

Interoperabilità dei dati pedologici: dalla classificazione internazionale (WRB) ad una semantica formale

Giovanni L'Abate¹, Caterina Caracciolo², Ferdinando Villa³, Edoardo Costantini¹

¹Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, ²Organizzazione delle Nazioni Unite per l'Alimentazione e l'Agricoltura (FAO), ³Centro basco per i Cambiamenti Climatici (BC3); IKERBASQUE, fondazione basca per la scienza

Abstract. Il World Reference Base for Soil Resources (WRB) è uno dei principali sistemi internazionali di classificazione dei suoli, ampiamente usato per classificare i profili di suolo osservati e generare mappe sulla distribuzione dei diversi tipi di suolo. Tuttavia, il WRB non è ancora disponibile in un formato standard, aperto e processabile automaticamente, in quanto viene pubblicato in formato elettronico ma la sua effettiva codifica di dati e metadati è lasciata ai singoli amministratori dei sistemi informatici che la utilizzano. Di conseguenza, gli applicativi che utilizzano questa classificazione mostrano una grande varietà di convenzioni di codifiche, strutture dati ed approcci al controllo delle versioni, che interferisce con il corretto scambio e l'integrazione dei dati machine based. Questo problema si applica ai dati pedologici nonché per altri insiemi di dati e domini che si riferiscono alla scienza del suolo, dalle scienze fisiche e naturali, a quelle economiche e sociali. Con questo contributo si discute sull'uso della classificazione WRB applicando principi di apertura, interoperabilità e correttezza semantica, destinate a sostenere uno scambio di dati affidabile nel campo scientifico ed, in ultima analisi, politico, al fine di semplificare il processo decisionale tramite l'ottimizzazione delle informazioni esistenti.

Keywords. Linked Open Data, World Reference Base for Soil Resources, Vocabolari, Normativa INSPIRE, Servizi web

Introduzione

Il suolo, strato più superficiale della terra, costituisce il substrato per tutte le attività umane, svolge un ruolo fondamentale nel settore agricolo, influisce sul clima, sul controllo dei deflussi superficiali e su molti altri processi biologici. La disponibilità di dati pedologici affidabili e univocamente interpretabili da diversi sistemi informativi è di fondamentale importanza per la comunità scientifica e per la definizione di politiche locali e comunitarie. In questo lavoro si presenta la proposta di un formato di pubblicazione e uso delle classificazioni pedologiche che ne permetta il riuso in diversi sistemi informativi e che elimini le possibilità di fraintendimento tradizionalmente generate dalle implementazioni ad-hoc di tali classificazioni.

Le tassonomie pedologiche riassumono sinteticamente i dati analitici mediante l'assegnazione di porzioni del suolo (caratteri, orizzonti e profili) a una determinata classe gerarchica del siste-

ma adottato. Tra queste, il World Reference Base for Soil Resources (WRB) è uno dei sistemi più ampiamente utilizzati, insieme alla USDA Soil Taxonomy (SOIL SURVEY STAFF, 2014). Il WRB è riconosciuto dall'Unione Internazionale delle Scienze sul suolo (IUSS) e dalla FAO, ed è il risultato di una collaborazione internazionale coordinata dall'International Soil Reference and Information Centre (ISRIC), che ne pubblica aggiornamenti regolari (1998, 2006, 2014).

Il WRB interpreta la morfologia del suolo principalmente come espressione dei processi pedogenetici. Al primo livello, quello dei Reference Soil Groups (RSGs), le classi sono differenziate in base a caratteri prodotti dai processi primari, salvo nel caso in cui risultino di importanza primaria materiali parentali specifici. Al secondo livello (RSGs con qualificatori) le caratteristiche derivanti da processi pedologici secondari sono utilizzate per specificare ulteriormente i caratteri primari. In molti casi sono pre-

si in considerazione tutti quei caratteri del suolo che hanno un effetto significativo sulla destinazione di uso dei terreni. Il sistema di classificazione consente l'uso combinato di un insieme di capitale e qualificatori supplementari (qualificatori ai quali è possibile associare sia prefissi che suffissi). Questo meccanismo consente una caratterizzazione molto precisa dei singoli profili pedologici, nonché di collegare efficacemente i vari sistemi di classificazione nazionali.

Ad oggi il WRB è stato pubblicato in formati elettronici principalmente orientati alla lettura e alla stampa (pdf, html e tramite applicazione Android), mentre la definizione delle strutture di database e gli schemi di dati per uso in sistemi di gestione dell'informazione viene lasciata a singoli utenti e istituzioni. La conseguenza è una mancanza di chiarezza quando si cerca di utilizzare i dati in applicazioni diverse da quelle originali, e una ancora più limitata interoperabilità quando devono essere integrati dati provenienti da diversi domini. Ciò limita anche la semplicità di aggiornamento delle classificazioni implementate quando nuove versioni vengono pubblicate.

L'obiettivo finale di questo lavoro è quello di promuovere l'interoperabilità di dati e servizi. A questo scopo proponiamo un vocabolario formale e pubblico che esprima le entità fondamentali della classificazione WRB, rispettando i principi di apertura dei dati e di coerenza semantica dei loro metadati. Il nostro obiettivo è quindi quello di consentire alle organizzazioni di gestione dei dati pedologici di fornire servizi compatibili con un modello di dati condiviso e senza ambiguità. Questi dati standardizzati per una varietà di scopi, dalla ricerca scientifica a fini commerciali allo sviluppo di politiche locali o comunitarie. Esempi di applicazioni possibili includono: l'analisi dei servizi ecosistemici, per i quali gli analisti hanno bisogno di ottenere gamme di proprietà del suolo, come pH, salinità o contenuto di carbonio dalle mappe nazionali, continentali o mondiali disponibili; la scelta di varietà colturali adatte alla coltivazione in una certa area; la valutazione economica di un terreno; nell'ambito della ricerca agricola, come pure nella ricerca in ecologia e nelle scienze natu-

rali e la realizzazione di mappe digitali dei suoli. Il problema posto da vocabolari locali utilizzati per descrivere i dati analitici relativi al suolo è stato affrontato in passato da alcuni degli autori (L'Abate et al., 2015) già all'interno del progetto agINFRA. Tuttavia il lavoro del 2015 presenta una limitazione, data dal fatto che il vocabolario creato era basato su di una struttura del database pedologico compatibile con la normativa INSPIRE (INSPIRE Thematic Working Group Soil, 2015) ma non perfettamente compatibile con l'approccio faceted su cui si basa invece la classificazione WRB. Nel presente lavoro ci proponiamo di affrontare questa limitazione chiarendo la semantica della classificazione WRB e descrivendone i suoi vincoli compositivi. L'implementazione proof-of-concept qui presentata è stata realizzata utilizzando le funzionalità della piattaforma open source di modellazione semantica k.LAB (Villa et al., 2014) che, come descritto più avanti, permette sia di definire ontologie che di utilizzarne gli assiomi per annotare e convalidare l'informazione contenuta in dati e modelli.

1. Metodi e Risultati

L'idea alla base del nostro lavoro è che i fondamentali gruppi tassonomici del suolo e le loro proprietà debbano essere descritti separatamente dalle regole utilizzate per la loro combinazione e uso nel lavoro classificatorio vero e proprio. A nostro parere la parte concettuale e terminologica va distinta da quella classificatoria: la prima va resa pubblica come vocabolario adatto al riuso in applicazioni informatiche, mentre le regole classificatorie vanno espresse attraverso una grammatica formale, in modo da rendere le due componenti del sistema classificatorio azionabili programmaticamente in applicazioni informatiche e riutilizzabili dalla comunità di utenti. Per quanto riguarda la parte concettuale e terminologica, siamo partiti dal vocabolario prodotto all'interno del progetto agINFRA, selezionandone la parte pertinente alla versione del 2006 del WRB. Per la parte applicativa e per la definizione delle regole classificatorie abbiamo invece usato la piattaforma informatica k.LAB. La versione WRB scelta è quella del 2006 per-

Database	Owner	Feature/Raster
SoilProfile	CREA	Point
SoilSamples	CREA	Point
SpectralLibrary	CREA	Point
WOSIS	ISRIC	Point
Soil_Regions	CREA	Polygon
Soil type classification WRB	SBG	Polygon
SoilGrids250m	ISRIC	Raster

Tab. 1 Banche dati online che adottano la classificazione WRB 2006, compatibili con il servizio WFS dell'Open Geospatial Consortium Web Feature

ché ampiamente utilizzata in banche dati pubblicate on-line tramite lo standard Web Feature Service dell'Open Geospatial Consortium (Tabella 1), sebbene l'ultima versione sia del 2014. Un vocabolario di 32 RSGs, qualificatori, prefissi e suffissi (Tabella 2) utilizzati in WRB 2006 è stato estratto dai vocabolari agINFRA già pubblicati in precedenza (Figura 1). Questo vocabolario include tutti i termini al primo e secondo livello del WRB ed utilizzabili per sviluppare le classi al terzo livello.

Le regole compositive della classificazione dei suoli WRB 2006 sono state analizzate e definite in una grammatica formale. Un software parser è stato scritto per la piattaforma k.LAB in grado di interpretare le specifiche della grammatica, analizzare i termini che vi appaiono e convalidarli in base al vocabolario agINFRA. In seguito alla convalida il software applica le regole di composizione secondo la grammatica, produce la definizione di un concetto che viene poi legata alle ontologie formali utilizzate in k.LAB (Figura 2). Il parser così costruito permette la creazione di concetti per tutti i tipi di suolo in base ad ontologie definite dall'utente.

	Term	URI
Reference Soil Groups	Acrisols	[<http://soilmaps.entecra.it/kos/c_768>]
	Albeluvisols	[<http://soilmaps.entecra.it/kos/c_769>]
	Alisols	[<http://soilmaps.entecra.it/kos/c_770>]
	Andosols	[<http://soilmaps.entecra.it/kos/c_773>]
	Anthrosols	[<http://soilmaps.entecra.it/kos/c_774>]
	Arenosols	[<http://soilmaps.entecra.it/kos/c_775>]
	Calcisols	[<http://soilmaps.entecra.it/kos/c_12420>]
	Cambisols	[<http://soilmaps.entecra.it/kos/c_777>]

Qualifiers	Abruptic	[<http://soilmaps.entecra.it/kos/c_12929>]
	Aceric	[<http://soilmaps.entecra.it/kos/c_12920>]
	Acric	[<http://soilmaps.entecra.it/kos/c_12918>]
	Acroxic	[<http://soilmaps.entecra.it/kos/c_12928>]
	Albic	[<http://soilmaps.entecra.it/kos/c_12933>]
	Alcalic	[<http://soilmaps.entecra.it/kos/c_12453>]
	Alic	[<http://soilmaps.entecra.it/kos/c_12923>]
...	...	
Suffixes	Bathy	[http://soilmaps.entecra.it/kos/c_801]
	Cumuli	[http://soilmaps.entecra.it/kos/c_802]
	Endo	[http://soilmaps.entecra.it/kos/c_803]
	Epi	[http://soilmaps.entecra.it/kos/c_804]
	Hyper	[http://soilmaps.entecra.it/kos/c_805]
	Hypo	[http://soilmaps.entecra.it/kos/c_806]
	Ortho	[http://soilmaps.entecra.it/kos/c_807]
	Para	[http://soilmaps.entecra.it/kos/c_808]
	Proto	[http://soilmaps.entecra.it/kos/c_809]
Thapto	[http://soilmaps.entecra.it/kos/c_810]	

Tab. 2 Esempio di classi tassonomiche WRB (2006) estratte dal vocabolario agINFRA

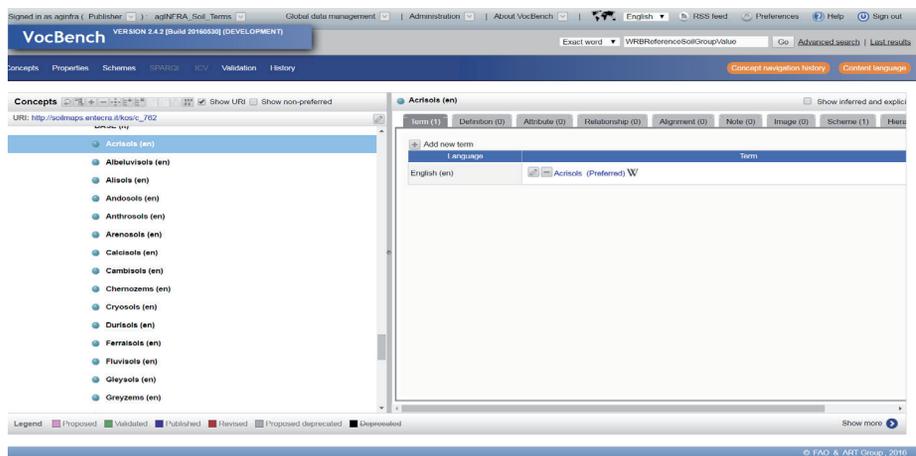


Fig. 1 Frammento del vocabolario agINFRA utilizzato, visualizzato attraverso l'interfaccia web per il suo aggiornamento (VocBench).

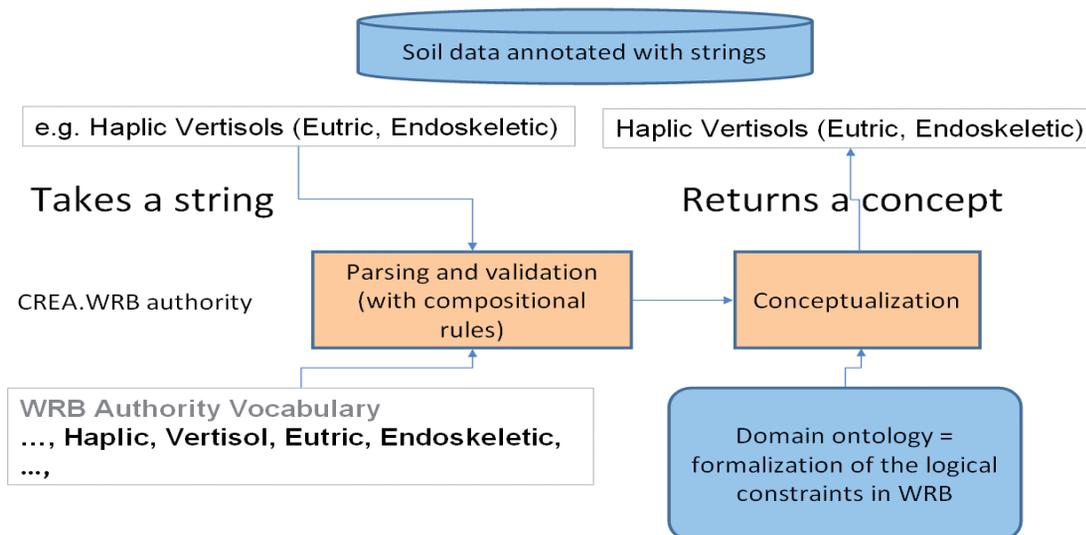


Fig. 2 Diagramma di flusso del processo di interpretazione delle specifiche WRB operato dal parser

Il passaggio dalle definizioni tassonomiche WRB ad una semantica formalizzata si realizza attraverso una componente software (nella terminologia k.LAB, si tratta di una authority, un plug-in che contiene sia il vocabolario che le regole di composizione della classificazione) che usa il vocabolario per convalidare una specifica e da questa produrre assiomi logici che saranno utilizzate dal sistema per integrare le corrispondenti ontologie di dominio.

La piattaforma software k.LAB include funzionalità di convalida (ovvero la componen-

te autorità descritta sopra) per varie classi di identità utilizzate negli studi dei sistemi socio-ambientali. Tra queste ricordiamo GBIF per le informazioni tassonomiche biologiche, IUPAC per gli identificatori chimici, AGROVOC per i termini agricoli, e WRB per la tassonomia del suolo. Il principio per tutte queste componenti è che un identificatore nel formato riconosciuto per l'autorità viene analizzato, convalidato e quindi trasformato in assiomi da incorporare nelle ontologie. Poiché gli assiomi sono identici per ogni utente, l'interoperabilità è resa possi-

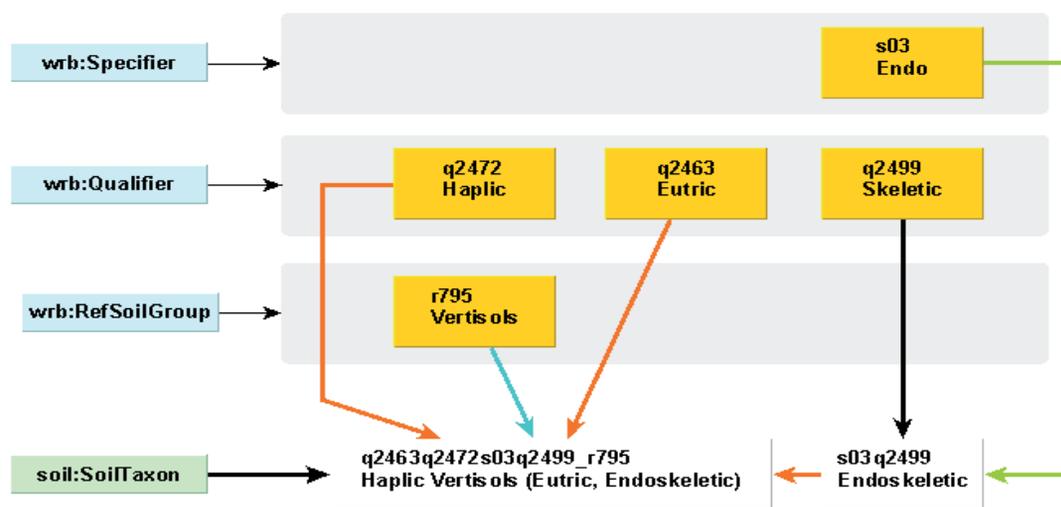


Fig. 3 Struttura logica derivante da una analisi della specifica WRB “Haplic Vertisols (Eutric, Endoskeletal)” da parte dell’autorità SOIL.WRB in k.LAB

```

/*
 * Example 1: aliasing
 */
identity HaplicVertisolTest is soil:SoilTaxon
  identified as "Haplic Vertisols (Eutric, Endoskeletal)" by SOIL.WRB;

/*
 * Example 2: annotation of a single observation
 */
thing HaplicVertisolSample
  is infrastructure:SamplingSite identified as "Haplic Vertisols (Eutric, Endoskeletal)" by SOIL.WRB;

/*
 * Example 3: annotation of dataset. This model will output the soil taxon as
 * a WRB concept across the context queried.
 */
model wfs(service="http://.../ows?service=WFS",
  id="SISI:SoilProfile",
  attr = "c_12395")
  as classify soil:SoilTaxon identified by SOIL.WRB;

/**
 * Example 4: use of reasoning to establish the presence of shallow soils in a region
 * The 'observing' line establishes a dependency on SoilTaxon (resolved by the model
 * above). The reasoner invoked in the expression will return true or false, indicating
 * presence or absence of the observable.
 */
model presence of soil:ShallowSoil within earth:Region
  observing (classify soil:SoilTaxon) named soiltaxon
  on definition set to [
    soiltaxon.hasQualifier("Leptic")
  ];

```

Fig. 4 Esempi di uso dell'autorità SOIL.WRB in k.IM. Fare riferimento al testo per la spiegazione di ogni esempio

bile per utenti indipendenti e non coordinati, anche quando i vocabolari siano troppo estesi per consentirne l'utilizzo pratico in una ontologia (è questo il caso delle tassonomie biologiche).

Si consideri l'esempio illustrato nella Figura 3. La specifica "Haplic Vertisols (Eutric, Endoskeletal)" (Taxon WRB 2006) è accettata dal parser e trasformata in assiomi che ne specificano la struttura logica. I rettangoli presenti in Figura 3 rappresentano concetti, di colori diversi a seconda della loro provenienza: in giallo i concetti elementari dal vocabolario WRB; in blu i concetti del WRB che costituiscono l'ontologia fondamentale creata da parte dell'autorità, in verde il concetto SoilTaxon dalla ontologia di dominio alla quale è collegata l'autorità. Le frecce rappresentano relazioni logiche: freccia nera per la relazione is-a, arancione per 'hasWRBQualifier', blu per 'hasWRBReferenceSoilGroup' e verde per 'hasWRBSpecifier'.

Il linguaggio k.IM permette l'uso di autorità in molti contesti differenti. La Figura 4 mostra alcuni esempi di utilizzo della autorità SOIL.WRB per costruire descrizioni formali da utilizzare nel modello dati e nelle annotazioni.

L'esempio in "Example 1" (Figura 4) mostra la definizione di un concetto da WRB, dove HaplicVertisolTest rappresenta un alias per il

concetto definito, da usarsi per facilità di lettura ed uso. Quando HaplicVertisolTest viene utilizzato in un modello o per l'annotazione locale di dati, il concetto ad esso associato consente la piena elaborazione e validazione semantica. Se un utente diverso annota la stessa definizione del concetto con un diverso identificativo, e poi condivide artefatti che utilizzano tale annotazione con il primo utente, i prodotti sono riconosciuti come equivalenti. L'esempio in "Example 2" mostra il caso in cui un concetto viene definito per un campione di terreno (oggetto fisico); la definizione crea un nuovo concetto per un campione di suolo con il taxon WRB specificato, che può successivamente essere utilizzato ad esempio per descrivere il contenuto di una raccolta di campioni. Nell'esempio "Example 3" una fonte di dati (in questo caso i dati vettoriali resi disponibili attraverso un servizio WFS) è annotata come contenente specifiche WRB nell'attributo c_12395. Ogni interrogazione per un'osservazione del soil:SoilTaxon in un contesto compatibile con la copertura del set di dati invocherà l'autorità SOIL.WRB che tradurrà le stringhe contenute nel dataset in concetti. L'esempio "Example 4" mostra anche come le proprietà del concetto possano essere utilizzate per interrogare la base di conoscenza durante l'ese-

cuzione del modello. Ponendo il concetto Soil-Taxon come dipendenza, un dataset appropriato verrà identificato (esempio 3) e la struttura logica del concetto viene analizzata nell'espressione on definition, stabilendo la presenza di una particolare tipologia di suolo (ShallowSoil) nel caso in cui sia presente il qualificatore Leptic.

2. Conclusioni e sviluppi futuri

Nel lavoro riportato in questo articolo abbiamo identificato e selezionato la porzione riutilizzabile del vocabolario agINFRA contenente le categorie chiave della tassonomia WRB e abbiamo codificato le regole compositive per la classificazione WRB utilizzando la piattaforma open source k.LAB che è stata così estesa con le componenti software corrispondenti. Il vocabolario WRB così composto può essere riutilizzato in una varietà di applicazioni di terze parti. La pubblicazione di una autorità WRB e la definizione della grammatica compositiva implementata in k.LAB ha consentito la lettura di diverse fonti dati (virtualmente tutte quelle pubblicate online) in modo non ambiguo e comprensibile alle macchine. La fase di test, svolta all'interno della piattaforma k.LAB, ha consentito di pubblicare il codice sorgente del parser WRB come un servizio indipendente per l'utilizzo anche all'interno di progetti non relativi ai dati k.LAB. I dati pedologici (Observed Soil Profiles) relativi alle tipologie italiane sono stati pubblicati online secondo lo standard interoperabile (WFS) in conformità con quanto definito da INSPIRE Thematic Working Group Soil Data Specification Technical Guidelines.

Gli sviluppi futuri saranno diretti da un lato, all'esposizione dei vocabolari attraverso un endpoint stabile, affidabile, dall'altro ad includere nei vocabolari una descrizione formale dei significati di concetti e oggetti del WRB, anziché esprimerne esclusivamente i termini.

Riconoscimenti

Gli autori riconoscono il fondamentale sostegno del 7° PQ, progetto agINFRA nel precedente lavoro. Il presente lavoro è stato eseguito con tempo donato dagli autori al di fuori di qualsiasi finanziamento.

Riferimenti bibliografici

Villa F., Bagstad K. J., Voigt B., Johnson G.W., Portela R., Honzák M., Batker D. (2014), A Methodology for Adaptable and Robust Ecosystem Services Assessment. PLoS ONE 9(3): e91001. doi:10.1371/journal.pone.0091001

INSPIRE Thematic Working Group Soil (2013), D2.8.III.3 INSPIRE Data Specification on Soil – Draft Technical Guidelines. http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SO_v3.0rc3.pdf

IUSS Working Group WRB (2006), World reference base for soil resources 2006. World Soil Resources Reports No. 103. FAO, Rome. <http://www.fao.org/3/a-a0510e.pdf>

IUSS Working Group WRB (2007), World Reference Base for Soil Resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome. http://www.fao.org/fileadmin/templates/nr/images/resources/pdf_documents/wrb2007_red.pdf

IUSS Working Group WRB (2015), World Reference Base for Soil Resources 2014, update 2015 International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome. <http://www.fao.org/3/a-i3794e.pdf>

L'Abate G., Caracciolo C., Pesce V., Geser G., Protonotarios V., Costantini E.A.C. (2015), Exposing vocabularies for soil as Linked Open Data. Info Proc Agri <http://dx.doi.org/10.1016/j.inpa.2015.10.002>

OrlovDO (2016), WRB tool for Android, Google Play. <https://play.google.com/store/apps/details?id=ru.orlovdo.wrbv2&hl=en>

Panagiotis A. V. (2005), Web Feature Service Implementation Specification, Open Geospatial Consortium Inc. https://portal.opengeospatial.org/files/?artifact_id=8339

Soil Survey Staff (2014), Keys to Soil Taxonomy, 12th ed. USDA-Natural Resources Conservation Service, Washington, DC.

Giovanni Labate

giovanni.labate@crea.gov.it



Laurea in Scienze ambientali, indirizzo terrestre il 25/10/1999 presso l'Università Cà Foscari di Venezia, attualmente è Tecnologo III livello TD dal 2013 presso il CREA - Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Firenze.

Caterina Caracciolo

caterina.caracciolo@fao.org



Specialista in information technology presso la FAO. Ha precedentemente lavorato presso il Consiglio Nazionale delle Ricerche (CNR) di Pisa e presso l'Università di Amsterdam, dove ha conseguito un dottorato di ricerca in informatica.

Ferdinando Villa

ferdinando.villa@bc3research.org



Dottore in Ecologia nel 1993 presso l'Università di Parma, attualmente è professore presso Ikerbasque, Basque foundation for science. Ha condotto ricerca interdisciplinare in ecoinformatica ed ecological economics presso le Università di Maryland e Vermont.

Edoardo Costantini

edoardo.costantini@crea.gov.it



Primo ricercatore presso il CREA - Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Firenze. Docente di Pedologia e Geopedologia nel Dipartimento di Scienze della Terra dell'Università degli Studi di Siena dal 1999 al 2008.