

## II

LINGUAGGI, STANDARD E METADATI  
QUESTIONI DI METODO E TERMINOLOGICHE

LANGUAGES, STANDARDS AND METADATA  
METHODOLOGICAL AND TERMINOLOGICAL PROBLEMS



## LINGUAGGI DICHIARATIVI PER LA RICERCA ARCHEOLOGICA

### 1. INTRODUZIONE

Un celebre motto di Marshall McLuhan, uno dei più illustri teorici della comunicazione di questo secolo, recita provocatoriamente: il medium è il messaggio<sup>1</sup>. McLuhan intendeva che ogni mezzo di comunicazione influisce sul pensiero e sulla cultura di una società non solo attraverso i messaggi che esso diffonde, ma anche in funzione delle sue caratteristiche tecniche. Il medium usato nella comunicazione definisce la scala e i modelli entro cui i pensieri umani si esprimono; questa “non neutralità” del mezzo comunicativo è nota fin dall’antichità: già Platone, con l’affermarsi della scrittura, temeva cambiamenti culturali<sup>2</sup>.

Così le macchine informatiche agiscono con inarrestabile rapidità sui paradigmi culturali e trasformano il nostro modo di comunicare e di pensare. Le applicazioni informatiche, pensate inizialmente quali normali strumenti di lavoro, stanno influenzando anche le metodologie di ricerca. Lo studio del territorio, ad esempio, non può più prescindere da una precisa prospettiva geografica; i progetti archeologici devono ormai tutti guardare all’integrazione della conoscenza acquisita in contesti territoriali più ampi, magari studiati in collaborazione da diverse Istituzioni, interessate a promuovere la conoscenza e la tutela del patrimonio di un contesto paesaggistico e storico sulla base della sua precisa identificazione spaziale.

Le applicazioni informatiche impongono che i dati siano conformati ai modelli logici ideati per la loro gestione. I database, ad esempio, richiedono l’inserimento dei dati in una struttura logica e in campi ben definiti. In alcuni casi è necessaria la destrutturazione del formato originale dei dati e questa procedura può causare la perdita di informazioni e il rischio di erronee interpretazioni, in quanto essa trasforma in forma sintetica e granulare ciò che era nato per essere concepito unitariamente.

<sup>1</sup> *The Medium is the Message* (McLUHAN 1967): l’autore fu così entusiasta e divertito dell’errore del grafico sul prototipo della prima copertina – message/massage – da decidere di tenere il titolo sbagliato, che si prestava anche a giochi di parole con message/mess-age.

<sup>2</sup> Platone, per esempio, nel Fedro, a proposito dell’alfabeto, faceva dire al re Thamus: «esso ingenererà oblio nelle anime di chi lo imparerà: essi cesseranno di esercitarsi la memoria perché fidandosi dello scritto richiameranno le cose alla mente non più dall’interno di se stessi, ma dal di fuori, attraverso segni estranei: ciò che tu hai trovato non è una ricetta per la memoria ma per richiamare alla mente. Né tu offri vera sapienza ai tuoi scolari, ma ne dai solo l’apparenza perché essi, grazie a te, potendo avere notizie di molte cose senza insegnamento, si crederanno d’essere dottissimi, mentre per la maggior parte non sapranno nulla; con loro sarà una sofferenza discorrere, imbottiti di opinioni invece che sapienti» (PLATONE, *Fedro*, 274c-276a, in *Opere*, Vol. I, Bari 1967, Laterza, 790-792).

Larga parte dell'attenzione è dunque oggi puntata a quei linguaggi che permettono il trattamento informatico dei documenti senza alterarne la struttura originaria. Per quanto riguarda i documenti testuali, è importante usare linguaggi che permettano di interpretare il valore connotativo di parti del testo a fronte di disomogeneità del valore denotativo dei segni adottati. Nel progetto di sistemi informativi archeologici, inoltre, che mirano alla realizzazione di forme di conoscenza persistenti, stante l'unicità e la non riproducibilità dell'azione di scavo, l'adozione di linguaggi non proprietari – cioè indipendenti da hardware o software specifici – è un requisito strategicamente imprescindibile nella prospettiva dell'usabilità dei dati e della persistenza degli archivi nel lungo periodo.

Quale ulteriore requisito i dati che vengono inseriti nel circuito della conoscenza devono poter contenere specifiche descrizioni del loro significato, anche note come metadati. I metadati – e il modo di organizzarli e congiungerli – sono oggi molto discussi nella prospettiva del Semantic Web, il web del prossimo futuro, inteso a permettere migliori indicizzazioni delle risorse presenti sulla rete. Ciò sarà reso possibile proprio grazie ai metadati e alle ontologie, che potranno essere prodotte con la sintassi di linguaggi specifici, trasformando i metadati in sistemi di conoscenza interpretabile dalle macchine.

All'insieme di tutti questi requisiti rispondono oggi solo i linguaggi informatici definiti “dichiarativi”, cioè quei linguaggi in cui i marcatori descrivono in modo chiaro il significato semantico del testo marcato. Nati con SGML (Standard Generalized Markup Language, cfr. *infra* § 4.1) oggi derivano sempre più numerosi dal suo potente erede: XML (Extensible Markup Language, cfr. *infra* § 4.2).

Se la confusione dei linguaggi è stata d'ostacolo, nel corso della storia umana, alla diffusione della conoscenza, l'informatica è intrinsecamente vocata a riprodurre in forme ancor più straordinarie l'antica babilonia<sup>3</sup>, poiché i suoi linguaggi mutano con grande facilità e rapidità. Nei gruppi di discussione si sente dire spesso ironicamente: «la cosa bella degli standard è che ce ne sono tantissimi tra cui scegliere». Sebbene la tecnologia sia uno dei pilastri dell'era della globalizzazione, essa deve essere utilizzata con ragionevole attenzione, almeno per quanto riguarda i linguaggi. Un progetto scientifico che miri a gestire efficacemente i suoi dati e a dividerli non può prescindere da un'attenta valutazione degli strumenti adottati per strutturare l'informazione digitale.

Questo lavoro si prefigge pertanto di far luce su alcuni linguaggi dichiarativi, tra quelli accettati e supportati dall'autorità del W3C (World Wide Web Consortium), che paiono adatti al trattamento dell'informazione archeologica o che già sono stati utilizzati in questo ambito. Essi promettono di

<sup>3</sup> Per la Bibbia fu il risultato di un progetto divino: «Ecco, essi sono un solo popolo e hanno tutti una lingua sola; questo è l'inizio della loro opera e ora quanto avranno in progetto di fare non sarà loro impossibile. Scendiamo dunque e confondiamo la loro lingua, perché non comprendano più l'uno la lingua dell'altro» (*Genesi* 11, 6-7).

garantire la persistenza dei dati anche sul lungo periodo e appaiono ragionevolmente destinati a diventare gli standard futuri della comunicazione nelle discipline umanistiche.

## 2. IL LINGUAGGIO UMANO E IL COMPUTER: DALL'ASCII ALL'UNICODE

Tutto quello che accade in un computer è identificato da stringhe di numeri binari. Come è noto, per il computer l'unità atomica di informazione è il bit, che può assumere il valore 0 o 1. I bit sono tuttavia raggruppati in byte, costituiti da 8 bit, oppure da loro multipli. Un computer identifica i caratteri di un testo associando ad essi un numero intero; il numero assegnato a ciascun segno è univoco all'interno di un insieme, definito *character set*. Poiché esistono tante lingue diverse, ciascuna con le proprie esigenze segniche, esistono tanti set di caratteri. Il più diffuso (del 1960) è l'ASCII (American Standard Code for Information Interchange) costituito da 128 caratteri. Date le limitazioni dell'ASCII base, si è reso presto necessario creare dei set di caratteri estesi. Tra questi (detti anche SBCS, Single Byte Character Sets) ricordiamo ISO-8859-1, che fornisce la maggior parte dei caratteri necessari alle lingue dell'Europa occidentale.

Dai SBCS si è giunti oggi ai set di caratteri definiti MBCS (Multiple Byte Character Sets): in fase di progressiva affermazione è il set Unicode (Universal Character Set) che include tutti i set attualmente definiti. L'emergere dello standard Unicode, unito alla recente disponibilità di strumenti che lo supportano, è fra i più significativi sviluppi della tecnologia della globalizzazione del software. La codifica Unicode a 32 bit è tuttavia molto costosa in termini di memoria, per cui si sono diffusi suoi sottoinsiemi, definiti UTF-16 e UTF-8, quest'ultimo perfettamente compatibile con lo standard ASCII esteso.

## 3. LINGUAGGI DI MARCATURA PROCEDURALI E DICHIARATIVI PER LA RAPPRESENTAZIONE ELETTRONICA DEI DOCUMENTI

### 3.1 *Linguaggi procedurali*

La redazione e la composizione grafica di un testo, con un word processor, sono operazioni sincroniche. Nei moderni sistemi di scrittura elettronica quello che appare sullo schermo è ciò che si otterrà in fase di stampa. I programmi che permettono di avere questo tipo di visualizzazione si definiscono WYSWYG (What You See is What You Get).

L'utilizzo di un word processor introduce nel testo istruzioni operative, che definiscono la struttura del testo e dei caratteri che lo compongono. Questa "marcatore" è definita procedurale. Il documento diventa leggibile e rappresentabile solo da un programma in grado di interpretare i codici di formattazione. Il formato della codifica, se appartiene ad una società che può

cambiarne la sintassi a suo piacimento, si definisce anche formato proprietario (ad esempio quello di Microsoft Word o PageMaker).

Dal punto di vista della persistenza dei dati sul lungo periodo, l'affidarsi a linguaggi proprietari è molto pericoloso e sicuramente da evitare. Larghe collezioni documentali possono divenire perfettamente illeggibili nel volgere di pochi anni. Contro i formati proprietari e il software "opaco" esiste oggi un crescente movimento di pensiero, alimentato da strenui oppositori di Microsoft, che fa capo all'iniziativa Open Source<sup>4</sup> per il software e promuove l'uso di standard documentali non proprietari.

### 3.2 Linguaggi dichiarativi

Un linguaggio dichiarativo è un linguaggio in cui i marcatori non servono solo a descrivere specifiche operazioni da compiere per la formattazione e la stampa (come in un linguaggio procedurale) ma anche a esporre il significato semantico del contenuto. I linguaggi dichiarativi originano da un modello di William Tunncliffe del 1967, nel quale si proponeva di separare il contenuto dei documenti elettronici dalle istruzioni per la loro rappresentazione. A tal fine sul finire degli anni '60 fu istituito un comitato di studio, definito GenCode Committee.

Il primo linguaggio dichiarativo fu presentato da lì a poco da Charles Goldfarb, R. Mosher e R. Lorie. Il linguaggio si chiamava GML (Generalized Markup Language, acronimo non incidentalmente formato dalle iniziali dei cognomi dei tre creatori) e fu inizialmente adottato dall'IBM. Nel 1978 l'American National Standard Institute (ANSI) incaricò Goldfarb di sviluppare ulteriormente il linguaggio, con l'obiettivo di produrre uno standard internazionale: i lavori procedettero speditamente e nel 1980 fu emessa la prima versione di SGML (Standard Generalized Markup Language). Il linguaggio in pochi anni raggiunse la forma definitiva, anche grazie al supporto dell'ISO (International Organization for Standardization), che lo registrò con la sigla ISO 8879:1986.

## 4. I METALINGUAGGI SGML E XML

### 4.1 Standard Generalized Markup Language (SGML)

SGML<sup>5</sup> è basato esclusivamente su caratteri ASCII standard, per cui assicura la persistenza dei documenti nel lungo periodo; un documento SGML è perfettamente leggibile con qualsiasi editor di testo elettronico;

<sup>4</sup> <http://www.opensource.org>.

<sup>5</sup> Cfr. a titolo introduttivo C.M. SPERBERG-McQUEEN, L. BURNARD, *A gentle introduction to SGML*, in *Guidelines for Electronic Text Encoding and Interchange* (TEI P3), 1994, cap. 2 (<http://www.isgmlug.org/sgmlhelp/g-index.htm>). Per una esaustiva bibliografia sul linguaggio cfr. <http://xml.coverpages.org/sgml.html>.

permette di costruire schemi gerarchici interni al documento e definire collegamenti tra diversi documenti; pur essendo rigido e definito nella sintassi è aperto ed estensibile dal punto di vista semantico. Ogni documento SGML prevede un documento accessorio definito DTD (Document Type Definition), all'interno del quale è necessario definire la struttura e la gerarchia degli elementi.

SGML non si occupa della rappresentazione del documento: è un problema troppo concreto per il livello di astrazione di SGML. Poiché tuttavia la proprietà più importante che viene riconosciuta a un documento è il suo valore tangibile e visivo non può sorprendere che SGML abbia avuto in generale un ruolo marginale per il mondo del web, che si regge proprio sulla diffusione e visualizzazione dei documenti. SGML è stato utilizzato per scrivere altri linguaggi, definibili “applicazioni” o “linguaggi applicativi” (che per ereditarietà sono anch'essi dichiarativi e non proprietari), indirizzati a risolvere problemi ben specifici. Tutte queste caratteristiche permettono di definire SGML come un “metalinguaggio”, cioè un linguaggio che permette la creazione di altri linguaggi di tipo descrittivo: la sua applicazione più nota è sicuramente HTML (cfr. *infra* § 5.1).

Da SGML deriva anche XML, che lo ha sostanzialmente sostituito e promette di diventare lo standard futuro della comunicazione globale.

#### 4.2 *Extensible Markup Language (XML)*

Fino a quando il web ha veicolato contenuti statici, indirizzati alla “pubblicazione” di contenuti, HTML si è rivelato abbastanza adeguato, anche grazie alle continue implementazioni di nuovi elementi. Tuttavia, con la nascita del commercio elettronico e con la necessità di sviluppare forme di interazione client server più evolute, HTML si è dimostrato troppo limitato.

Molte tecnologie sono state sviluppate per potenziare HTML. È il caso delle Applet Java, degli Activex, dei filmati Flash, dei Javascript e dei VBScript, che hanno introdotto le potenzialità dei linguaggi di programmazione nel mondo del web, senza risolvere tuttavia la problematica più profonda di HTML: la sua inestensibilità semantica.

La necessità di trovare un piano di comunicazione globale, capace di rendere veramente condivisibili i dati tra diverse applicazioni e su diverse piattaforme hardware ha richiesto, sulla scorta dell'esperienza fatta nei primi anni di vita del www e dell'information technology, la definizione di un linguaggio universale, non proprietario, dichiarativo, estensibile e a differenza del vecchio SGML pensato esplicitamente per essere supportato dal web.

È nato così XML<sup>6</sup>, sviluppato da XML Working Group (originaria-

<sup>6</sup> <http://www.w3.org/XML/>. Per una esaustiva bibliografia cfr. <http://xml.coverpages.org/biblio.html>.

mente noto come SGML Editorial Review Board) sotto gli auspici del W3C nel 1996.

Il nuovo metalinguaggio è stato realizzato mantenendo la compatibilità con SGML: un documento XML valido è un documento SGML valido. XML non sostituisce HTML, perché i due linguaggi sono complementari. Il compito di XML è descrivere i dati sul web, non solo visualizzarli. I documenti basati su XML non sottintendono una loro rappresentazione obbligatoria, anzi uno stesso documento può essere facilmente utilizzato per scopi diversi; i dati possono essere utilizzati da qualsiasi applicazione che comprenda il linguaggio XML: un browser, un database, un telefono cellulare, un word processor.

XML permette di creare tag personalizzati (nuovi elementi). Questa caratteristica, oltre a configurare XML come un metalinguaggio, permette di semplificare la creazione di applicazioni che svolgono operazioni intelligenti con i documenti elettronici, quali ad esempio i sistemi di information retrieval basati sulla semantica.

L'interesse suscitato da XML è stato subito enorme, anche per i validi supporti che hanno permesso di renderlo "praticamente" utilizzabile. In qualità di metalinguaggio XML ha generato applicazioni per obiettivi specifici, quali XSL (Extensible Stylesheet Language), nato per permettere la rappresentazione di documenti XML sul web. La tecnologia XML-XSL è stata definitivamente integrata nei browser più recenti, quindi i file XML possono essere utilizzati per la pubblicazione diretta sul web.

## 5. LINGUAGGI DI MARCATURA PER IL TRATTAMENTO DI TESTI

Come si è detto SGML e XML sono metalinguaggi. Ciò ha consentito di sviluppare numerose loro "applicazioni" che definiscono linguaggi per ambiti specifici. Per quanto riguarda SGML, segnaliamo HTML e lo stesso XML. Per quanto riguarda XML, tra le applicazioni di particolare interesse per gli studi umanistici sono da segnalare i linguaggi XHTML, SMIL, SVG, ma anche la trascrizione della DTD SGML del progetto TEI (Text Encoding Initiative; cfr. *infra* § 5.5) in XML.

Nella prospettiva del Semantic Web assume poi particolare interesse la sintassi RDF, che si propone in qualità di standard per la codifica dei metadati e OWL, il linguaggio per la definizione delle ontologie.

### 5.1 *HyperText Markup Language (HTML)*

Nell'anno 1990, Tim Berners-Lee, un software consultant del CERN, per promuovere un efficace scambio dei documenti tra i ricercatori ginevrini propose un suo progetto all'amministrazione dei laboratori. Dato lo scarso interesse suscitato, decise di pubblicizzare il suo lavoro su alcuni newsgroup

e di mettere a disposizione il suo lavoro in Internet. Esso consisteva nel protocollo HTTP, nello schema della sintassi URL, nel primo webBrowser (che chiamò WWW), nel primo software per webServer e nel linguaggio HTML. HTML<sup>7</sup> è un'applicazione ottenuta dallo standard ISO 8879 (SGML); è un linguaggio di marcatura dichiarativo, non proprietario, "orientato al video", con capacità ipertestuali. Un documento HTML è un unico file codificato in ASCII, che include sia il testo sia gli elementi di marcatura dichiarativa usati per rappresentare il contenuto.

La straordinaria diffusione che il linguaggio ha ottenuto è dovuta alla sua semplicità. Uno standard semplice e condiviso offre indiscutibili vantaggi. Con HTML, inoltre, è nata la figura dell'utente programmatore, che cura la redazione dei contenuti e la codifica per la diffusione lavorando in un unico ambiente software. HTML è stato nei primi anni sottoposto ad una continua espansione. I due maggiori produttori di tecnologie per il web, Microsoft e Netscape, si sono dati battaglia sul campo dell'HTML, espandendone senza controllo la DTD e operando scelte tecniche spesso in netta contraddizione. Con l'obiettivo di stabilire delle regole comuni e indirizzare l'evoluzione del web, nel 1994 è nato il W3C, un'organizzazione internazionale fondata da Berners-Lee, supportata dal MIT (Massachusetts Institute of Technology), dal CERN, dalla Comunità Europea e dalla DARPA (Defense Advanced Research Projects Agency). Per quanto attiene HTML, il W3C, fino alla versione 4.0 del linguaggio, è stato poco ascoltato dai grandi produttori di software per il web. Poi, grazie anche ai buoni propositi di Microsoft e Netscape – che sono entrati nel consorzio – il W3C è riuscito a riprendere il controllo della DTD del linguaggio.

Al W3C oggi aderiscono anche Adobe, Hewlett Packard, IBM, Novell, SoftQuad, Sun Microsystems, e molte altre aziende leader dell'informatica. Nella sua ultima versione, la 4.01, HTML è uno standard ISO/IEC abbastanza stabile, identificato dal codice 15445:2000.

## *5.2 Extensible HyperText Markup Language (XHTML)*

Se HTML è un'applicazione SGM, XHTML<sup>8</sup> è HTML definito in XML. Il vocabolario del linguaggio base resta quello di HTML 4.01, l'ultima versione rilasciata dal W3C. Il linguaggio XHTML è però estensibile – con l'utilizzo dei *namespaces* – con altri schemi di marcatura.

XHTML è più rigido di HTML. Mentre i browser cercano di visualizzare il codice HTML anche quando non è ben formato, una pagina codificata in XHTML per essere valida deve essere formalmente corretta, nell'assoluto

<sup>7</sup> <http://www.w3.org/MarkUp/>.

<sup>8</sup> <http://www.w3.org/TR/xhtml1/>.

rispetto delle specifiche del linguaggio originario. Ciò, se da una parte può creare qualche difficoltà in più, garantisce che un documento XHTML sia portabile su piattaforme e programmi diversi.

Questa tecnologia ha attratto le aziende che intendono raggiungere diverse tipologie di utenti, che accedono alla rete oltre che con i browser più diffusi per i PC, con telefoni cellulari WAP, che dispongono di schermi ridottissimi, o con sistemi di navigazione web particolari, addirittura privi di schermo grafico, come quelli destinati ai non vedenti, dove il testo è letto da un sintetizzatore vocale. Si stima che nei prossimi anni larga parte del traffico di rete sarà generata da nuovi tipi di dispositivi elettronici; un linguaggio di marcatura a prova di errori sintattici può semplificare i software di questi apparecchi e quindi abbattere i costi della loro realizzazione su larga scala.

### 5.3 Extensible Stylesheet Language (XSL)

XSL (Extensible Stylesheet Language)<sup>9</sup> è un'applicazione in sintassi XML creata per visualizzare i documenti XML. XSL ha le stesse capacità dei fogli di stile CSS (Cascading Style Sheets) ma permette anche di ristrutturare l'albero gerarchico dei documenti. Un documento XML può dichiarare nella sua intestazione l'URL di uno specifico foglio di stile XSL.

XSL funziona attraverso "template" da applicare agli elementi XML e utilizza un doppio processo: prima ricostruisce un albero strutturale, poi procede formattando gli elementi. Nel ricostruire l'albero, gli elementi del documento sorgente possono essere ricomposti in modo significativamente diverso. Per esempio, disponendo di un documento testuale marcato in XML che descriva uno scavo archeologico, nel documento generato da XSL si potrebbe aggiungere una tabella iniziale con funzioni di sommario, creata dal filtraggio selettivo degli elementi <settore dello scavo> trasformati in ancore ipertestuali che puntano al contenuto di tali settori, riorganizzati sotto forma di presentazione tabellare. Con un altro foglio di stile, si potrebbe invece presentare il documento originale sotto forma di semplice testo, così come redatto dai responsabili dello scavo.

Il linguaggio è costituito da tre parti: XSLT, Xpath, XSL:fo. Per selezionare gli elementi del documento sorgente si usa la sintassi definita in Xpath. Per modificare l'albero strutturale si usa la sintassi di XSLT, mentre XSL:fo si utilizza prevalentemente nelle trasformazioni XML → altri formati (PDF, TEX, RTF). Per generare la più semplice trasformazione XML → HTML si usano i template; tutto quello che è inserito in un template viene infatti trasferito direttamente nel file di uscita in sostituzione dell'elemento identificato dal template.

<sup>9</sup> <http://www.w3.org/TR/xsl/>.

#### 5.4 *Mathematical Mark-Up Language (MathML)*

MathML<sup>10</sup> è stato creato dal W3C Math Working Group con lo scopo di permettere l'inclusione di espressioni matematiche nelle pagine web. L'attuale versione è la 2.0. L'ultima versione delle specifiche è stata rilasciata nell'ottobre 2003. Poiché MathML è scritto in XML esso si integra specialmente con XHTML, con il quale condivide l'infrastruttura. I namespace permettono infatti di mescolare facilmente le semantiche di due diverse DTD.

MathML è descritto da una DTD che limita elementi ed attributi a quelli dichiarati. Il linguaggio è direttamente supportato da Netscape 7 e da Explorer 6, che lo visualizzano abbastanza bene. I browser che non lo supportano direttamente usano invece plug-in o applet specifiche.

#### 5.5 *Le linee guida TEI per la codifica di testi letterari*

Nel 1988 con il supporto di numerose istituzioni accademiche internazionali è stato avviato un progetto per sviluppare uno standard per la codifica dei testi in formato elettronico. Questo progetto è stato denominato Text Encoding Initiative (TEI)<sup>11</sup> e ha visto la partecipazione di moltissimi studiosi. Le linee guida TEI forniscono un ambiente per creare, sulla base di frammenti di DTD modulari, specifiche DTD per la marcatura di testi. Di fatto TEI propone schemi di codifica definiti, comunque personalizzabili, che costituiscono un vero e proprio linguaggio per la marcatura dichiarativa. Le DTD TEI nascono in ambito SGML; le prime specifiche sono state pubblicate nel 1991 con il titolo Guidelines for Electronic Text Encoding and Interchange, TEI P1. Nel 1992 è stata emessa la versione P2 e nel 1994 la P3. Nel 2000 si è costituito il TEI Consortium, con sede a Bergen (Norvegia) e, nel 2002, è stata pubblicata la versione P4 (SPERBERG-McQUEEN, BURNARD 2002) che presenta la versione XML della DTD e i relativi fogli di stile.

La versione P5 attualmente in preparazione è attesa per la fine del 2004; dovrebbe essere espressa in XML Schema e dovrebbe permettere, tra le altre novità, di usare i namespace XML per inserire metadati espressi con altre semantiche nell'intestazione del documento e di aggiungere altri elementi, definiti in altri namespace, nel testo. Per queste esigenze, relativamente al contesto archeologico, chi scrive aveva già cercato di trovare una possibile soluzione (BARCHESI *et al.* 2003).

<sup>10</sup> <http://www.w3.org/Math/>.

<sup>11</sup> <http://www.tei-c.org/>.

## 6. I LINGUAGGI DEL WEB SEMANTICO

Attualmente, nelle idee del W3C, il web dovrebbe evolvere verso un nuovo modello logico, di cui si parla sempre più intensamente: il *Semantic Web*<sup>12</sup>. La graduale trasformazione dovrebbe avvenire attraverso l'uso dei metadati legati da relazioni logiche finalizzate a descrivere profondamente o come si dice, ontologicamente, il loro significato. I pilastri del web semantico saranno XML, che abbiamo già descritto, la struttura RDF e il linguaggio OWL.

### 6.1 Resource Description Framework (RDF)

RDF è una raccomandazione W3C pubblicata nel febbraio del 1999. È una struttura per descrivere e scambiare metadati. È molto semplice, in quanto usa solo tre concetti: risorsa (*resource*), proprietà (*property*), valore della proprietà (*value*). I tre concetti legati assieme costituiscono un'affermazione (*statement*).

Una **risorsa** è identificata da un URI (Uniform Resource Identifier); un tipo particolare di URI è ad esempio l'URL (Uniform Resource Locator) di un sito web: <http://www.progettocaere.rm.cnr.it>. Anche la collocazione di un libro in una biblioteca, o il suo codice ISBN è un URI, poiché rappresenta un valido identificativo di una risorsa, che non necessariamente deve essere disponibile via web.

Una **proprietà** è una caratteristica di una risorsa, cui può darsi un nome ed un valore specifico, ad esempio: `<Web Editor></Web Editor>`.

Un **valore** è il contenuto della proprietà, ad esempio: `<Web Editor>Claudio Barchesi </Web Editor>`.

Ogni risorsa può essere descritta con appropriate coppie proprietà-valore, utilizzando una sintassi di questo tipo:

```
<rdf:Description about='risorsa'>  
<proprietà 1> valore 1 </proprietà 1>  
<proprietà 2> valore 2 </proprietà 2>  
...  
<proprietà n> valore n </proprietà n>  
</rdf:Description>
```

ad esempio:

```
<rdf:Description about='http://www.progettocaere.rm.cnr.it'>  
<Content Editor>Paola Moscati</Content Editor>  
<Web Editor>Claudio Barchesi</Web Editor>  
<Date> 7/7/2004 </Date>  
</rdf:Description>
```

<sup>12</sup> Il sito W3C del Semantic Web è accessibile presso <http://www.w3.org/2001/sw/>.

Le specifiche RDF non si limitano certamente a questa breve descrizione; tuttavia il semplice sistema a tre membri qui descritto resta immutato.

## 6.2 *Web Ontology Language (OWL)*

L'obiettivo del web semantico è quello di trasformare il web in un sistema di conoscenza distribuito, ove le informazioni siano presenti, accessibili e comprensibili alle macchine. Attualmente l'interrogazione di un motore di ricerca web restituisce moltissimi risultati, ma la maggior parte solo parzialmente significativi. Le macchine infatti non sono in grado di ragionare senza adeguate istruzioni sul significato dei testi. L'inserimento di una serie di parole nella finestra di un motore di ricerca ontologico permetterà invece ricerche intelligenti. I metadati saranno la base del web semantico e per descrivere le relazioni dei metadati si userà OWL. Questo linguaggio, la cui ultima specifica edita (ad oggi) è del febbraio 2004, dovrebbe permettere di rappresentare in forma esplicita le relazioni dei termini usati in un vocabolario di metadati, e il loro significato, costituendo quella che è definita un'ontologia.

Le ontologie, pur restando liberamente definibili, dovrebbero trovare poi modo d'affermarsi all'interno di specifiche comunità, che potranno così interpretare meglio i loro dati.

Un esempio di questo nuovo scenario, nell'ambito delle applicazioni interessanti per l'archeologia, è rappresentato dal progetto sviluppato dal CIDOC CRM<sup>13</sup> – in corso di registrazione come standard ISO – una struttura formale per descrivere i concetti impliciti ed espliciti e le relazioni usate nella documentazione dei beni culturali.

C.B.

## 7. LINGUAGGI DI MARCATURA LEGATI ALLA GRAFICA E ALL'IPERMEDIALITÀ

Le applicazioni dei linguaggi di marcatura non si limitano ai testi: uno degli aspetti più dibattuti per la diffusione dei dati grafici consiste nella formalizzazione di linguaggi che permettano di ovviare ai limiti dovuti all'utilizzo di piattaforme di lavoro diverse, con software ed hardware specifici, che spesso hanno difficoltà a dialogare fra loro<sup>14</sup>. In aggiunta, applicazioni GIS e di realtà virtuale sono ormai normali strumenti di lavoro nella maggior parte dei progetti archeologici, i cui risultati vengono diffusi attraverso diversi media, con una netta prevalenza di Internet.

Uno dei maggiori limiti del web è legato alla limitata gestione della grafica consentita da HTML. In una pagina HTML, infatti, è possibile inseri-

<sup>13</sup> <http://cidoc.ics.forth.gr/index.html>.

<sup>14</sup> Sono note le difficoltà di scambio di dati fra sistemi operativi Microsoft, Macintosh e Linux.

re immagini soltanto come file esterni in formato GIF, BMP, JPG. Per potenziare HTML sono state proposte tecnologie ausiliarie e nuovi linguaggi: alcune soluzioni hanno avuto seguito, mentre altre sono state nel tempo abbandonate.

La base di partenza è comunque individuabile in XML, le cui potenzialità sono state utilizzate per la creazione di nuovi linguaggi, in particolare per sviluppare applicazioni GIS e per la rappresentazione di oggetti grafici sia vettoriali che raster, bidimensionali o tridimensionali.

Per la distribuzione di dati vettoriali in rete sono state avanzate numerose proposte, soprattutto di formati che, tuttavia, richiedono plug-in per la visualizzazione e non hanno possibilità di animazione; si tratta in particolare di SVF (Simple Vector Format), DWF (Drawing Vector Format) e PDF (Portable Document Format).

### 7.1 *Synchronized Multimedia Interface Language (SMIL)*

Il W3C ha presentato le raccomandazioni di SMIL 2.0 nell'agosto 2001<sup>15</sup>. Quelle emesse precedentemente, nel 1998, relative alla versione 1.0, erano state respinte da Microsoft che sosteneva l'inutilità del linguaggio. SMIL è un linguaggio per la creazione di documenti con contenuti multimediali. Nasce per aggiungere animazioni alle pagine web senza dover ricorrere a Java, Javascript o al Dynamic HTML. SMIL non sostituisce i molti programmi che permettono la realizzazione di file audio e video ma li integra in unico ambiente di presentazione. È un linguaggio basato su XML, nato per creare il web multimediale e può essere usato per estendere il linguaggio XHTML con i marcatori del proprio namespace. SMIL, infatti, permette l'inclusione in un unico documento di più file multimediali, la loro riproduzione in parallelo o in serie, la specifica del timing d'esecuzione e il formato di rappresentazione (layout). Inizialmente la versione 1.0 del 1998 era supportata solo da Real Player, mentre la versione 2.0 è stata integrata già parzialmente nella versione 5.5 di Internet Explorer e poi pienamente nella versione 6.0.

SMIL è implementato da Microsoft con il nome Timed Interactive Multimedia Extensions for HTML, HTML+TIME, (XHTML+SMIL), che fa riferimento al relativo standard del W3C.

### 7.2 *Geography Markup Language (GML)*

Data l'inadeguatezza degli attuali formati a gestire e trasferire in rete informazioni geografiche e dati strutturati di un database, è stato sviluppato dall'Open GIS Consortium<sup>16</sup> il linguaggio GML. Si tratta di un linguaggio

<sup>15</sup> <http://www.w3.org/TR/smil20/>.

<sup>16</sup> <http://www.opengis.org>.

per rappresentare i dati geografici ed i loro rapporti topologici basato su XML. Ha una struttura aperta, indipendente dalla piattaforma di lavoro, che supporta la descrizione di applicazioni spaziali e di banche dati e la loro condivisione in rete. L'ultima versione è GML 3.0, presentato il 29 gennaio 2003<sup>17</sup>. Come aspetto innovativo ha una struttura modulare ad oggetti, gestisce geometrie e topologie anche in 3D, ha una componente temporale, un sistema di riferimento di coordinate e offre diversi sistemi di unità di misura.

Occorre puntualizzare che ad oggi il linguaggio GML codifica esclusivamente il contenuto dell'informazione geografica e descrittiva, mentre per quanto riguarda la rappresentazione degli oggetti grafici, all'interno degli attuali browser, richiede l'utilizzo di altri linguaggi di rappresentazione quali, ad esempio, Scalable Vector Graphics (SVG), Vector Markup Language (VML), X3D (Extensible 3D Language) oppure le Applet Java.

Una critica che si può muovere al linguaggio è che è stata utilizzata la sigla GML, già esistente per rappresentare il Generalized Markup Language (cfr. *supra* § 3.2).

### *7.3 Vector Markup Language (VML)*

È un linguaggio basato su XML per la grafica vettoriale bidimensionale, proposto da Microsoft al W3C il 13 maggio 1998. Incluso nel browser Internet Explorer, a partire dalla versione 5.0, è compatibile con gli standard Microsoft ed è stato infatti introdotto in Front Page XP. La visualizzazione dei dati codificati in VML richiede una trasformazione di XML in HTML tramite XSL e ha, pertanto, bisogno dei CSS (Cascading Style Sheets)<sup>18</sup>. Il limite di questo linguaggio, essendo proprietario di Microsoft, consiste nel fatto che non tutti i browser lo supportano e pertanto i file vettoriali vengono visualizzati come immagini raster in formato GIF.

### *7.4 Precision Graphics Markup Language (PGML)*

Si tratta di un linguaggio per la grafica vettoriale bidimensionale proposto da Adobe, Sun e Netscape al W3C il 10 aprile del 1998<sup>19</sup>. Si basa sul modello dei dati del linguaggio descrittivo PostScript e del formato PDF (Portable Document Format). Il limite di questo linguaggio consiste nel fatto che, essendo proprietario di Adobe, non è stato implementato e presenta a volte difficoltà di interazione con XML. Il PGML è stato pertanto generalmente abbandonato per orientarsi verso la proposta del W3C del nuovo linguaggio SVG.

<sup>17</sup> <http://www.opengis.org/docs/02-023r4.pdf>.

<sup>18</sup> <http://msdn.microsoft.com/library/default.asp?url=/workshop/author/vml/default.asp>.

<sup>19</sup> <http://www.w3.org/TR/1998/NOTE-PGML-19980410>.

## 7.5 Scalable Vector Graphics (SVG)

Il W3C ha cercato di legare le due proposte di VML e PGML, pubblicando le linee guida per SVG (Scalable Vector Graphics). Si tratta di un linguaggio per descrivere oggetti grafici bidimensionali, basato su XML<sup>20</sup>. Ammette tre tipi di oggetti grafici: dati vettoriali, immagini raster e testi. Richiede una DTD e deve avere un CSS in un formato grafico ASCII.

Viene definita un'area di lavoro (viewport), in cui si rappresentano gli oggetti grafici tramite coordinate cartesiane ed unità di misura. Le unità di misura consentite sono *px* Pixels (dimensione di default), *pt* Punti (1/72 di pollice), *pc* Pica (1/6 di pollice), *cm* Centimetri, *mm* Millimetri, *in* Pollici. Gli oggetti grafici possono essere raggruppati e trasformati (ruotati, ridimensionati, traslati). I vantaggi di SVG sono molteplici: si tratta di un linguaggio Open Source e i file sono codificati con tag basati su XML; inoltre permette la distribuzione dei dati nel web e può venire usato come formato grafico per la rappresentazione dei dati GML.

Il limite è costituito dal fatto che al momento nessun software GIS sul mercato consente attualmente di esportare i dati in SVG, per cui è necessario convertirli in un formato testuale e poi intervenire con un programma di trasformazione per ottenere il codice SVG. Attualmente SVG può essere visualizzato solo attraverso un plug-in gratuito della Adobe<sup>21</sup>.

Al momento è comunque il linguaggio più utilizzato nell'ambito della diffusione in rete dei risultati di indagini archeologiche, come ad esempio gli scavi, in quanto permette di visualizzare in modo dinamico anche la componente geografica dei dati<sup>22</sup>.

## 8. LINGUAGGI PER LA REALTÀ VIRTUALE

Le ricostruzioni tridimensionali e la realtà virtuale si pongono come strumenti di conoscenza per le loro potenzialità comunicative nel settore dei beni culturali. Tuttavia i problemi legati alla loro distribuzione sono molteplici e, come per le applicazioni GIS, la rete viene indicata come la soluzione da perseguire. La rappresentazione dei dati tridimensionali in Internet non ha però avuto una grande diffusione per una serie di motivazioni non tutte dovute a limiti dello sviluppo tecnologico. Inoltre i fattori che determinano il vasto utilizzo di un formato o di una tecnologia in rete sono il risultato di una sintesi fra efficienza e diffusione.

<sup>20</sup> La versione più recente è SVG 1.2 pubblicata dal W3C il 18 marzo 2004 (<http://www.w3.org/TR/SVG12/>).

<sup>21</sup> Adobe SVG Viewer: <http://www.adobe.com/svg/viewer/install/>.

<sup>22</sup> Ad esempio, The York Archaeological Trust ha presentato i risultati delle campagne di scavo urbane utilizzando SVG: <http://www.yorkarchaeology.co.uk/dccook/main.htm> (progetto presentato anche al CAA 2004).

Il limitato impiego del 3D nella rete è dovuto principalmente alla mancanza di un comune linguaggio grafico di descrizione dei modelli. Si è operato un tentativo di standardizzazione per le rappresentazioni tridimensionali con il linguaggio VRML, che tuttavia non è immune da difetti in quanto richiede una elevata potenza di calcolo e velocità di trasmissione in rete.

La mancanza di una piattaforma comune di visualizzazione tridimensionale ha dato origine a diverse tecniche, con il conseguente moltiplicarsi di plug-in necessari per la visualizzazione di singoli formati grafici proprietari.

### 8.1 *Virtual Reality Modeling Language (VRML)*

Si tratta di un linguaggio sviluppato dal W3C con lo scopo di descrivere la geometria e il comportamento di una “scena 3D” o “mondo virtuale”. Esso non è particolarmente flessibile e richiede un plug-in per la visualizzazione.

Il primo prototipo di VRML è stato presentato alla First International Conference of Internet nel marzo del 1994. Il prototipo, denominato Labyrinth, era la prima realizzazione di un browser indipendente dalla piattaforma di utilizzo. In quella occasione si utilizzò per la prima volta il termine VRML. Nel maggio 1995, dopo aver scelto Open Inventor come formato ufficiale di file, venne presentata la versione VMRL 1.0.

Il VRML è un linguaggio descrittivo di oggetti tridimensionali, i file sono in formato ASCII e possono essere visualizzati mediante un interprete (parser) o browser VRML. Fra i browser VRML più diffusi si può citare Cosmo Player (Cosmo Software, Computer Associates), Blaxxun Contact (Blaxxun Inc.) e Cortona (Parallel Graphics). Essendo il VRML un codice interpretato, i risultati della visualizzazione possono variare a seconda del browser utilizzato.

Nell'agosto del 1996 è stata presentata la versione 2.0, che accoglieva le proposte di Silicon Graphics nell'ambiente di sviluppo Moving Worlds. L'aspetto innovativo della versione 2.0 consisteva nelle funzioni di animazione e interazione con l'utente. L'ulteriore sviluppo del linguaggio è avvenuto con la presentazione di VRML97 che è diventato di fatto uno standard<sup>23</sup>. Per la sua semplicità semantico-sintattica nella strutturazione della scena e per la possibilità di integrazione con i sistemi operativi e i software più diffusi rappresenta ancora oggi il formato più utilizzato, soprattutto per la distribuzione via Internet di ambienti tridimensionali. Sono numerosi, infatti, i programmi che esportano i dati in formato .wrl<sup>24</sup>.

Notevoli limiti, comunque, derivano dalla dimensione dei file in for-

<sup>23</sup> *International Standard ISO/IEC 14772-1:1997: VRML 97* (<http://www.iso.ch/iso/en/ISOOnline.frontpage>).

<sup>24</sup> Ad esempio i principali 3D Studio Max (Autodesk) e Maya (Alias).

mato VRML, dalla necessità di utilizzare visualizzatori e dal fatto che l'aggiornamento dei modelli è estremamente laborioso.

Nell'archeologia virtuale si sono utilizzate le capacità di programmazione del linguaggio VRML immediatamente dopo la sua presentazione e alcuni studiosi hanno realizzato modelli sfruttando interamente le possibilità di VRML. Tuttavia la direzione verso cui ci si muove nei progetti di realtà virtuale è piuttosto l'utilizzo del linguaggio di programmazione C++ – in associazione alle librerie OpenGL – il quale è stato privilegiato al VRML, in quanto più flessibile e adeguato alle nuove schede grafiche.

## 8.2 Extensible 3D Language (X3D)

Si tratta di una rielaborazione di VRML con una sintassi XML. Dal 2001 il W3C si è orientato verso la formalizzazione di un nuovo standard, che ha preso il nome di X3D<sup>25</sup>, inizialmente denominato VRML-NG (Next-Generation). Tale standard ha come obiettivo il superamento di alcune caratteristiche che avevano limitato la diffusione di VRML in Internet, quali, ad esempio, la dimensione dei file dei modelli. Una caratteristica importante di X3D consiste nel fatto che non richiede l'installazione di moduli aggiuntivi per la visualizzazione, in quanto si basa sul motore 3D di Java, che è già supportato dalle ultime versioni dei vari browser, o comunque su delle semplici Applet Java che funzionano da viewer.

A partire dal febbraio 2002 il W3C ha rilasciato la versione iniziale del linguaggio X3D, che da allora è in continua evoluzione. Esso si presenta come lo standard più utilizzato in molti campi di applicazioni; ad esempio è stato adottato da parte del Moving Picture Experts Group (MPEG) per lo standard multimedia MPEG-4.

Le ricerche in corso in campo archeologico sono volte sia alla formalizzazione di modelli applicativi, sia alla definizione di piattaforme software<sup>26</sup> che consentano l'integrazione tra diversi linguaggi di modellazione degli ambienti tridimensionali e il trasferimento di informazioni strutturate sulla rete Internet in modo tale da poter sfruttare le qualità specifiche di VRML e X3D (GOODRICK, EARL 2004).

## 9. CONCLUSIONI

In questo contributo si è cercato di delineare la storia della evoluzione dei linguaggi dichiarativi che trovano applicazione in archeologia e nel settore

<sup>25</sup> [http://www.web3d.org/x3d/specifications/x3d\\_specification.html](http://www.web3d.org/x3d/specifications/x3d_specification.html).

<sup>26</sup> Ad esempio da aprile 2004 l'Archaeology Technologies Laboratory (ATL) dell'Università del North Dakota sta sviluppando un X3D Open Source che può essere visualizzato con Maya 4.0 e seguenti: [http://atl.ndsu.edu/resources/maya\\_x3d.php](http://atl.ndsu.edu/resources/maya_x3d.php).

dei beni culturali. Come emerge chiaramente da quanto delineato, la ricerca nel campo delle scienze umanistiche ha fatto notevoli progressi a livello sia tecnico sia cognitivo, grazie all'applicazione di metodologie informatiche, che hanno permesso di recuperare ed elaborare un notevole patrimonio di informazioni. Questo processo ha indirizzato quindi la ricerca verso l'elaborazione di strumenti idonei a preservare la struttura logica dell'informazione, con un approccio che cerca di mediare la dualità fra i linguaggi del computer e quelli che si basano sulla sintassi e la semantica del linguaggio naturale (GARDIN 2003).

La codifica dei dati, ovvero il processo di trasformazione delle informazioni per renderle interpretabili non solo dalle persone ma anche dalle macchine, ha generato molti linguaggi che vanno dall'ASCII ai linguaggi procedurali e dichiarativi e alle loro applicazioni, sia nel campo dei dati testuali che di quelli grafici bidimensionali e tridimensionali.

Da tutto ciò si evince l'utilità di una riflessione teorica e tecnica sui linguaggi che permettono il trattamento informatico dei dati senza alterarne la struttura originaria. Per quanto riguarda i documenti testuali è importante usare linguaggi che consentano di interpretarne il valore connotativo. Nell'ambito dei beni culturali, le proposte e le applicazioni più innovative dei linguaggi di marcatura testuale sono state generalmente sviluppate nel settore archivistico, dove la marcatura testuale dei documenti è finalizzata a trasformare un testo narrativo in una fonte di informazioni strutturate.

Ciò che appare, però, più difficoltoso nelle applicazioni non è tanto la fase di marcatura delle informazioni, quanto il passaggio allo sviluppo di metodologie di interrogazione dei dati. In Italia si segnalano, ad esempio, i progetti del Centro di Ricerche Informatiche per i Beni Culturali della Scuola Normale di Pisa (CORTI, PAOLI, SPRUGNOLI 1998 e da ultimo COLAZZO *et al.* 2002)<sup>27</sup> e dell'Istituto di Studi sulle Civiltà Italiane e del Mediterraneo Antico del CNR (cfr. da ultimo BARCHESI *et al.* 2003; MOSCATI c.s.)<sup>28</sup>, che hanno realizzato una metodologia di codifica XML e specifiche procedure di interrogazione dei testi in rete.

Infatti, se da una parte la codifica dei dati è un aspetto fondamentale del tema trattato, l'altro aspetto consiste nell'analisi degli standard di diffusione delle informazioni, in cui Internet assume un ruolo importante. L'interesse in ambito archeologico per le potenzialità di XML legate alla grafica raster, vettoriale e ad applicazioni GIS, soprattutto verso la rete, è ancora piuttosto recente. Tuttavia SVG è già diventato uno standard per l'acquisizione di dati vettoriali come dimostra ad esempio la sua adozione da parte dell'AHDS<sup>29</sup>.

<sup>27</sup> [http://www.cribecu.sns.it/analisi\\_testuale/settore\\_informatico/tresy/](http://www.cribecu.sns.it/analisi_testuale/settore_informatico/tresy/).

<sup>28</sup> <http://www.progettocaere.rm.cnr.it/>.

<sup>29</sup> <http://ahds.ac.uk/depositing/deposit-formats.htm>.

Il campo in cui maggiormente si è discusso sull'utilità di adottare degli standard è quello della realtà virtuale e della realtà incrementata (BARCELÓ 2000), sia per la pubblicazione scientifica, che per le implicazioni culturali. Questo è infatti uno dei maggiori obiettivi dell'UCLA Cultural VR Lab, ovvero la creazione di modelli di realtà virtuale scientificamente convalidati che diventino strumenti di conoscenza (FRISHER c.s.). Sono stati anche scritti veri e propri manuali per l'utilizzo di linguaggi e standard per la realtà virtuale nell'ambito di progetti archeologici (FERNIE, RICHARDS 2002)<sup>30</sup>.

Inoltre, poiché ci si è resi conto della necessità primaria di utilizzare un approccio metodologico che consenta di rendere trasparenti i codici che generano i modelli di realtà virtuale, l'interesse si è recentemente incentrato sulle problematiche relative alla normalizzazione delle procedure di lavoro, alla portabilità dei dati e al confronto dei risultati. A tale proposito, ad esempio, l'Istituto Centrale per il Catalogo e la Documentazione indica VRML come standard di riferimento<sup>31</sup>, così come la proposta del CVRO (The Cultural Virtual Reality Organisation), che suggerisce l'utilizzo di linguaggi Open Source quali VRML e X3D e la documentazione dei modelli con metadati in XML (RYAN 2001; FRISCHER *et al.* 2002).

L.C.

CLAUDIO BARCHESI

LETIZIA CECCARELLI

Istituto di Studi sulle Civiltà Italiane  
e del Mediterraneo Antico  
CNR – Roma

## BIBLIOGRAFIA

- BARCELÓ J., FORTE M., SANDERS D. (eds.) 2000, *Virtual Reality in Archaeology*, BAR International Series 843, Oxford, Archaeopress.
- BARCELÓ J. 2000, *Visualizing what might be: An introduction to virtual reality techniques in archaeology*, in BARCELÓ, FORTE, SANDERS 2000, 9-35.
- BARCHESI C., MOSCATI P., SANTORO P., SCARPATI D. 2003, *Ricerche archeologiche sul campo e archivi digitali: il manoscritto di Ercole Nardi*, «Archeologia e Calcolatori», 14, 245-273.
- COLAZZO D., SARTIANI C., ALBANO A., GHELLI G., MANGHI P., LINI L., PAOLI M. 2002, *Atyped text retrieval query language for XML documents*, «Journal of American Society for Information Science and Technology», Special Topic Issue on XML and Information Retrieval, 53, 6, 467-488.

<sup>30</sup> Nel 2002 è stato anche pubblicato da parte dell'AHDS nell'ambito della serie Guides to Good Practice, *Creating and Using Virtual Reality: a Guide for the Arts and Humanities* ([http://vads.ahds.ac.uk/guides/vr\\_guide/index.html](http://vads.ahds.ac.uk/guides/vr_guide/index.html)).

<sup>31</sup> [http://www.iccd.beniculturali.it/download/norme\\_300/normative\\_300.pdf](http://www.iccd.beniculturali.it/download/norme_300/normative_300.pdf).

- CORTI F., PAOLI M., SPRUGNOLI R. 1998, *TReSy un motore di ricerca testuale per documenti strutturati*, «Bollettino d'Informazioni – Centro di Ricerche Informatiche per i Beni Culturali», 8, 1, 49-71.
- FERNIE K., RICHARDS J. 2003, *Creating and Using Virtual Reality: A Guide for the Arts and Humanities*, Oxford, Oxbow Books.
- FRISCHER B., NICCOLUCCI F., RYAN N., BARCELÓ J. 2002, *From CVR to CVRO. The past, present, and future of cultural virtual reality*, in F. NICCOLUCCI (ed.), *Proceedings of the VAST Euroconference (Arezzo 2000)*, BAR International Series 834, Oxford, Archaeopress, 7-18.
- FRISCHER B. c.s., *Mission and recent projects of the UCLA Cultural Virtual Reality Laboratory*, in *Acts of Virtual Retrospect 2003 (Biarritz, France, 2003)* (<http://www.cvrlab.org/research/images/FrischerVirtRetro2003.pdf>), in corso di stampa.
- GARDIN J.C. 2003, *Archaeological discourse, conceptual modelling and digitalisation: an interim report of the logicist program*, in M. DOERR, A. SARRIS (eds.) 2003, *CAA. The Digital Heritage of Archaeology, Heraklion*, Hellenic Ministry of Culture, 5-12.
- GOODRICK G., EARL G. 2004, *A manufactured past: virtual reality in archaeology*, «Internet Archaeology», 15 ([http://intarch.ac.uk/journal/issue15/earl\\_toc.html](http://intarch.ac.uk/journal/issue15/earl_toc.html)).
- MCLUHAN M. 1967, *The Medium is the Massage: An Inventory of Effects*, New York, Simon & Schuster.
- MOSCATI P. c.s., *Linguaggi di marcatura per la conservazione e la valorizzazione dell'informazione archeologica*, in *Atti del Convegno Internazionale Archivi informatici per il patrimonio culturale (Roma, Accademia Nazionale dei Lincei, 2003)*, in corso di stampa.
- RYAN N. 2001, *Documenting and validating virtual archaeology*, «Archeologia e Calcolatori», 12, 245-273.
- SPERBERG-MCQUEEN C.M., BURNARD L. (eds.) 2002, *TEI P4: Guidelines for Electronic Text Encoding and Interchange, Text Encoding Initiative Consortium. XML Version*, Oxford, Providence, Charlottesville, Bergen.

## ABSTRACT

This article presents a detailed overview of the principal languages for the representation, interchange and exploitation of data, both textual and graphical. In particular, a detailed discussion is made of the procedure of text encoding. The approach taken in the article emphasises the importance of the World Wide Web for data dissemination and the fundamental issue of standards: HTML, XML and its derivate languages are analysed in detail.

Importance has been given to the languages that represent not only the characters that textual sources contain but also the structure, content and appearance of the data. Two types of markup languages are presented: procedural and descriptive. A procedural markup specifies how the document should be presented. Descriptive (or logical) markup languages describe the structure of a document, such as SGML. The article considers the topics of international standards as the TEI Guidelines for Electronic Text Encoding and Interchange for the description of marked-up electronic texts and the RDF metadata recommendation. The first section concludes with a presentation of the innovative aspects of the Semantic Web.

The second part focuses on spatial, graphical and multimedia data, and their display and exchange over the Web. The development of the Geography Markup Language (GML) is introduced and discussed, as well as other vector formats derived by XML, such as SVG, to construct structured spatial and non-spatial information for data sharing over the Web. Importance has also been given to the virtual reality languages such as VRML, an ISO standard, and the XML-based X3D. In conclusion the article aims to present a broad view not only of the technical aspects of data encoding but also the analysis of the standards, which are fundamental in the light of data interoperability and exchange.

