

Armonizzazione, arricchimento semantico e condivisione interoperabile di dati geospaziali multi dimensionali per la pianificazione e gestione del paesaggio rurale

Simone Lanucara¹, Paola Carrara¹, Alessandro Oggioni¹ e Giuseppe Modica²

¹ Consiglio Nazionale delle Ricerche, Istituto per il Rilevamento Elettromagnetico dell'Ambiente

² Dipartimento di Agraria, Università "Mediterranea"

1. Introduzione

La caratterizzazione dinamica del paesaggio rurale è effettuata attraverso l'interpretazione di immagini satellitari/aeree storiche e la successiva produzione, elaborazione, analisi e mappatura di dati geospaziali. Tradizionalmente, mappe e dati erano creati attraverso strumenti software desktop proprietari e condivisi su supporti cartacei. Oggi, grazie alla diffusione di Internet, strumenti software open source, standard internazionali per la condivisione dei dati, ricercatori e professionisti possono condividere dati ed informazioni nel *World Wide Web* (WWW). Le informazioni scientifiche ed i relativi dati sui paesaggi rurali e sui loro cambiamenti possono essere condivise sotto forma di conoscenza scientifica in diversi modi: pubblicazioni scientifiche, condivisione di archivi e attraverso servizi web. Queste diverse metodologie di condivisione della conoscenza: favoriscono processi pianificatori partecipati e tracciabili, offrono agli *stakeholders* ed ai decisori le informazioni necessarie per comprendere le dinamiche del paesaggio, permettono ad altri ricercatori di confrontare la dinamica del paesaggio di aree differenti. Per permettere la condivisione dei dati (mappe) e metadati nel Web una soluzione ampiamente diffusa è lo sviluppo di cyber infrastrutture come le *Spatial Data Infrastructures* (SDI) (Foley e Maynooth, 2009) che permettono di ottenere una interoperabilità a diversi livelli. I servizi ed i relativi standard dell'*Open Geospatial Consortium* (OGC) permettono la distribuzione interoperabile di: Web Map Service (WMS) per mappe e relative legende, Styled Layer Descriptor (SLD) per la definizione di stili grafici di dati e legende, Web Feature Service (WFS) per la condivisione di dati vettoriali, Web Coverage Service (WCS) per la condivisione di dati raster e Catalogue Service for the Web (CSW) per la condivisione di cataloghi di metadati.

In termini generali, le SDI possono essere definite come piattaforme tecnologiche per la gestione, ricerca, condivisione ed utilizzo di dati e metadati geospaziali. Le SDI forniscono un ambiente ideale per connettere le applicazioni ai dati, influenzando allo stesso tempo la creazione dei dati stessi e lo sviluppo di applicazioni basate su appropriati standard e procedure. Varie ed ampiamente diffuse solo le soluzioni adottate per implementare le architetture SDI (Steiniger e Hunter, 2012). Le soluzioni possono essere proprietarie, ad esempio @ESRI (Maguire, 2008), FOSS4G o ibride, in particolare lo sviluppo di applicazioni FOSS4G è supportato dall'ampia disponibilità di suite software che rendono possibile condividere mappe, dati e metadati (Brovelli et al., 2016).

Questo approccio di condivisione, pur interoperabile, comunque difetta nell'armonizzazione dei dati e nell'abilitazione semantica (Lanucara et al., 2017), un divario che rende difficile la ricerca, comparazione ed analisi di dati ottenuti da diverse fonti. L'interoperabilità, infatti, può essere suddivisa in quattro categorie principali: sistemica, sintattica, schematica e semantica. L'interoperabilità semantica facilita la trasformazione dei dati in conoscenza all'interno delle infrastrutture tecnologiche.

A riguardo negli ultimi 15 anni, la Commissione Europea, tramite la Direttiva 2007/2/EC (European Commission, 2007) "Infrastructure for spatial information in Europe" (INSPIRE), ha sviluppato un quadro giuridico e le linee guida per l'interoperabilità tese a facilitare l'accesso transnazionale ai dati geospaziali di diversi domini come geologici, rischi naturali, uso del suolo, copertura del suolo, ecc..

Le linee guida per l'interoperabilità di INSPIRE forniscono gli indirizzi per: la meta datazione, i servizi di ricerca dei dati e metadati, l'accesso e la visualizzazione dei dati, lo scaricamento e la trasformazione dei dati, l'interscambio dei dati, l'utilizzo di liste di termini. Esse costituiscono la base per la strutturazione delle regole di attuazione sull'interoperabilità dei set di dati territoriali; forniscono un quadro di sviluppo

in modo che i dati possano essere riutilizzati in altri ambienti su scala locale, regionale, contribuendo a livello nazionale e globale.

Quindi l'utilizzo delle linee guida INSPIRE è un buon modo per condividere, attraverso il web, la conoscenza scientifica derivante da diversi domini (uso del suolo, idrografia, geomorfologia, geologia). Conoscenza necessaria alla caratterizzazione del paesaggio rurale, al suo cambiamento e come base per sistemi di supporto alle decisioni (Modica *et al.*, 2016) per l'ambiente, la gestione delle risorse e lo sviluppo sostenibile.

In questo articolo descriviamo una metodologia, gli standard internazionali, i vocabolari controllati e gli strumenti software open source che abbiamo utilizzato per condividere sul web, in maniera interoperabile, semanticamente arricchita ed armonizzata, dati multi dimensionali di copertura del suolo (Land Cover - LC) rilevati nel contesto di uno studio multi disciplinare teso a caratterizzare dinamicamente i cambiamenti del paesaggio negli ultimi sessant'anni di una zona del sud Italia denominata "Costa Viola" (Modica *et al.*, 2015).

L'articolo è organizzato nel seguente modo: la prossima sezione (2.1) presenta brevemente i dati prodotti nello studio multidisciplinare; la sezione 2.2 descrive l'architettura software ideata ed implementata per la condivisione dei dati geospaziali e delle mappe. Le sezioni da 2.3 a 2.4 descrivono la metodologia per l'armonizzazione e l'abilitazione semantica dei dati. L'ultima sezione (3) descrive i risultati e conclude l'articolo.

2. Materiali e metodi

Le seguenti sezioni descrivono la metodologia adottata per la condivisione dei dati geospaziali.

2.1. Contesto: i dati

I dati di LC, su cui abbiamo testato la metodologia per condividere sul web, in maniera interoperabile, semanticamente arricchita ed armonizzata, sono stati rilevati nel contesto di uno studio multi disciplinare teso a caratterizzare dinamicamente i cambiamenti del paesaggio negli ultimi sessant'anni di una zona del sud Italia denominata "Costa Viola" (Modica *et al.*, 2015).

I dati rilevati, prodotti con tecniche proprie del telerilevamento e dei sistemi GIS, consistono in sette dataset vettoriali di LC, uno per ogni anno investigato: 1955, 1976, 1989, 1998, 2008, 2012, 2014. Tutti i dataset sono stati registrati in sette distinte tabelle in un database geospaziale PostgreSQL-PostGIS (<http://postgis.net>).

2.2. Architettura software

Per facilitare l'accesso ai dati geospaziali ricavati nel contesto dello studio multidisciplinare e per condividerli in maniera interoperabile con *stakeholders*, decisori e ricercatori abbiamo utilizzato una infrastruttura hardware e software, composta da *software Free and Open Source for Geospatial* (FOSS4G), sviluppata in una precedente ricerca (Modica *et al.*, 2016).

L'architettura adottata è stata ideata per permettere l'interscambio di dati geospaziali sul web e per fornire agli utenti un'applicazione "user-friendly", caratterizzata da accessibilità e versatilità. Seguendo l'ideazione, abbiamo implementato la SDI con un'architettura multi livello (fig. 1) composta da quattro strati con differenti funzioni: 1) repository per memorizzare dati e metadati in un database geospaziale sviluppato in PostgreSQL (<https://www.postgresql.org/>) con estensione PostGIS (<http://postgis.net>); 2) server, composto da GeoServer (<http://geoserver.org>) e GeoNetwork opensource (<http://geonetwork-opensource.org>), per gestire dati e metadati memorizzati e pubblicarli sul web; 3) interfacce standard OGC per pubblicare i dati e metadati in maniera interoperabile; 4) front-end client (WebGIS client) sviluppato in GeoExt (<http://geoext.org>), che fornisce un'interfaccia grafica per visualizzare ed analizzare i dati.

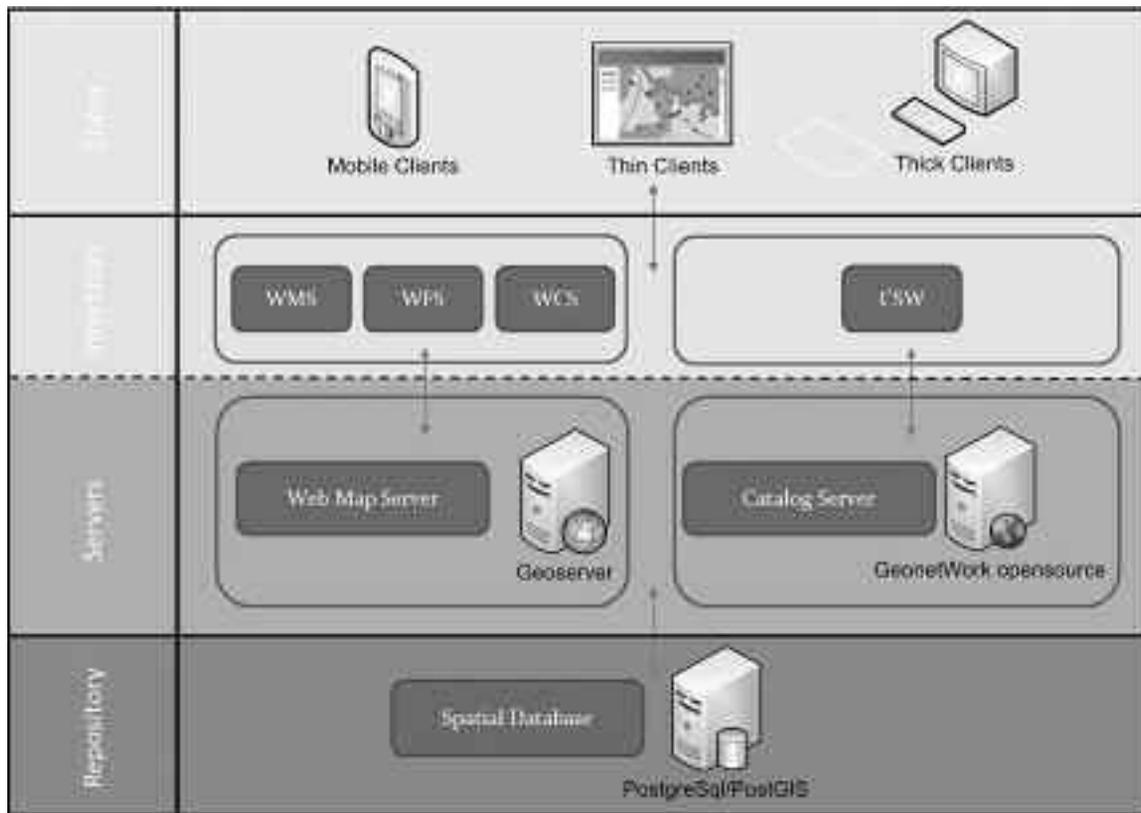


Figura 1. Architettura della SDI

Il livello *repository* identifica l'area di memorizzazione contenente tutti i dati e metadati. L'accesso a questo livello è permesso solo ai componenti hardware e *software* definiti a livello di *storage area network*, per assicurare sia l'integrità e consistenza dei dati stessi sia un alto livello di sicurezza informatica.

Il livello *server* rappresenta l'ambiente *software* che permette di organizzare i dati e renderli disponibili sul Web: fornisce mappe, dati geospaziali e metadati, da una varietà di formati di origine a interfacce standard OGC, in modo che siano disponibili per client come WebGIS e GIS Desktop che gestiscano tali interfacce. In questo modo è possibile archiviare i dati geospaziali ed i relativi metadati in qualsiasi formato di origine. Per tale livello abbiamo utilizzato: GeoServer, un server open source, che riveste un ruolo centrale all'interno della SDI per la condivisione dei dati geospaziali; GeoNetwork, un server open source per la gestione di metadati e condivisione di cataloghi di metadati e che permette all'utente di creare, editare metadati e condividerli secondo gli standard OGC ed in linea con INSPIRE. Dal punto di vista tecnico, GeoServer è una delle principali implementazioni degli standard WFS, WMS o WCS definiti da OGC. In tal modo, ad esempio, gli utenti possono accedere ai dati geospaziali utilizzando qualsiasi client (Desktop, Web, Mobile) che sia capace di gestire tali standard. GeoNetwork opensource abilita la creazione di metadati e la loro condivisione creando un servizio di catalogo esposto tramite le interfacce CSW definite da OGC.

Relativamente al livello front-end, accedendo all'interfaccia grafica del WebGIS, attraverso un qualsiasi browser, l'utente può visualizzare, interrogare, stampare mappe, scenari e grafici. In particolare, per visualizzare i dati e le mappe, lo standard WMS è ampiamente utilizzato per condividere mappe geotematiche e dati georiferiti e per rispondere alle interrogazioni sul contenuto degli stessi.

I summenzionati componenti sono implementati in un ambiente virtuale (macchina virtuale) basato sul sistema operativo Ubuntu e gestiti da un @Microsoft Server Datacenter con Hyper-V.

2.3. Processi di armonizzazione dei dati

I dati relativi alla LC, che hanno fornito un quadro delle tendenze evolutive del paesaggio dal 1955 al 2014, sono stati archiviati come tabelle nel database geospaziale PostGIS e condivisi dalla SDI tramite gli standard WMS-WFS ottenendo un'interoperabilità a livello sintattico. Anche se condivisi, tali dati sono difficilmente riutilizzabili da altri ricercatori e decisori, infatti le informazioni sul LC quali descrizione in

lingua italiana e valori codificati, rendono difficilmente comprensibili i dati. Inoltre, non facendo uso di schemi dati standardizzati, sono difficilmente riutilizzabili o integrabili con altri dati di LC provenienti da altre fonti.

Per colmare queste lacune, armonizzare i dati ed ottenere un interoperabilità schematica, a scala Europea, abbiamo trasformato dati e relativo schema sorgente secondo le specifiche INSPIRE Land Cover (http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_LC_v3.0.pdf); questa procedura è comunemente definita come ETL (Extract, Transform, Load) (Vassiliadis, 2009). Le procedure ETL estraggono i dati geospaziali dalla sorgente, trasformano il modello dati sorgente in un nuovo modello e premettono agli utenti di richiedere il nuovo modello di dati in diversi formati. Queste procedure inoltre permettono di ristrutturare i valori, nome e tipo degli attributi, i sistemi di coordinate, le geometrie ed il tipo di dato.

Per effettuare le procedure ETL abbiamo utilizzato il software open source Hale Studio (HS), già utilizzato dal Data Harmonization Panel di INSPIRE (<http://www.dhpanel.eu/>). HS permette agli utenti di importare schemi sorgenti e dati da differenti sorgenti e formati, importare schemi di output e mappare le trasformazioni attraverso un'interfaccia grafica. Le sorgenti e gli output possono essere file presenti su personal computer (shapefiles, GML, XML, XSD), condivisi online, servizi web (WFS), o presenti su basi di dati che presentino connessioni del tipo Java Database Connectivity (JDBC) application programming interface (API).

Per condividere i dati conformemente alle specifiche INSPIRE Land Cover, attraverso servizi web esposti dal componente GeoServer della SDI, abbiamo utilizzato il modulo di app-schema che permette di mappare lo schema dati sorgente allo schema dati di output tramite un file di mappatura in formato XML.

Utilizzando HS e GeoServer abbiamo armonizzato i dati originari secondo le schema dati INSPIRE Land Cover; il flusso di lavoro adottato è il seguente:

- installazione del plug-in app-schema nel componente GeoServer della SDI;
- importazione degli schemi sorgente e dei relativi dati in HS;
- importazione dello schema INSPIRE Land Cover, in formato XSD (<http://inspire.ec.europa.eu/schemas/lcv/4.0/LandCoverVector.xsd>), in HS;
- mappatura di schemi da sorgente a output facilitata dall'interfaccia grafica di HS;
- esportazione del file di configurazione XML della mappatura di schemi e sua importazione in GeoServer;
- condivisione del servizio WFS secondo lo schema INSPIRE Land Cover.

La seguente sezione descrive, in dettaglio, la metodologia utilizzata per l'armonizzazione, l'arricchimento semantico e la condivisione dei dati di LC.

2.4. Armonizzazione dei dati di copertura del suolo

I dati di LC forniscono informazioni sulla copertura fisica della superficie terrestre, incluse le aree agricole, le foreste, i laghi, le aree naturali e artificiali. Le specifiche dati Land Cover di INSPIRE forniscono le linee guida per l'armonizzazione delle informazioni di LC e per l'implementazione dei servizi web. Esse definiscono due modelli concettuali o modelli core, uno per dati vettoriali e uno per dati raster, rispettivamente LandCoverVector e LandCoverRaster. È da evidenziare che entrambi i modelli prevedono la data di osservazione (observationDate nello schema delle specifiche dati) della LC (ad esempio data di acquisizione della foto aerea o dell'immagine satellitare nei casi in cui il rilevamento remoto è il metodo di osservazione). La data di osservazione fornisce importanti metadati e consente la rappresentazione del cambiamento temporale del LC. I dati di LC possono anche essere collegati a code list (ad es. elenco dei codici CORINE Land Cover), ma la specifica dei dati non prescrive o raccomanda una particolare nomenclatura da utilizzare.

Per armonizzare i nostri dati, registrati su tabelle del database geospaziale, e renderli conformi alle specifiche dati Land Cover di INSPIRE abbiamo mappato lo schema di origine allo schema di destinazione, LandCoverVector. Esso definisce come i dati LC sono descritti da una rappresentazione vettoriale. Questo schema di applicazione contiene quattro classi descritte con un linguaggio di modellazione unificato

(UML): LandCoverDataset, LandCoverUnit, LandCoverObservation e LandCoverValue (fig.3).

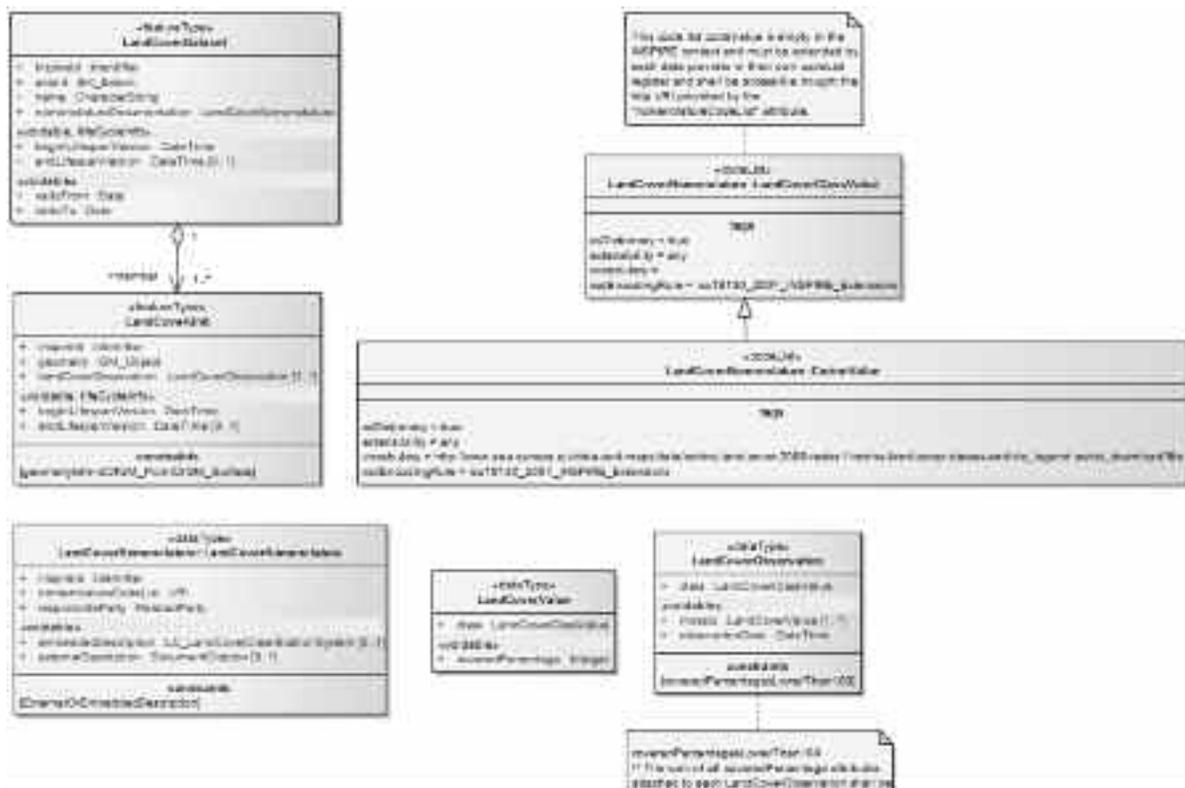


Figura 2. Diagramma UML per la Classe LandCoverVector

Lo schema LandCoverVector modella i set di dati LC (LandCoverDataset) come raccolte di LandCoverUnit. La LandCoverUnit rappresenta una sezione di spazio ben definita ed è definita da una geometria, limitata a punti o superfici, che consente la descrizione spaziale dell'unità. Una LandCoverUnit può supportare più osservazioni distribuite in date differenti; nello schema questo è rappresentato da una relazione uno a molti con la classe LandCoverObservation.

LandCoverObservation definisce l'attributo classe che consente un codice di classificazione, ad esempio codice Corine Land Cover o un altro codice corrispondente a una nomenclatura nazionale, istituzionale o locale. I valori del codice sono definiti nell'elenco di codici definito dalla classe LandCoverClassValue. I nostri dati di origine sono costituiti da diversi set di dati vettoriali di LC registrati su singole tabelle PostGIS della SDI, uno per ogni anno di osservazione: 1955, 1976, 1989, 1998, 2008, 2012 e 2014. Lo schema di queste tabelle è composto da una rappresentazione semplice costituita dai campi: identificativo della geometria, geometria, uso del suolo descrittivo, uso del suolo codificato secondo i codici Corine Land Cover per i diversi livelli da 1 a 4, presenza o meno di superfici terrazzate (fig. 3). Per armonizzare i nostri set di dati alle specifiche dati Land Cover di INSPIRE abbiamo mappato i diversi schemi di origine, uno per ogni tabella, ad un unico schema di destinazione, LandCoverVector, tramite HS. Questa operazione, denominata schema mapping, ha utilizzato diverse tecniche come la corrispondenza relazionale e la mappatura condizionale che comprendono la ridenominazione, il filtraggio, la riclassificazione, la suddivisione, la conversione delle caratteristiche geometriche e delle loro proprietà.

Grazie all'utilizzo delle specifiche INSPIRE, alle procedure ETL, alla mappatura degli schemi e all'utilizzo di vocabolari controllati, abbiamo armonizzato i dati di partenza ottenendo un'interoperabilità di livello semantico a scala Europea.

2.5. Client *EwbGIS*

Consultare, interrogare e visualizzare le informazioni e le mappe multi-temporali di LC da un unico punto di accesso implica lo sviluppo di un client di tipo WebGIS. Infatti, le capacità di un WebGIS di mostrare relazioni geospaziali e cambiamenti spazio-temporali sono ampiamente riconosciute: esse permettono agli utenti di visualizzare ed interrogare le informazioni geospaziali in modo semplice.

Per implementare il WebGIS abbiamo adottato gli ambienti software offerti da GeoExt e da D3js. GeoExt è un potente strumento open source che permette di combinare le librerie di web-mapping di OpenLayers (<https://openlayers.org/>) con l'interfaccia utente di Ext JS (<https://www.sencha.com/products/extjs/>). OpenLayers è una libreria open source JavaScript, utilizzata per visualizzare mappe interattive su un browser web: fornisce le interfacce applicative di programmazione che permettono l'accesso a varie sorgenti di informazioni geospaziali sul Web, ad esempio le interfacce OGC, le mappe di base satellitari (Google Maps, Bing Maps, DigitalGlobe, ecc.), le immagini OpenStreetMap (OSM). Ext JS è un framework JavaScript che permette di costruire applicazioni web per sistemi desktop, tablet e smartphones. D3js è una libreria JavaScript per creare grafici dinamici ed interattivi partendo da dati organizzati, visibili attraverso un browser.

La soluzione adottata permette di acquisire i servizi OGC ed i relativi dati pubblicati dalla SDI e da altre fonti di informazioni geospaziali esistenti sul Web (immagini satellitari DigitalGlobe e mappe OSM). Il client WebGIS permette di visualizzare, ricercare ed interrogare le mappe multi-temporali di LC e di visualizzare i grafici di cambiamento di uso del suolo dall'anno 1950 all'anno 2014.

3. Risultati e conclusioni

I dati di LC forniscono informazioni sulla copertura fisica della superficie terrestre, incluse le aree agricole, le foreste, i laghi, le aree naturali e artificiali. Lo studio dei cambiamenti dinamici di LC è di primaria importanza per decisioni politiche, azioni di pianificazione e successive attività di utilizzo del suolo delle aree rurali.

La condivisione a livello globale dei dati geospaziali e delle informazioni sul LC pone diversi problemi di interoperabilità: sintattica, schematica e semantica. Per risolvere queste problematiche abbiamo sviluppato e testato una metodologia che si basa sull'utilizzo di standard OGC WMS e WFS, specifiche dati Land Cover di INSPIRE e vocabolari controllati. Per implementare la metodologia abbiamo sviluppato, con software FOSS4G una SDI, armonizzato i dati multi-temporali di LC secondo le specifiche dati grazie a procedure di ETL ed abilitato semanticamente i dati stessi grazie al vocabolario fornito dalla *European Environment Agency*. Inoltre, abbiamo progettato ed implementato un client WebGIS che, accedendo ai servizi OGC esposti dalla SDI, permette, in maniera semplice, di visualizzare, interrogare mappe di LC e di visualizzare i grafici di cambiamento di uso del suolo.

Bibliografia

- Brovelli M.A., Minghini M., Moreno-Sanchez R. e Oliveira R., 2016, «Free and Open Source Software for Geospatial Applications (FOSS4G) to Support Future Earth», in *International Journal of Digital Earth* 0(0): 1–19. <https://www.tandfonline.com/doi/full/10.1080/17538947.2016.1196505>.
- European Commission, 2007. «Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)», in *Official Journal of the European Union* 50 (January 2006): 1–14. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>.
- Foley R e Maynooth N., 2009, *Spatial Data Infrastructures*, Elsevier 5: 507–11.
- Modica G., Praticò S., Lanucara S. e Di Fazio S., 2015, «Characterisation and Change Detection of the Agricultural Terraced Landscape of Costa Viola (Calabria , Italy) in View of Its Sustainable Management»

in *EGU General Assembly Conference Abstracts*, 14729.

Lanucara S., Oggioni A., Modica G. e Carrara P., 2017, «Interoperable Sharing and Visualization of Geological Data and Instruments: A Proof of Concept», in *Computational Science and Its Applications – ICCSA 2017*. Lecture Notes in Computer Science, 584–99. http://link.springer.com/10.1007/978-3-319-62401-3_42.

Maguire D.J., 2008, «ArcGIS: General Purpose GIS Software System», in *Encyclopedia of GIS*, Boston, MA: Springer US, 25–31. http://link.springer.com/10.1007/978-0-387-35973-1_68.

Modica G., Pollino M., Lanucara S., La Porta L., Pellicone G., Di Fazio S. e Fichera C.R., 2016, «Land Suitability Evaluation for Agro-Forestry: Definition of a Web-Based Multi-Criteria Spatial Decision Support System (MC-SDSS): Preliminary Results» in *Computational Science and Its Applications - ICCSA 2016*. Lecture Notes in Computer Science, Vol 9788, eds. Osvaldo Gervasi et al. Cham: Springer International Publishing, 399–413. http://dx.doi.org/10.1007/978-3-319-42111-7_31.

Stefan S. e Hunter A.J.S., 2012, «Free and Open Source GIS Software for Building a Spatial Data Infrastructure» in *Geospatial Free and Open Source Software in the 21st Century*, 247–61. <http://www.springerlink.com/index/K373407242815273.pdf>.

Vassiliadis P., 2009, «A Survey of Extract–Transform–Load Technology» in *International Journal of Data Warehousing and Mining* 5(3): 1–27. <http://www.igi-global.com/article/survey-extract-transform-load-technology/3894>.