

Il Semantic Web per la Data Harmonization nei sistemi di supporto decisionale ambientale

Stefano DE LUCA (*)

(*) Enterprise Digital Architects, viale Egidio Galbani 55, 00156 Roma, tel. 06/7202.3774
fax. 06/7202.3153 stefano.deluca@enterpriseda.com

RIASSUNTO

I Sistemi di Supporto Decisionale Ambientali (EDSS) sono sempre più importanti nel controllo dei disastri o il controllo di sistemi ecologici. Per poter funzionare correttamente, questi sistemi devono supportare l'integrazione con numerose sorgenti dati, come l'Earth Observation, le reti di sensori, fonti web, e usare i più potenti sistemi di simulazione, tra cui gli automi cellulari e i sistemi intelligenti, in particolare i sistemi di agenti. L'integrazione di tutti questi aspetti può essere ottenuta estendendo il modello del semantic web.

ABSTRACT

The Environmental Decisional Support Systems are becoming critical to disaster control, in the ecological control etc. To fully work, these systems must support the integration with all interesting data sources, as the Earth Observation, the sensor networks and the web, and must support the most powerful simulation systems, including cellular automata, intelligent systems and Multi Agent Systems (MAS). The overall integration can be reached using the semantic web model.

KEYWORDS: DSS, EDSS, *semantic web, semantic grid, spatial web, multi agent systems, cellular automata, INSPIRE*

INTRODUZIONE

Negli anni recenti si è vista la crescita esponenziale dei sistemi di supporto alle decisioni (Decisional Support Systems, DSS), in particolare negli ambiti in cui la complessità delle decisioni è molto alta. Questi sistemi si basano prevalentemente su modelli statistici, business intelligence e la gestione di business rules, ed hanno raggiunto il massimo della diffusione nei sistemi di controllo manageriale delle aziende.

Consequenziale a questa tendenza si è visto l'insorgere dei DSS anche in ambito ambientale (Environmental DSS, EDSS; cfr. Rizzoli et al. 1997), ad es. per l'identificazione delle zone pericolose, la valutazione del rischio, il monitoraggio e la scelta di azioni per la riduzione di incendi o inquinamento, o il controllo di un impianto di trattamento delle acque. Il passaggio da DSS a EDSS non è soltanto il cambio di un dominio, ma introduce problematiche del tutto nuove: da un sistema ambientale ci aspettiamo che usi nelle analisi i dati che provengono dal territorio, dalle immagini satellitari (Earth Observation), dalle reti di sensori, da sistemi accessibili via web services (ad es. le condizioni climatiche); ci troviamo inoltre a confrontarci con sistemi che vanno a ragionare su dimensioni estremamente grandi, sia in termini di variabili da prendere in considerazione che anche come numerosità di osservazioni e dimensioni spaziali.

Questi temi sono ancora ad uno stato embrionale di soluzione. Possiamo cercare un aiuto dall'approccio del Semantic Web e, ancor di più, nel Semantic Grid.

IL SEMANTIC WEB E IL SEMANTIC GRID

Gli EDSS sono descrivibili come framework distribuiti, dinamici, complessi, spesso con caratteristiche randomiche e periodiche, ed eterogenei (Rizzoli, Young, 1997). Un sistema reale non può quindi prescindere da tutte queste caratteristiche ma, ad oggi, questo è vero? Non è affatto facile realizzare un EDSS che incorpori tutti questi elementi, poiché un sistema che tratti l'ambiente risulta dalla somma delle esperienze, dei modelli, delle definizioni e dei termini, del formato dei dati di una quantità considerevole di discipline, ognuna evolutasi indipendentemente.

Ancora Rizzoli identifica come elementi necessari di un EDSS:

- L'acquisizione della conoscenza e la sua rappresentazione.
- L'integrazione del modello e il suo riuso.
- La gestione dei dati spaziali.
- Il planning, la gestione e l'ottimizzazione.
- La definizione del problema e la sua soluzione.
- Il supporto degli esperti.

Quando si cerca di unificare i contributi delle discipline in un unicum si hanno due problemi, dovendo armonizzare:

- le informazioni in un formato omogeneo, che non si fermi dinanzi alle terminologie e alle diverse strutture dati;
- i modelli e i sistemi di ragionamento di ogni sistema coinvolto.

Il tutto sarà reso più complesso in quanto gli strumenti software attuali presentano nomenclature e specializzazioni che rendono sovente difficile l'armonizzazione anche a fronte degli stessi dati reali. D'altronde, questo problema ha molti punti in comune con quello incontrato in sistemi solo parzialmente formalizzati, come nel caso della pagine web e preso in considerazione dal Semantic Web, come presentato da Berners-Lee.

L'idea del Semantic Web è di avere in ogni pagina web (HTML) una descrizione in termini concettuali (semantici) del contenuto, usando delle ontologie che possono essere utilizzati da agenti intelligenti come *playground* per poter eseguire ricerche concettuali e non più limitate alla ricerca del testo (e quindi potendo discriminare le pagine che parlano di "terremoto" in senso proprio da quelle che ne parlano in senso metaforico, come i "terremoti politici"), e dar luogo ad agenti che agiscano in nostra vece richiamando ed offrendo servizi più complessi dell'abituale, ad es. riconoscendo in una pagina una conferenza interessante come AMFM ed automaticamente riservare lo spazio in agenda e magari prenotare anche i mezzi di locomozione e gli alberghi preferiti.

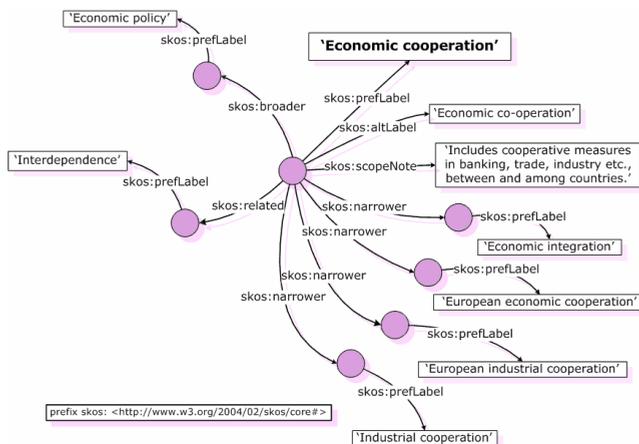


Figura 1: un esempio di ontologia in SKOS

Il punto di partenza di questo approccio è dunque la categorizzazione dei termini e delle loro relazioni, ad es. di ampliamento di un concetto, o di generica relazione. In figura 1 vediamo un esempio che usa le terminologie di SKOS, Simple Knowledge Organisation System, del consorzio W3C, *The World Wide Consortium*. Vediamo che è possibile definire i termini, le relazioni tra loro (un termine è più preciso, *narrower*, o più ampio, *broader*, o correlato, *related*), i termini alternativi e in altre lingue.

La rappresentazione delle ontologie è basata sull'ormai de facto standard per l'interoperabilità, XML. A partire da questo si è sviluppato in prima istanza l'RDF (*Resource Description Framework*), che consente di rappresentare termini e relazioni in modo disambiguo tramite triple di un concetto, una relazione e un altro concetto. In RDF è possibile descrivere: classi, sotto-classi, proprietà, sotto-proprietà, domini, range per le proprietà etc.

RDF è stato grandemente esteso da OWL (*Web Ontology Language*), un altro standard del W3C. Incrementa RDF (da cui parte: i dati OWL sono anche dati RDF, per quanto con una semantica diversamente definita) aggiungendo anche la cardinalità per gli attributi, la proprietà inversa, la disgiunzione tra classi, la sinonimia, proprietà di secon'ordine etc. OWL esiste in tre varianti, progressivamente più potenti e complesse: OWL Lite, OWL DL ed OWL Full. Il più grande vantaggio di OWL rispetto a RDF è la possibilità di inserire abbastanza conoscenza già nella rappresentazione dei dati per poter attivare dei *reasoners*, ovvero dei programmi che fanno uso delle regole inserire per sviluppare inferenze, possibilità d'uso, possibili incongruenze.

In effetti, spesso si confonde il Semantic Web con la rappresentazione dei dati utilizzata, mentre nella visione originaria questi dati hanno lo scopo di attivare proprio i sistemi di ragionamento, di ricerca e diffusione di servizi realizzati tramite agenti intelligenti, ovvero dei software autonomi capaci di rispondere a degli obiettivi fissati in modo autonomo e pro-attivo.

Il Semantic Web si è di recente andato ad integrare con un altro potente concetto, quello del Grid Computing. Con questo ultimo termine intendiamo la possibilità di far cooperare una grande quantità di risorse (computer, sensori etc.) su un obiettivo unico; ad es. molte migliaia di computer distribuiti su tutta la Terra hanno contribuito alla realizzazione di un modello di diffusione della SARS, contribuendo a tenere sotto controllo questa terribile malattia.

La somma della potenza di calcolo del Grid computing e la capacità di ragionamento e condivisione dei dati del Semantic Web dà luogo al **Semantic Grid**. Grazie all'origine storica di questo tipo di architettura (usata inizialmente per simulazioni meteorologiche, di fisica, in ambito medico), vediamo che l'orientamento è proprio quello della e-Science, con la visione (citando Roure et al.): “the Semantic Grid vision is to achieve a high degree of easy-to-use and seamless automation to facilitate flexible collaborations and computations on a global scale, by means of machine-processable knowledge both on and in the Grid.”.

Ontologie ed e-Science

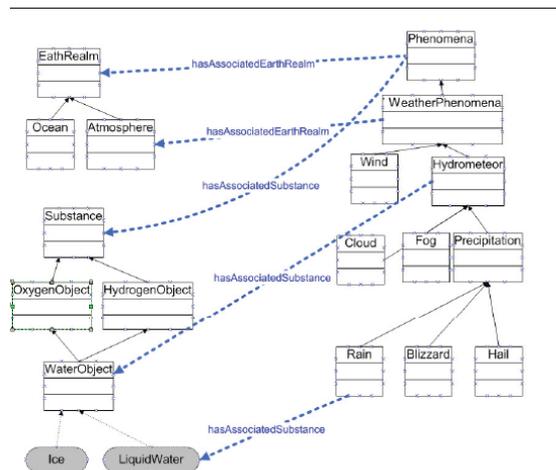
Quanto questo processo è stato recepito dalle Scienze della Terra? L'iniziativa Europea INSPIRE – Infrastruce for Spatial Information in Europe – riconosce l'importanza di avere dei dati indipendenti dall'ontologia, e si pone come obiettivo il superamento di questi blocchi concettuali; la produzione reale non è stata però ad oggi equivalente all'ampiezza del tema: nel documento ESDI – European Spatial Data Infrastructure – si parla di “[the] standards [that] will address concepts of geographical information and information technology”. Il fatto che nei termini non si usi né *semantic* né *ontology* mostra quanto non vi sia una forte coscienza delle tecnologie e degli approcci che possono portare al raggiungimento del goal.

ISO ha uno specifico comitato, ISO TC/211, per la standardizzazione del GIS: “The focus of geo-spatial standards is to define the basic semantics and structure of geographic information for data management and data interchange purposes, and to define geographic information service components and their behaviour for data processing purposes” (ISO 19101). Nello standard ESDI si cita questo passo per aggiungere che ESDI dovrà essere compatibile con TC/211.

L'Open GIS Consortium (OGC) affronta in modo serio il problema in un whitepaper dal titolo “Spatial Web”. L'approccio è corretto (tra l'altro c'è anche una comparazione con gli altri attori), ma il semantic web viene visto solo al primo livello, come fonte di armonizzazione dei dati, e quindi adeguata per i web services, ma non si va oltre, ovvero all'aspetto operativo.

Decisamente più avanti sono gli statunitensi della NASA con il loro progetto SWEET, *Semantic Web for Earth and Environmental Terminology*. L'approccio qui è molto più pragmatico (cfr. Raskin 2004): partendo dalle keyword del sistema GCMD, sono state sviluppate alcune ontologie che vanno a coprire gli aspetti salienti dell'Earth Science:

- Earth Realm
- Non-Living Element Living Element
- Physical Property
- Units
- Numerical Entity
- Temporal Entity
- Spatial Entity
- Phenomena
- Human Activities



- Data

Queste diverse ontologie sono poi correlate tramite macro-relazioni (come quelle mostrate in figura 2). Le ontologie sono pubbliche e disponibili su web. La NASA riconosce esplicitamente il valore dei reasoners applicati a queste ontologie, al punto di averle sviluppate nella versione più completa di OWL, OWL Full.

I SISTEMI DI SUPPORTO DECISIONALE AMBIENTALE

Avendo delineato brevemente le caratteristiche del Semantic Web e Grid a noi più pertinenti, nonché lo stato degli standard, possiamo finalmente chiederci a cosa queste metodologie e tecnologie sono utili per un EDSS e se lo stato dell'arte è adeguato.

Gli EDSS e le ontologie

Un EDSS può gestire appropriatamente ambienti limitati, come ad es. un sistema di smaltimento delle acque o di gestione dei rifiuti, o tentare di coprire eventi su aree più vaste, come per il controllo ecologico di una regione o il forecasting degli incendi in una determinata area. In entrambi i casi, per operare efficacemente il DSS dovrà ricevere input aggiornati della situazione reale, per produrre previsioni dei comportamenti futuri, per acquisire dati storici che possono migliorare i modelli previsionali, per poter operare delle simulazioni di tipo *what-if* cambiando solamente alcuni dei fattori.

È necessario dunque far convergere dati da tutte le sorgenti disponibili: reti di sensori, sensori specifici per l'applicazione, Earth Observation.

La convergenza di EDSS ed ontologie si è rivelata critica in un framework completo per EDSS realizzato da chi scrive (De Luca et al. 2006), che implementa l'intero processo (in fig. 3), dalla Earth Observation ad un sistema ad agenti (MAS) per simulazioni.

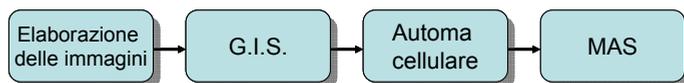


Figura 3: schema a blocchi del framework dell'EDSS

In particolare:

- prende i propri input dalle immagini satellitari;
- automaticamente queste vengono processate per classificare i temi di interesse (ad es. lo stato di stress della vegetazione, quanto questa rischi di incendiarsi; l'effettiva presenza di costruzioni);
- i dati vengono armonizzati in una struttura che raccoglie i dati raster, i dati classificati, gli eventuali dati dai sensori, i dati vettoriali dal GIS; questa struttura è un grid basato su esagoni, punto di partenza per un automa cellulare;
- l'automata cellulare usa delle funzioni di transizione che sono l'espressione matematica del modello evolutivo dell'ambiente; nel nostro lavoro suggeriamo l'uso di layer cellulari automatati, che consentono l'uso parallelo di diverse funzioni di transizioni, una per ogni aspetto saliente (ad es. il controllo dell'umidità, la velocità di diffusione dell'incendio etc.);

- sull'automata cellulare si attesta un sistema di agenti (multi agent system, MAS) che va a modellare il comportamento umano; il MAS prende gli input dall'automata e lo cambia con i suoi output, ad es. spargendo dell'acqua su un incendio.

Il modello è sufficientemente generalizzato da poter essere usato nelle più diverse applicazioni; il lavoro iniziale è stato testato in un sistema previsionale di incendi. Il collante di tutte le applicazioni è il GIS che mantiene i layer vettoriali, quelli classificati e le immagini raster.

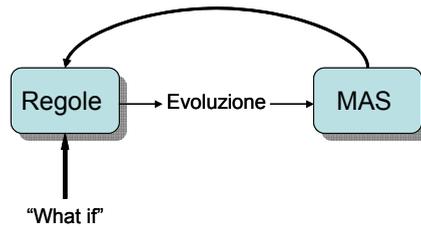


Figura 4: uso del sistema Multi Agenti (MAS) per il *what-if*

Il GIS e l'EDSS

Il GIS è quindi essenziale per ricomporre tutti i moduli. Le analisi eseguite sulle immagini raster satellitari sono trattate con prodotti ed algoritmi specifici che consentono di analizzare al loro interno i risultati. Dovendoli usare in un contesto più ampio, si devono poter trattare in un formato più standard. Avendo già a disposizione molti dati vettoriali, non rimane che georeferenziare i dati (cosa che avviene in automatico, avendo le coordinate nelle immagini satellitari) e trasferirli all'interno di un GIS.

Il GIS diviene quindi il repository attivo di tutti i dati, in particolare per i moduli più specialistici ed "intelligenti", ovvero il sistema di evoluzione del sistema (nel nostro caso, un automata cellulare) e il simulatore dei comportamenti umani, tramite un sistema di agenti (cfr. fig. 5).

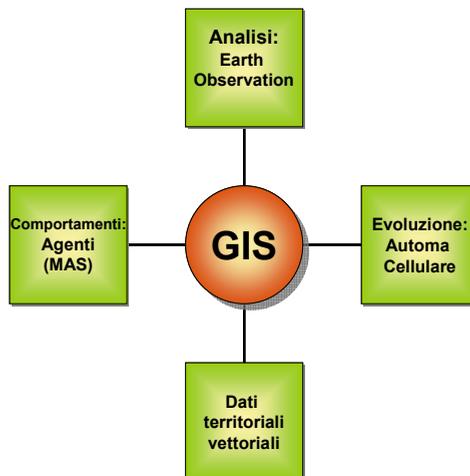


Figura 5: il GIS centrale all'EDSS

È infatti necessario che un modulo “terzo” rispetto a tutti gli altri mantenga le informazioni etichettate opportunamente (tramite i layer); nella simulazione è infatti necessario sapere dove si trovano gli attori, se i movimenti che fanno vanno in conflitto, se un determinato luogo è, ad es., in fiamme o no. Queste informazioni altamente dinamiche non possono essere gestite dai singoli attori, a ragione delle interazioni richieste: bisognerebbe programmare in ogni attore tutte le interazioni con tutti gli altri attori, il che renderebbe il sistema centrale e difficilmente estendibile. Mantenere queste informazioni in un luogo esterno consente:

- **il mantenimento della coerenza**, ovvero la rimozione di incongruenze (ad es. un camion che passa in un bosco infuocato senza riportarne danni);
- **la centralizzazione delle informazioni** per tutti i moduli, anche attualmente non previsti, senza che però vi sia un coupling tra i moduli aggiunti; ogni modulo comunicherà soltanto con il GIS, e con gli altri moduli solo se strettamente necessario;
- **la gestione delle performance** e della storicità delle informazioni; il GIS è infatti ottimizzato per questi compiti e può mantenere anche la storia dei cambiamenti.

Cosa manca: gli aspetti operativi delle ontologie e il multilinguismo

La definizione delle ontologie fatta nel progetto SWEET è particolarmente utile, perché parte da una base consolidata e perché è stata estesa per accettare dei reasoners, ma è ancora limitata nei suoi obiettivi, e quindi anche nelle sue realizzazioni.

Il linguaggio può essere visto sia nei suoi aspetti descrittivi, e quindi i termini, che in quelli operazionali, e quindi le azioni. Le ontologie definite in SWEET sono relative solamente al primo di questi due aspetti, e quindi limitati alla descrizione di una *situazione*, ma non alle *azioni*, anche in termini di previsioni o modifica, che su quella situazione possono essere apportate. La teoria performativa del linguaggio è stata introdotta dai filosofi Austin e da Searle e forma la base delle comunicazioni tra agenti; l'ipotesi è che le parole risultano in azioni, che possono informare, richiedere lo stato, suggerire od ordinare azioni, creare un nuovo stato delle cose.

Questo è proprio quanto difetta alle ontologie che abbiamo finora realizzato: non vi sono concetti per poter agire, non è preso in considerazione come una trattazione della semantica del mondo ambientale non possa prescindere dalle azioni umane o dagli effetti di un'azione sull'ambiente. Le ontologie vanno quindi arricchite in questo senso, ed è opportuno un lavoro di definizione delle *azioni* sui diversi domini ambientali.

Per chiarire meglio questi punti, proviamo a fare degli esempi, usando i termini e le ontologie di SWEET. Nel nostro sistema di modellazione e prevenzione degli incendi, abbiamo bisogno di sapere dove ci sono i depositi di legname perché possono essere un punto di rilancio forte di un incendio. I depositi possono essere definiti come subclass del concetto *ForestProduct* (nell'ontologia SWEET *human_activities*). In termini operativi, quello che va aggiunto è che questi particolari prodotti della foresta sono massimamente incendiabili (è un'azione) e quindi potenzialmente pericolosi (e quindi correlati al concetto di *Hazard*, già esistente) applicando però una regola che eviti la troppa genericità della relazione (la regola dovrebbe stabilire che sono pericolosi in presenza di un incendio, o di una forte temperatura e bassa umidità etc.).

Ancora più interessante il termine *Lake* (nell'ontologia *Earthrealm*), subclass di *SurfaceWater*. *Lake* dovrebbe essere correlato alla possibilità (azione) di spengimento di un incendio

(raggiungibile da un aereo Canadair, ovviamente solo se le sue dimensioni sono adeguate), e correlato anche all'impossibilità di essere usato per la fuga da persone a piedi o in auto (per le parti della simulazione relative all'evacuazione).

Come si vede, i termini ci sono, grazie al grande lavoro della NASA, ma vanno resi "vivi" dall'uso operativo e non solo di fotografia statica. Questo è quanto abbiamo iniziato con l'EDSS descritto, con le ontologie usate dagli agenti (l'ambiente usato, JADE, consente l'uso diretto e comodo delle ontologie per la comunicazione tra gli agenti, comunque previsto dalle specifiche di comunicazione ACL, Agent Communication Language, della FIPA, l'organizzazione IEEE per la standardizzazione della tecnologia ad agenti. Non è necessario che le ontologie vengano usate dai sistemi ad agenti, possono esserlo anche da altri tipi di tecnologie, ad es. sistemi di simulazione basati su catene di Markov, ma sicuramente la pienezza viene raggiunta con i sistemi multi agente, e anche qui rientriamo nello spirito pieno del semantic web, che intende usare la semantica come spazio concettuale per gli agenti che devono fare qualcosa per noi.

L'ultimo aspetto su cui non si è lavorato abbastanza è sul multilinguismo. In un sistema Grid transnazionale (almeno europeo) abbiamo molte lingue, e le ontologie devono avere delle etichette tali da consentire una traduzione veloce da una lingua all'altra e da queste ai concetti. Un lavoro in questa direzione è svolto attualmente in ambito W3C con il citato SKOS, e in ambito CEN dal progetto ADNOM, "European Network for Administrative Nomenclature". A nostro avviso un comitato equivalente all'ADNOM dovrebbe essere creato per quanto riguarda le ontologie geografiche ed ambientali.

Avere un'ontologia ambientale arricchita con le azioni e multilinguistica sarebbe un carburante eccezionale per gli EDSS europei: è una sfida che non possiamo che raccogliere!

BIBLIOGRAFIA

- Berners-Lee, T., Hendler, J. and Lassila, O., 2001, "The Semantic Web", *Scientific American*, 284, pages 34-43 May 2001.
- De Roure, D. Jennings, N.R. Shadbolt, N.R., 2005, "The Semantic Grid: Past, Present, and Future", *Proceedings of the IEEE*, Volume 93, Issue 3, March 2005, Pages 669-681
- De Luca, Balsi, Maggio, 2006, "An Integrated Environmental Decisional Support System Framework using Earth Observation, Cellular Automata and Multi-Agent System", International Summit on Environmental Modelling & Software, IEMSS 2006
- European Spatial Data Infrastructure (ESDI), 2001, Action Plan December, 2001, http://www.ec-gis.org/inspire/reports/ESDI_Action_plan1aa13.pdf
- FIPA, The Foundation for Intelligent Physical Agents, <http://www.fipa.org>
- OGC, 2004, "The Spatial Web. An Open GIS Consortium (OGC) White Paper"
- Rizzoli, A.E., Young W.J., 1997, "Delivering environmental decision support systems: software tools and techniques". *Journal of Environmental Modelling & Software*, 12(2/3) (1997), 237-249.
- Raskin, R., 2004, "Enabling Semantic Interoperability for Earth Science Data." NASA Earth Science Technology Office (ESTO) 2004.
- SKOS, W3C, Simple Knowledge Organisation System, <http://www.w3.org/2004/02/skos/>
- SWEET, Semantic Web for Earth and Environmental Terminology, NASA, Internet: <http://sweet.jpl.nasa.gov/>