenosols (en)
- Calcisols (en)
- Cambisols (en)
- Chernozems (en)
- Cryosols (en)
- Durisols (en)
- Ferralsols (en)
- Fluvisols (en)
- Gleysols (en)
- Greyzems (en)

World Reference Base for Soil Resources (WRB) (en)

Term (1)	Definition (0)	Attribute (0)	Relationship (0)	Alignments (4)	Note (0)	Image (0)	Scheme (1)	Hierarchy
Add new alignment								
Alignment				Value				
skos:broadMatch				<a href="#">http://inspire.ec.europa.eu/featureconcept/SoilBody</a>				
skos:exactMatch				<a href="#">http://lod.nal.usda.gov/nalt/216301</a>				
skos:narrowMatch				<a href="#">http://vocabularies.aginfra.eu/soil</a>				
				<a href="#">http://inspire.ec.europa.eu/document/WRBSpec/</a>				

the vocabulary created within AgINFRA project adopted a classical thesaurus-like structure, which did not reflect the faceted structure of WRB: compositional rules were missing

Legend Proposed Validated Published Revised Proposed deprecated Deprecated Show more

# Assumptions

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

WRB is in fact a faceted classification

A compositional/faceted approach  
associated with vocabularies would  
help solving the problem

Separate soil groups and properties  
from the rules used to combine them  
into a description (classification)

# current work

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

The reusable fragment of the agINFRA soil vocabularies was identified and selected.

Compositional rules for the WRB classification were defined and implemented in k.LAB platform.

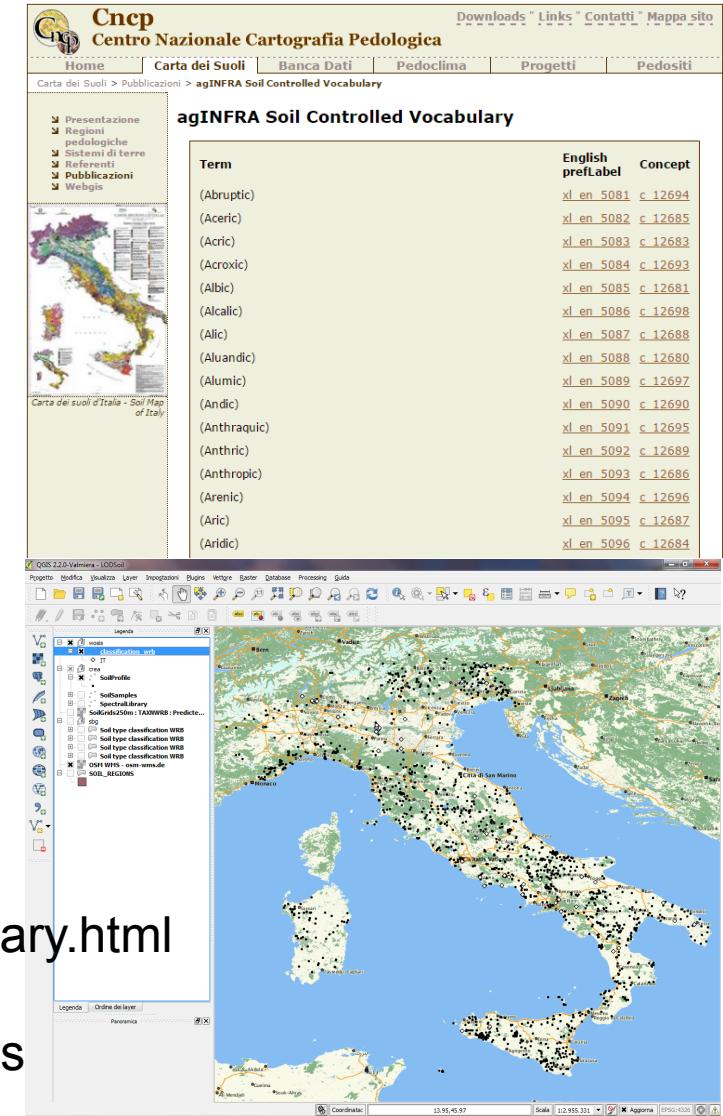
Soil data were exposed trough the Web Feature Service (WFS) standard

agINFRA Soil Controlled Vocabulary

<http://soilmaps.entecra.it/en/soilcontrolledvocabulary.html>

SoilProfile, SoilSamples, SpectralLibrary, Soil\_Regions

<http://93.63.35.107:8080/geoserver/web>



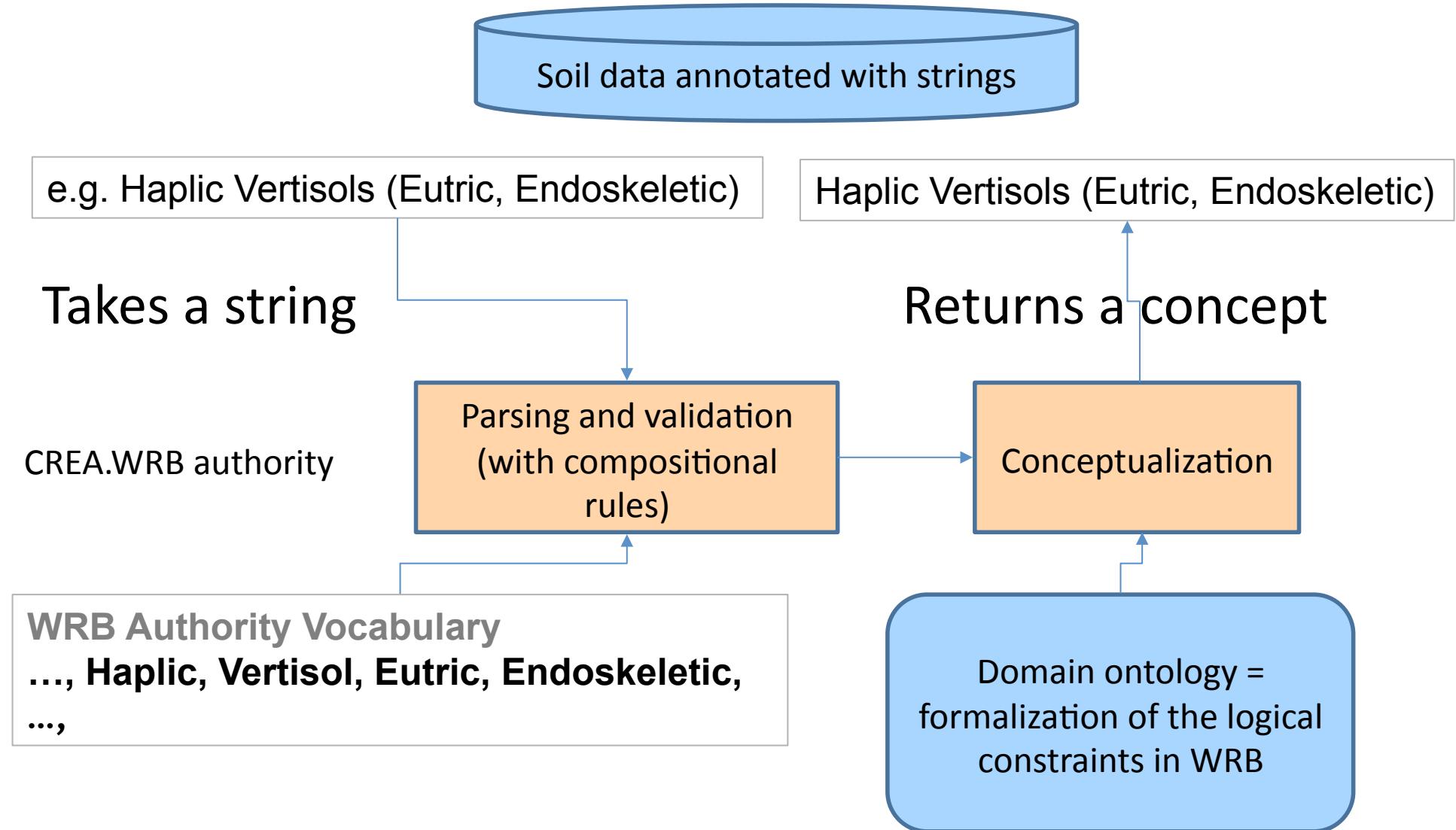
25/06/13

CONFERENZA GARR 2016 30 NOVEMBRE - 2 DICEMBRE 2016, FIRENZE

# Annotating soil data through the WRB authority

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies



# Next steps

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

While the software components of the work reported is near-final state, some issues remain before complete applications can be implemented

A phase of testing is taking place within the k.LAB platform, after which the source code of the WRB parser will be advertised as an independent service for use also within projects not related to k.LAB

# Conclusions

L'Abate

Opening Soil Data through Semantics and Linked Data Vocabularies

WRB vocabulary is aspected to become an “authority list” of concepts, to be reused by a variety of third party applications.

The use of WRB authorities, in conjunction with the compositional grammar implemented in k.LAB will allow for the annotation of diverse data sources in an unambiguous and machine actionable way.

25/06/13

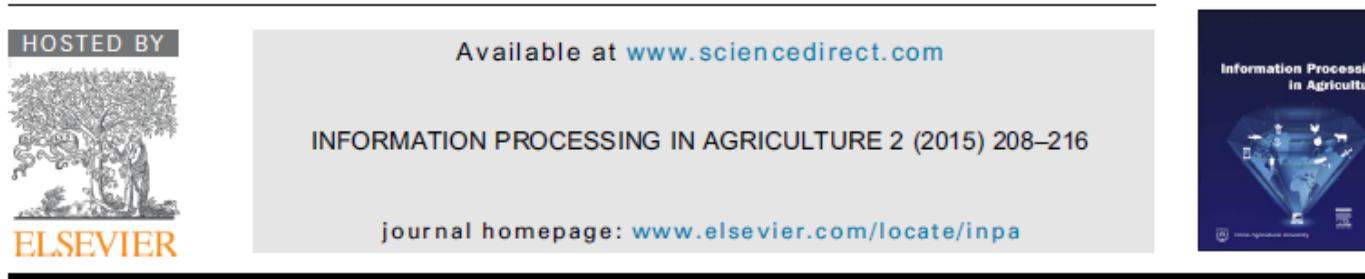
CONFERENZA GARR 2016 30 NOVEMBRE - 2 DICEMBRE 2016, FIRENZE

# Thanks!



Opening Soil Data through Semantics and Linked Data Vocabularies

We acknowledge the support of the 7th FP agINFRA project in previous work. The work described was performed with time donated by the authors.



## Exposing vocabularies for soil as Linked Open Data



Giovanni L'Abate<sup>a,\*</sup>, Caterina Caracciolo<sup>c</sup>, Valeria Pesce<sup>b</sup>, Guntram Geser<sup>e</sup>,  
Vassilis Protonotarios<sup>d</sup>, Edoardo A.C. Costantini<sup>a</sup>

<sup>a</sup> Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA), Agrobiology and Pedology Research Centre, Firenze, Italy

<sup>b</sup> Global Forum on Agricultural Research, Rome, Italy

<sup>c</sup> Food and Agriculture Organization of the United Nations, Rome, Italy

<sup>d</sup> University of Alcalá, Alcalá de Henares, Spain

<sup>e</sup> Salzburg Research, Salzburg, Austria

This presentation is made available under both the Italian Open Data License (IODL v2.0) [LicenseLogo] <http://www.dat.gov.it/iodl/2.0/> and Creative Commons License version 3.0 (CC-BY) [LicenseLogo]

### How to cite this presentation:

L'Abate, Caracciolo, Villa, Costantini (2016): Opening Soil Data through Semantics and Linked Data Vocabularies. GARR2016. PowerPoint presentation. Approver (Year). Identifier1

1 [https://www.popdata.bc.ca/publishing\\_research\\_materials/citingsources](https://www.popdata.bc.ca/publishing_research_materials/citingsources)