

UN PROFILO ISO 19115 PER *DATASET* AMBIENTALI GEOGRAFICI E NON-GEOGRAFICI PER LA V.I.A.

Elisabetta PECCOL

Università degli Studi di Udine, Dipartimento di Scienze Agrarie e Ambientali, Via delle Scienze 208, 33100 Udine, tel. 0432 558657, fax 0432 558603, peccol@uniud.it

Riassunto

Nell'ambito di un'indagine finalizzata alla scoperta e descrizione delle fonti di dati ambientali rilevanti per gli studi di V.I.A. in Friuli Venezia Giulia sono stati documentati *dataset* ambientali di tipo geografico e non-geografico sulla base di un profilo di metadati sviluppato *ad hoc* e basato sulla norma ISO 19115:2003 (E). Il profilo è stato calibrato per documentare i dati ad un livello di tipo "discovery", ovvero per fornire informazioni sufficienti per accertare l'esistenza di una risorsa informativa, conoscere alcuni elementi utili per effettuare una scelta sull'opportunità o meno di acquisirla e individuare i soggetti collegati ad essa. Il profilo viene presentato in questo lavoro corredato di un diagramma UML, è disponibile, a richiesta, con una documentazione ISO conforme che comprende un Data Dictionary e diagrammi UML e può essere utilizzato per documentare in modo semplice ed essenziale *dataset* geografici e non-geografici. Il profilo è stato applicato in ambito operativo per documentare 957 *dataset* di interesse per la V.I.A. Si riportano inoltre alcune considerazioni su opportunità e problemi riscontrati nell'applicazione della norma ISO 19115.

Abstract

In a survey aimed at discovering and describing strategic environmental data sources for E.I.A. studies in the Friuli Venezia Giulia region, geographic and non-geographic environmental *datasets* have been documented with a profile of the ISO 19115 standard specifically developed for this purpose. The profile has been maintained at a discovery level, to allow to ascertain the existence of an information resource and to make an informed decision on whether it is appropriate for the planned use. The profile is presented with an UML diagram and is available, on demand, with an ISO compatible documentation including a data dictionary and UML diagrams. The profile has been applied in a survey which documented 957 *datasets*. Finally, some comments on the opportunities and problems encountered while applying the ISO 19115 norm are reported.

Introduzione

I progetti di V.I.A. richiedono un'analisi basata su uno spettro d'informazioni ambientali ampio sia dal punto di vista qualitativo che quantitativo più che in altri comparti del settore ambientale (E.T.C.G., 2002). I dati geografici ed i sistemi informativi territoriali hanno un ruolo decisivo nelle analisi ambientali ed in particolare nelle V.I.A. a causa della dimensione spaziale della maggior parte dei fenomeni ambientali. Da un'indagine condotta recentemente dalla Commissione Europea emerge la rilevanza dei dati geografici nelle V.I.A. e nelle V.A.S. (Vanderhaegen, Muro, 2005). Nella stessa indagine si evidenziano difficoltà sperimentate dagli utenti di dati geografici, consistenti principalmente nell'accesso ai dati, nell'accertamento della disponibilità dei dati e nel riscontro della indisponibilità degli stessi. Anche in regione Friuli Venezia Giulia, l'accesso alle informazioni ambientali da parte degli operatori in ambito V.I.A. viene segnalato tra i punti critici (R.A.F.V.G.

2003), motivabile con una frammentazione e una mancanza di coordinamento delle competenze riguardanti la raccolta e la gestione dell'informazione ambientale unita alla mancanza di servizi come cataloghi di dati e metadati finalizzati ad agevolare il rapido accesso alle informazioni ambientali strategiche. Le principali conseguenze di tale situazione si riscontrano in un aumento del tempo e costo degli studi di impatto ambientale ed in una inferiore qualità e affidabilità nelle previsioni di impatto. Tale situazione si riscontra a scala europea (European Commission, 2003) e a scala locale, come conferma un'analisi condotta su un campione rappresentativo di studi di impatto ambientale nella regione Friuli Venezia Giulia (Piani et al., 2004). Tuttavia l'accesso al dato costituisce solo uno degli aspetti legati all'uso del dato che risulta possibile solo se corredato dei metadati che ne permettono la fruibilità ed un uso corretto. Il progetto "Catalogo delle fonti di dati ambientali a supporto della Valutazione di Impatto Ambientale" ha avuto l'obiettivo di rilevare fonti di dati ambientali prodotti da soggetti pubblici e privati di pertinenza del territorio regionale, rilevanti per la valutazione di impatto ambientale e la loro descrizione in modo omogeneo e standardizzato in un catalogo di metadati. La descrizione dei dati ambientali si è basata su un profilo di metadati sviluppato specificamente a tale scopo sulla base dallo standard ISO 19115:2003(E). In questo contesto si riporta l'esperienza relativa all'applicazione della norma ISO 19115:2003(E) per lo sviluppo di un modello dei dati utilizzato per la descrizione di *dataset* ambientali geografici e non-geografici, mentre, mentre i risultati dell'indagine sono riportati in altra pubblicazione (Peccol, Muzzolini, 2004).

Il modello dei metadati

La descrizione delle risorse informative è stata pensata per soddisfare principalmente gli utenti dei dati, con lo scopo di fornire informazioni sufficienti per accertare l'esistenza di una risorsa informativa, per conoscere alcuni elementi utili per effettuare una scelta sull'opportunità o meno di acquisirla e individuare i soggetti collegati ad essa. Il modello dei metadati (Fig. 1) costituisce un profilo della norma ISO 19115:2003 (E) Geographic information — Metadata (ISO/TC 211, 2003) ed è stato sviluppato con l'intento di fornire un metodo coerente e standardizzato per documentare risorse informative digitali e cartacee identificabili come dati geografici e dati tabellari non-geografici: infatti, è possibile applicare la norma ISO 19115:2003(E) a dati non-geografici, anche se non per tutti gli elementi di metadati.

Un profilo di metadati si basa su una selezione di alcuni o tutti gli elementi di metadati presenti nello standard, che possono essere utilizzati con diversi livelli di dettaglio e complessità per la documentazione di una risorsa informativa. In questo contesto, prima di sviluppare il modello dei metadati, è stata accertata l'applicabilità di modelli di metadati esistenti, sviluppati nell'ambito di progetti di cataloghi di dati ambientali e/o territoriali (Pick, 1997)(Centro Interregionale, 2001)(Carotenuto, 2003)(Regione Piemonte, 2001)(Günther et al., 1996)(Sfida, 2003)(Swoboda et al. 1999)(Kazakos et. al. 2003). Tuttavia, i modelli considerati non sono risultati soddisfacenti o perchè si appoggiavano a standard di metadati proprietari od obsoleti o perchè troppo estesi e dettagliati per le esigenze del presente progetto.

Lo sviluppo del profilo ha tenuto conto delle seguenti esigenze:

- soddisfare la necessità di rilevare l'esistenza di un dato da parte di un potenziale utilizzatore e fornire gli elementi essenziali ai fini di una decisione in merito all'acquisizione;
- includere tutti gli elementi di metadato "Core" dell'ISO; tali elementi non sono tutti obbligatori, ma sono considerati il "nocciolo" dell'informazione vivamente consigliata per descrivere una risorsa informativa;
- includere una parte degli elementi opzionali nella norma in quanto ritenuti importanti per il contesto regionale o per soddisfare i requisiti informativi espressi dal Servizio per la V.I.A.;
- aggiungere nuovi elementi di metadati (estensioni) ove la norma risultava carente. Le estensioni hanno interessato elementi di metadati e aggiunte a liste di codici. I nuovi elementi di metadati hanno riguardato l'aggiunta della data di modifica dei metadati a complemento della data di cre-

azione dei metadati già presente nella norma, dei metadati descrittivi della normativa collegata con la risorsa informativa, di eventuali note per la copertura geografica e temporale;

- restringere la condizione di compilazione da facoltativa ad obbligatoria per alcuni elementi (es. proprietario del dato, fonte delle parole chiave ecc.);
- comprendere il profilo anche da parte di non esperti di metadati e di documentazione di informazioni territoriali; in alcuni casi i modelli di metadati possono essere talmente complessi da richiedere per la loro compilazione e consultazione personale specializzato;
- contenere il numero degli elementi dei metadati per semplificare il processo di compilazione ed eventuale aggiornamento dei metadati;
- produrre un profilo comprensibile anche da parte di non esperti in metadati e informazione territoriale;
- obbligatorietà nell'utilizzo del thesaurus GEMET quale fonte di parole chiave.

Il modello dei metadati è stato espresso utilizzando la Unified Modeling Language (UML) che definisce un linguaggio standard e le annotazioni grafiche per la modellizzazione, come proposto dalla norma. I diagrammi UML sono stati integrati da un Data Dictionary (DD) che descrive le caratteristiche dei metadati. Sono stati utilizzati elementi di metadato contenuti nelle entità di metadati della norma ISO illustrate in figura 2.

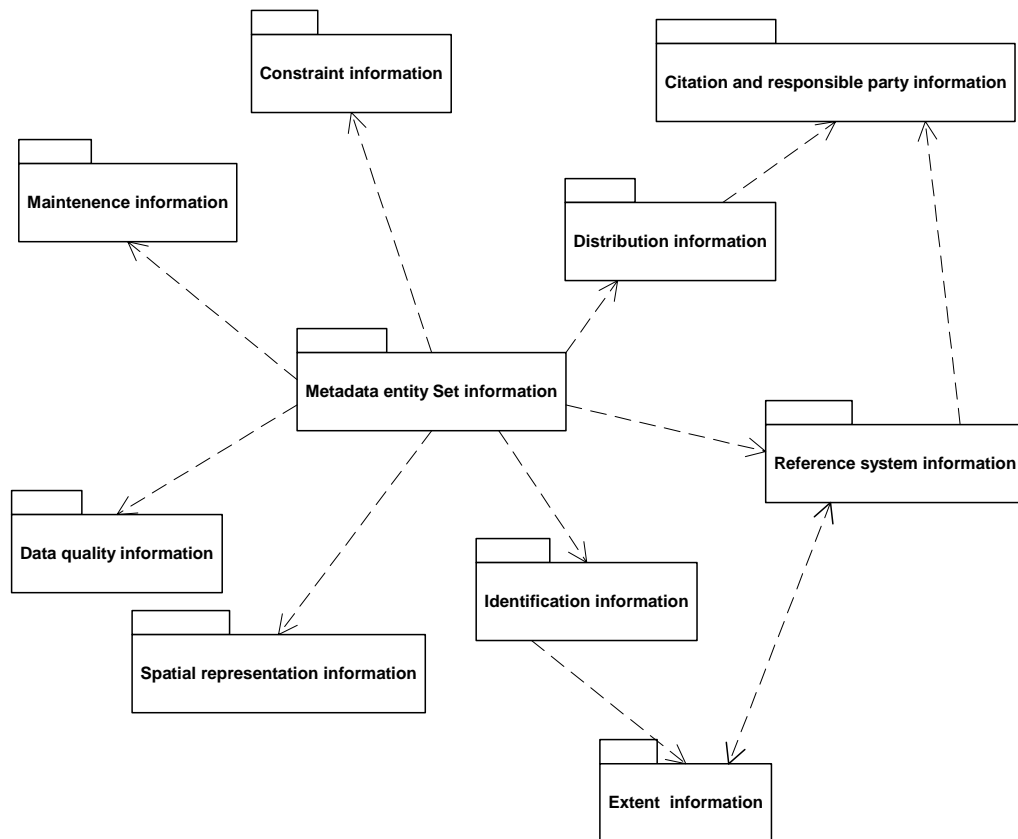


Figura 2 - I pacchetti di metadati 19115:2003(E) applicati nel Catalogo per la V.I.A

Considerazioni sull'applicazione della norma ISO 19115

La scelta dello standard ISO 19115:2003 (E) è derivata dalla necessità di documentare una prevalenza di risorse informative di tipo geografico e dal fatto che, nonostante la recente approvazione, si sta consolidando quale standard riconosciuto per i dati geografici a livello internazionale.

I vantaggi che si sono riscontrati con l'uso della norma ISO 19115 sono consistiti nella possibilità di poterla applicare a *dataset* geografici e non-geografici, nella presenza di descrittori per gli aspetti spaziali dei dati geografici che non si trovano in standard di metadati più generici e semplici come il

Dublin Core, dall'eshaustività della norma e dalla flessibilità consistente nella possibilità di creare profili *ad hoc* ed estensioni sia agli elementi di metadati, sia alle liste codificate.

Si segnalano anche alcuni aspetti problematici che suggeriscono l'opportunità di apportare alcune modifiche allo standard per rendere alcune definizioni più chiare e precise e per migliorare la sua applicabilità.

Un primo problema riguarda la definizione semantica del "*Dataset*". Nella sezione iniziale "Terms and Definitions" della norma, il "*Dataset*" viene definito come un "unità identificabile di dati", dalla quale si evince che non è preclusa la possibilità di specificare, se necessario, il *Dataset* nelle sotto-classi di "*dataset geografico*" e "*dataset non-geografico*". Tuttavia nell'Allegato G della norma, un *Dataset* viene definito come "...un' istanza di prodotto di dato spaziale che può essere generato o reso disponibile da un distributore di dati spaziali". Questa definizione esclude l'appartenenza di un "*dataset non-geografico*" al *Dataset*. Nello stesso allegato la parola "*Dataset*" indica anche un livello di astrazione in una gerarchia di metadati (es. Serie, *Dataset*, Classe di entità). Con la stessa parola si indica quindi una classe di risorse informative ed un livello gerarchico.

Un ulteriore problema incontrato è consistito nella mancanza di un elemento di metadato per specificare la tipologia o classe di risorsa informativa. In un modello di metadati può risultare utile specificare nella *root entity* che definisce i metadati applicati ad una risorsa informativa (nella norma ISO è MD_Metadata), sia la classe di informazione alla quale appartiene la risorsa (*dataset, non-geographic dataset, software, service, model*)(ISO/TC 211, 2003), sia il livello gerarchico (*hierarchy level*) di una risorsa informativa alla quale si applicano i metadati. Attualmente la norma ISO 19115:2003(E) prevede nella *root entity* MD_Metadata l'elemento N.6 "hierarchyLevel" che definisce il livello gerarchico della risorsa alla quale si applicano i metadati. I livelli gerarchici sono riportati nell'Allegato G nella norma ed identificabili nella Serie, *Dataset*, Classe di Entità, Entità, Classe di Attributi, Attributo e sono elencati nella lista codificata B.5.25 MD_ScopeCode <<CodeList>> alla quale l'elemento N.6 rimanda, insieme alle classi di informazione precedentemente citate. Al riguardo si ritiene che possa essere utile l'aggiunta di un'estensione allo schema proposto dalla norma ISO, consistente in un nuovo elemento di metadato nell'entità MD_Metadata per specificare espressamente, oltre al livello gerarchico, anche la classe di informazione della risorsa alla quale si applicano i metadati. Per esempio, i metadati possono essere applicati ad una risorsa informativa di tipo *Dataset non-geografico* (definito "*nonGeographicDataset*" nella norma ISO) al livello gerarchico di "*Dataset*". La lista di codici B.5.25 MD_ScopeCode <<CodeList>> della norma riporta sia classi di risorse informative geografiche e non-geografiche (es. "*nonGeographicDataset*" o "*software*"), sia livelli gerarchici (es. "*series*", "*dataset*", "*feature type*" ecc.). Poiché nella norma ISO lo stereotipo <<CodeList>> è usato per indicare una enumerazione aperta, è possibile, se necessario, aggiungere le tipologie di classi di risorse informative mancanti, purché secondo le indicazioni dell'Allegato F "*Metadata extension methodology*" della norma.

La norma offre, come già visto, la possibilità di applicare i metadati a livelli gerarchici diversi (es. *Series, Dataset, Feature Type*). Aggregati di *Dataset* che condividono caratteristiche simili di tema, data di rilevamento, risoluzione e metodologie di realizzazione possono essere documentati a livello di Serie e i metadati possono essere ereditati da ogni *Dataset* appartenente a quella Serie. Nel caso di cataloghi che documentano le risorse informative a livello *Discovery*, ovvero un livello per rilevare l'esistenza della risorsa, può essere sufficiente limitare l'applicazione dei metadati ai livelli gerarchici più alti (es. Serie o Aggregati di *Dataset*). Tuttavia la norma ISO 19115 richiede obbligatoriamente la creazione dei metadati al livello gerarchico di *Dataset*, lasciando come opzionali gli altri livelli gerarchici. Ciò entra tuttavia in contraddizione con l'esempio riportato nell'Allegato H "*Hierarchical levels of metadata*" ove si riporta che "i metadati possono essere creati esclusivamente a livello gerarchico della Serie" ed è necessario approfondire in futuro la documentazione a livello di *Dataset* solo per riflettere i cambiamenti apportati alla risorsa informativa nel tempo. Ne consegue che non è chiaro se è possibile limitare l'applicazione dei metadati al livello superiore della "Serie" o di altri aggregati di *Dataset*, o se la documentazione della risorsa informativa è sempre

necessaria al livello gerarchico di *Dataset*. La nostra conclusione è che sarebbe utile poter limitare, se necessario, la creazione dei metadati a livello della "Serie" mettendo questa quale condizione obbligatoria se la documentazione non viene effettuata a livello di "*Dataset*".

Infine tra gli elementi *Core* della norma ISO 19115 esiste il "Formato di Distribuzione" che è un elemento opzionale, nel senso che è possibile escluderlo o includerlo nel modello dei metadati. Per "Formato di Distribuzione" la norma intende la "descrizione del linguaggio informatico usato per rappresentare i dati in un record, file, messaggio, dispositivo per archiviazione e canale di trasmissione". Nel caso di risorse digitali è opportuno inserire questo elemento nel modello dei metadati anche perchè è uno degli elementi *Core*. Ne consegue l'imposizione della norma di documentare obbligatoriamente nome e versione del formato, azione che diventa inapplicabile sia nel caso di risorse informative distribuite su supporto cartaceo sia nel caso frequente in cui non sono noti o stabiliti formati e versioni di distribuzione della risorsa elettronica. Per tale motivo si riterrebbe opportuno rendere opzionale anche la documentazione degli elementi nome e versione del formato di distribuzione.

Bibliografia

European Commission, 2003. Adoption of the 5 Years Report on the Application and Effectiveness of the EIA Directive (Directive 85/337/EEC as amended by Directive 97/11/EC) How successful are the Member States in implementing the EIA Directive.

Carotenuto M., 2003. Estratto da Documento di progetto della versione 1.0 CARISMA: Catalogo delle Risorse Informative di interesse per SINAnet e Metadati Ambientali, APAT

Centro Interregionale, 2001. Architettura del Repertorio Cartografico. Convenzione Regione Basilicata, Centro Interregionale, Sistema Cartografico di Riferimento.

Environmental Thematic Coordination Group (ETCG), 2002. INSPIRE Environmental thematic user needs - Position Paper, Version 2. Editor Arvid Lillethun, European Environmental Agency..

Günther, O., Lessing, H., Swoboda, W. 1996. UDK: A European Environmental Data Catalogue, in: Proceedings of the Third International Conference/Workshop on Integrating GIS and Environmental Modelling, Santa Fe, New Mexico.

ISO/TC 211, 2003. ISO 19115:2003 (E): Geographic Information – Metadata.

Kazakos W., Akhounov A., Paoli H., Behrens S., Niesing H., Lombardo S., 2003. Editing ISO 19115 compliant Metadata in EUROSION, in Proceeding, "The Information Society and Enlargement of the European Union", 17th International Conference Informatics for Environmental Protection, Cottbus 2003.

Pick T., 1997. The Catalogue of Data Sources Core Data Model ETC/CDS of the EEA

Regione Autonoma Friuli Venezia Giulia - Servizio dell'Emergenza Ambientale, 2003. Il libro bianco sull'emergenza ambientale. Regione Autonoma FVG.

Peccol E. e Muzzolini V., 2004. Un rilevamento delle fonti di dati per la V.I.A.: problemi e proposte per un'infrastruttura di dati territoriali e ambientali a scala locale, in Atti VIII Conferenza Nazionale ASITA. Roma, 14 -17 dicembre 2004.

Piani L., Modolo C., Ianni E., 2004. La valutazione d'impatto ambientale in Friuli Venezia Giulia. Rapporto finale Progetto "Catalogo delle fonti di dati ambientali in Friuli Venezia Giulia a supporto della valutazione di impatto ambientale"

Regione Piemonte, 2001. PROGETTO INTERREGIONALE GAIA - Sistema informativo di governo dell'Ambiente e flussi informativi ambientali verso gli utenti. Rapporto di Conclusione Lavori. <http://www.regione.piemonte.it/ambiente/sina/gaia/index.htm>

Sfida, 2003. Progetto del catalogo dati. (URL www.sfida-life.it)

Swoboda W., Kruse F., Nikolai R., Kazakos W., Nyhuis D., Rousselle H., 1999. The UDK Approach: the 4th Generation of an Environmental Data Catalogue Introduced in Austria and Germany, Proceedings of the Third IEEE Meta-Data Conference, Bethesda, Maryland, USA.

