

LA GESTIONE DELLA CONOSCENZA IN ARCHEOLOGIA: MODELLI, LINGUAGGI E STRUMENTI DI MODELLAZIONE CONCETTUALE DALL'XML AL SEMANTIC WEB

1. INTRODUZIONE

Nonostante nel corso degli ultimi anni un numero sempre crescente di ricercatori, che svolgono la loro attività nel settore dei beni culturali, si stia rendendo conto dei vantaggi che possono derivare dall'utilizzo di metodologie e strumenti di gestione della conoscenza (Knowledge Management-KM), è tuttora percepibile una certa carenza di consapevolezza dei principi di base di questa disciplina. Concetti chiave come rappresentazione della conoscenza (Knowledge Representation-KR), metadati, modellazione concettuale, interoperabilità sintattica, interoperabilità semantica, linguaggi di marcatura, ontologie possono infatti risultare di non facile comprensione (e ancor più di difficile utilizzo) per ricercatori provenienti da una formazione di tipo umanistico.

Il KM deriva – e mutua teorie, metodologie e strumenti – da 50 anni di ricerca nel campo dell'Intelligenza Artificiale (Artificial Intelligence-AI). È infatti nell'estate del 1956 che Herbert Simon, Marvin Minsky, Claude Shannon e altri ricercatori si riunirono al Dartmouth College di Hanover, nel New Hampshire, per discutere la possibilità di simulare apprendimento e ragionamento umani, utilizzando macchine di recente invenzione: i computer. Così, con molto ottimismo, essi ritennero che i tempi fossero maturi perché questi studi costituissero una disciplina autonoma e, non senza contrasti, gli attribuirono il nome di "Intelligenza Artificiale". E proprio l'IA fornì i primi, pionieristici, studi sulla conoscenza e specialmente sulla sua rappresentazione, producendo una serie di formalismi che avrebbero preparato il terreno alle più recenti applicazioni di tecniche di KM nel settore del business (SENGE 1990; NONAKA, TAKEUCHI 1995) e all'ancor più recente "vision" del Semantic Web (BERNERS-LEE *et al.* 2001).

Il contesto applicativo attuale è sempre più basato sul web, che per sua natura è un mondo aperto in cui interagiscono realtà diverse per organizzazione e cultura e che richiede strumenti e tecnologie che consentano una reale "interoperabilità" delle applicazioni (un aspetto non meramente tecnologico, ma soprattutto semantico). Il presente articolo tende dunque a fare chiarezza sul quadro di riferimento teorico, attraverso l'analisi concreta di materiali archeologici. Infatti, mentre può risultare abbastanza semplice mutuare definizioni e concetti astratti o, addirittura, creare artificialmente modelli concettuali anche molto articolati (cfr. ad

esempio il CIDOC CRM¹, ormai giunto quasi a livello di standard ISO), ben più arduo risulta superare il banco di prova del confronto con oggetti del mondo reale.

In accordo con questa analisi, e con la necessità di passare dagli aspetti teorici a quelli applicativi, il contributo è strutturato in tre parti: la prima offre una base teorica che mette a disposizione, anche dei non esperti, gli strumenti per poter affrontare aspetti più operativi; la seconda descrive, attraverso esempi concreti, sia il modello di rappresentazione della conoscenza sia lo strumento che sono stati poi utilizzati per analizzare concretamente una classe di materiali, le urne etrusche, come esemplificato nella terza parte. L'obiettivo perseguito è dunque di fornire un punto di riferimento che faciliti l'avvicinamento al KM e contribuisca a fare chiarezza in una disciplina che sta riscuotendo un sempre maggiore successo ma anche mostrando, finora, un elevato livello di entropia.

Come quadro di riferimento per affrontare gli aspetti legati alla prima parte, e cioè agli aspetti teorici, ci si è serviti del cosiddetto "Semantic Web Wedding Cake" (Torta Nuziale del Web Semantico) che, partendo dai caratteri ammessi (Unicode), guida attraverso l'identificazione delle risorse sul web (Universal Resource Identifier-URI) per arrivare ad analizzare l'ambito sintattico prima (eXtensible Mark-up Language-XML) e semantico poi (Resource Description Framework-RDF), per terminare con le ontologie, la possibilità di operare ragionamenti (livello logico) e il "trust", cioè l'affidabilità delle risorse sul web.

La seconda parte è invece strutturata in accordo con il "metamodello" OPAL (Object, Process, and Actor Language), basato su un'architettura a tre livelli: 1) livello "meta", destinato alla specifica delle "idee di modellazione" (proprietà, relazione, ecc.), indipendenti dal dominio applicativo; 2) livello "intensionale", che contiene i concetti principali specifici di un determinato dominio; 3) livello "estensionale", in cui vengono raccolti gli oggetti del mondo reale che corrispondono ai concetti del dominio considerato.

La terza parte, infine, presenta un progetto sperimentale dedicato alla classificazione automatica di una classe di oggetti appartenenti al dominio considerato, cioè all'archeologia: i coperchi delle urne lapidee di fabbricazione etrusca, conservati nel Museo Guarnacci di Volterra.

2. KNOWLEDGE MANAGEMENT E RICERCA DI INFORMAZIONI

I processi essenziali nel KM sono legati alla possibilità da un lato di "reperire" le fonti di conoscenza rilevanti per il problema specifico, e dall'altro lato di "fornire" le fonti di conoscenza da utilizzare per risolvere tale proble-

¹ The CIDOC Conceptual Reference Model (<http://cidoc.ics.forth.gr/>). Attualmente è ISO/DIS 21127 (Draft International Standard) e prossimamente sarà portato a livello di FDIS (Final Draft International Standard).

ma. Schematicamente, possiamo individuare cinque processi: acquisizione, rappresentazione, elaborazione, condivisione e utilizzo della conoscenza. Il web – e in particolare il Semantic Web (cfr. da ultimo WORKSHOPSW 2004), che ne è la naturale evoluzione – costituisce un formidabile componente per supportare gran parte di questi processi. La conoscenza codificata nel Semantic Web è, infatti, rappresentata in maniera “elaborabile” dalla macchina, e quindi può essere utilizzata da componenti automatizzati, denominati “agenti software”.

La ricerca di informazioni sul web resta però tuttora un’esperienza spesso frustrante, perché i motori di ricerca esistenti sono poveri di semantica sia in fase di indicizzazione che in fase di ricerca. In fase di indicizzazione delle pagine web, non sempre è possibile identificare e utilizzare metainformazioni come l’autore, le parole chiave, etc., e spesso anche le informazioni che i progettisti inseriscono nei tag <meta> vengono ignorate dagli strumenti di indicizzazione. In fase di ricerca, viene consentito di combinare le parole con operatori di contesto (tutte le parole, una parola qualunque, nel titolo), ma in definitiva il risultato scaturisce sempre da una ricerca sulla presenza di parole chiave e dall’identificazione dei documenti più affini alla domanda posta. Resta comunque il problema di fondo che l’utente non sempre conosce o è posto in condizione di comprendere la semantica della strutturazione delle informazioni.

L’esistenza di proposizioni più ricche dal punto di vista espressivo permette invece agli utenti di ritrovare in maniera più facile ed efficace le informazioni necessarie per risolvere i problemi, e consente di indicizzare le risorse esistenti sul web in maniera più ricca rispetto al metodo tradizionale di associare alle risorse parole chiave o concetti. Diventa allora possibile formulare richieste più sofisticate, migliorando sia la precisione delle risposte ottenute, che il richiamo dei documenti pertinenti. Giusto a titolo di esempio, in assenza di questo tipo di arricchimento una query che richiedesse la restituzione di tutti i documenti in cui compaiono le parole chiave: “*Giulio II*” AND “*Michelangelo*” restituirebbe sia i documenti che descrivono opere/eventi in cui *Giulio II* è *committente* di un’opera di *Michelangelo* che quelli in cui *Giulio II* è *soggetto rappresentato* in un’opera di *Michelangelo*.

In presenza di documentazione strutturata, l’utente potrebbe effettuare una ricerca più precisa indicando che i termini ricercati devono comparire in specifiche parti del documento (per esempio, nel campo soggetto o committente). Tuttavia, questo richiede che l’utente sia a conoscenza della struttura dell’informazione (quello che tradizionalmente veniva indicato come tracciato record o struttura della scheda), e che questa struttura sia comprensibile indipendentemente dalla sua lingua madre o cultura. Nell’infrastruttura del Semantic Web, i concetti sono identificati in maniera univoca e indipendente dalla lingua, e agenti software sono in grado di realizzare

le associazioni e identificare le equivalenze terminologiche, grazie a una rappresentazione formalizzata della conoscenza e alla capacità di eseguire dei ragionamenti.

3. LE ONTOLOGIE

Nell'organizzare le informazioni si ricorre spesso al meccanismo della classificazione, che può però presentare diversi problemi, derivanti essenzialmente dalle differenze di dominio, terminologia, scelta e significato delle varie caratteristiche ritenute importanti, e dalle differenze tra le relazioni più significative.

È importante distinguere le ontologie dai meccanismi di classificazione. Mentre questi ultimi prestano attenzione alle esigenze di accesso all'informazione, basandosi su criteri predefiniti codificati mediante elementi "sintattici", le ontologie si concentrano sul "significato" dei termini e sulla "natura" e "struttura" di un dominio. Ne deriva che il problema essenziale è sostanzialmente quello della corrispondenza semantica (*semantic matching*) e dell'integrazione semantica.

Anche sottili differenze tra le diverse terminologie risultano importanti, e la definizione di vocabolari standard non è la soluzione al problema: essi sono spesso il risultato di anni di lavoro da parte di esperti, ma non sempre si adattano bene, soprattutto in presenza di domini di conoscenza eterogenei, e non sempre vengono utilizzati in maniera corretta. Un'ontologia, almeno nel senso in cui questo termine viene impiegato nel settore informatico, è un oggetto specifico, progettato per esprimere il "senso inteso" (*intended meaning*) di un vocabolario. Una caratteristica importante di un'ontologia è il suo livello di precisione, cioè il modo in cui è in grado di esprimere i concetti in maniera non ambigua e con un buon grado di copertura e di precisione rispetto al modello inteso.

La formalizzazione della conoscenza ha una lunga tradizione nel contesto dei beni culturali, e molti anni di lavoro sono stati spesi in questa direzione, realizzando glossari, cataloghi, thesauri. Il livello di formalizzazione dei concetti e delle loro relazioni, e quindi di precisione ontologica, cresce a mano a mano che si passa da cataloghi, glossari, tassonomie (sistemi di classificazione in cui i termini sono in relazione gerarchica, come in Iconclass²), a thesauri (vocabolari controllati strutturati in modo da evidenziare le relazioni semantiche tra termini e concetti, come AAT³) o, infine, a teorie assiomatizzate (DIGICULT 2003). In effetti, va sottolineato come lessici, tassonomie, vocabolari strutturati e thesauri non si prestino adeguatamente a collegare tra di loro

² <http://www.iconclass.nl/>.

³ http://www.getty.edu/research/conducting_research/vocabularies/aat/.

alberi concettuali diversi, mentre approcci formalmente più solidi (tra i quali va senz'altro citato CIDOC CRM) promettono di rappresentare la conoscenza in modo da poter supportare ragionamento e deduzioni.

In un contesto ampio e decentralizzato come quello del patrimonio culturale e del web, assume una particolare importanza l'integrazione dell'informazione. In questo processo risulta essenziale il ruolo giocato da una *core ontology*, il cui obiettivo è fornire un modello globale ed estensibile in cui possono essere messi in corrispondenza ed integrati i dati provenienti da fonti eterogenee. Questa forma canonica è in grado di fornire una singola base di conoscenza per strumenti e servizi cross-domain (resource discovery, browsing, data mining), riducendo la complessità combinatoria che deriverebbe dal tentativo di mettere in corrispondenza a due a due i singoli formati di metadati o le ontologie.

DOERR 2003 sottolinea la distinzione sottile, ma importante, tra una *core ontology* e la definizione di *core metadata* (come Dublin Core). Entrambe sono finalizzate all'integrazione dell'informazione, ma differiscono per l'importanza attribuita alla comprensibilità da parte di un lettore umano. I metadati sono compilati e utilizzati principalmente da esseri umani, mentre una *core ontology* è un modello formale utilizzato da strumenti che provvedono all'integrazione di varie fonti di dati e svolgono varie altre funzioni. Di conseguenza, mentre i fattori umani, e principalmente la leggibilità, costituiscono un elemento cardine nella definizione dei *core metadata*, una *core ontology* può accettare un livello di complessità maggiore, privilegiando la completezza e la correttezza logica rispetto alla comprensibilità umana.

Un interessante esempio di *core ontology* è CIDOC CRM, costituita da circa 80 classi e 130 proprietà, che si pone come modello di riferimento per il patrimonio culturale e per altri contesti, e consente di supportare il ragionamento spaziale e temporale. Per esempio, utilizzando informazioni annotate semanticamente con le classi di questa ontologia, sarebbe possibile, anche se non banale, individuare i ritratti immaginari, definiti come quei dipinti che ritraggono personaggi che non erano più in vita al momento in cui è nato l'autore del dipinto, o non sono mai stati contemporaneamente in vita.

4. IL SEMANTIC WEB

4.1 L'architettura

La comunità scientifica sta investendo molte energie nel Semantic Web⁴ (BERNERS-LEE 1997, 1998, 1999), la cui architettura a livelli è ripor-

⁴ <http://www.semanticweb.org/>.

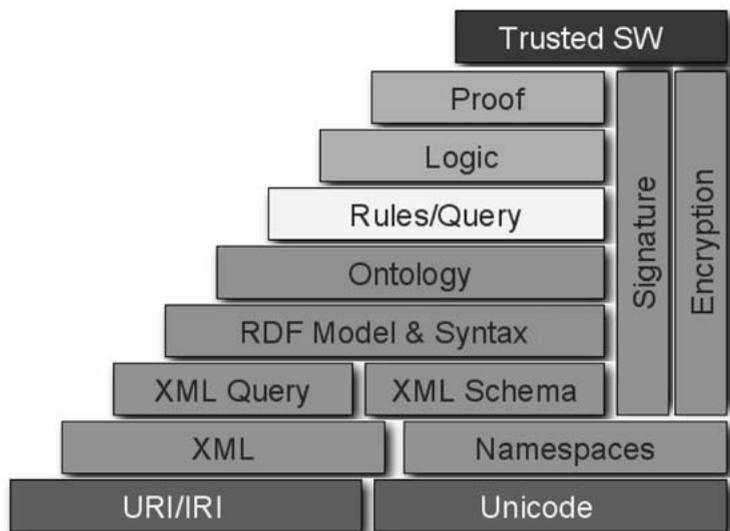


Fig. 1 – L'architettura a strati del Semantic Web.

tata in Fig. 1. Nel contesto del Semantic Web, il termine semantico assume sostanzialmente la valenza di “elaborabile dalla macchina”. La semantica dei dati consiste nelle informazioni utili perché la macchina possa utilizzarli nel modo corretto.

A tale proposito, va ricordato che la filosofia di base del web è quella di uno spazio informativo universale, navigabile, con un mapping da URI⁵ alle risorse. Il Semantic Web (BERNERS-LEE, HENDLER, LASSILA 2001) potrà funzionare solo se le macchine potranno accedere ad un “insieme strutturato di informazioni” e a un “insieme di regole di inferenza” da utilizzare per il ragionamento automatico. La sfida del Semantic Web, quindi, è fornire un linguaggio per esprimere “dati” e “regole” per ragionare sui dati, che consenta l’“esportazione” sul web delle regole da qualunque sistema di rappresentazione della conoscenza.

Diamo qui una breve descrizione del ruolo dei vari componenti:

– XML (eXtensible Markup Language), che con Namespace e XML Schema consente di dare ai documenti una “struttura” arbitraria, gioca un ruolo di base.

⁵ URI (Uniform Resource Identifier) è il generico insieme di tutti i nomi/indirizzi che costituiscono le brevi sequenze di caratteri che fanno riferimento ad una risorsa; URL (Uniform Resource Locator) è un termine informale, non più utilizzato nelle specifiche tecniche, associato con gli schemi URI più noti e diffusi (http, ftp, mailto, etc.). Per ulteriori dettagli cfr. <http://www.w3.org/Addressing/>.

- RDF (Resource Description Framework), che può essere usato per esprimere il “significato”, asserendo che alcuni particolari elementi hanno delle proprietà (per es. *autore-di*).
- L’ontologia (o “livello ontologico”), che è il contenitore che definisce in modo formale le relazioni fra i termini e che permette di descrivere le relazioni tra i tipi di elementi (per es. la transitività di alcune proprietà), senza però fornire informazioni su come utilizzare queste relazioni dal punto di vista computazionale. Il linguaggio definito dal W3C (World Wide Web Consortium)⁶ per definire ontologie strutturate, in architettura web, e per consentire una migliore integrazione dei dati tra applicazioni in settori diversi è OWL (Web Ontology Language).
- Il “livello logico”, che è il livello immediatamente superiore a quello ontologico. A questo livello le asserzioni esistenti sul web possono essere utilizzate per derivare nuova conoscenza. Dato che i sistemi deduttivi non sono normalmente interoperabili, si può pensare di definire un linguaggio universale per rappresentare le dimostrazioni. I sistemi potrebbero quindi autenticare con la firma digitale queste dimostrazioni ed esportarle ad altri sistemi che le potrebbero incorporare nel Semantic Web.
- La firma digitale (*digital signature*), che è importante in diversi strati nel modello astratto del Semantic Web, nel quale si ritiene necessaria una infrastruttura in cui le parti possano essere riconosciute e accettate come credibili in specifici domini. Con una granularità fine, la firma digitale potrebbe essere utilizzata per stabilire la provenienza delle ontologie e delle deduzioni, oltre che dei dati.

4.2 Alcune considerazioni

Nel settore dell’Intelligenza Artificiale già da molti anni esistono sistemi per la gestione della conoscenza che sono in grado di operare ragionamenti e deduzioni. Una caratteristica importante del Semantic Web è, coerentemente con i principi del web, la sua totale “decentralizzazione”, e l’“interoperabilità” tra applicazioni, macchine, ontologie diverse. In questa visione, l’opera di armonizzazione delle ontologie e degli schemi descrittivi viene affidata ad agenti software che, disponendo di una rappresentazione della conoscenza e di regole di deduzione espresse con un linguaggio interoperabile, operano per armonizzare conoscenze diverse.

Per l’interoperabilità semantica, un problema rilevante è l’armonizzazione degli schemi descrittivi e delle ontologie, due aspetti fortemente interconnessi. Infatti, va tenuto presente che l’organizzazione dei concetti (aspetto ontologico) non può essere considerato a sé stante, scorrelato dallo schema adottato per rappresentare le informazioni. In altri termini, la complessità derivante, per

⁶ <http://www.w3.org/>.

esempio, dalla specializzazione degli oggetti per funzione o tipo (es. arco rampante, vasetto per unguenti, colonna a tortiglione) deve necessariamente trovar posto o in un sistema di classificazione (thesaurus) o in una articolazione della scheda descrittiva (quindi campo, sottocampo, attributo). Ne consegue che l'interoperabilità semantica tra collezioni non può essere affrontata semplicemente trovando degli equivalenti terminologici, impresa peraltro già non semplice, ma deve tener conto di come una determinata organizzazione dei concetti può essere tradotta verso uno specifico schema di rappresentazione.

Un'architettura possibile, coerente con quella del Semantic Web, è presentata in *SIGNORE 2005*, in cui si ipotizza un meccanismo associativo guidato da un modello mentale dell'utente, e da una caratterizzazione semantica dei contenuti, entrambi basati su una ontologia formale.

4.3 L'interoperabilità

La decentralizzazione e l'interoperabilità sono due dei principi cardini del web. Accade spesso di vedere applicazioni web che di fatto non sono concettualmente dissimili dalle applicazioni tradizionali basate su un elaboratore centrale, usuali negli anni '70. Lo spirito del web è invece l'assenza di poli centrali che ne possano rendere critico il funzionamento. Ne consegue che difficilmente si può accettare un'architettura in cui una qualche autorità centrale sia investita del potere di definire termini, concetti, strutture dati, formati di scambio, mentre è essenziale l'interoperabilità tra applicazioni e istituzioni diverse.

Si ritiene spesso che la definizione di uno schema di rappresentazione dell'informazione, e una codifica che ne permetta la fruizione su qualunque piattaforma, siano la soluzione ottimale. Si tratta tuttavia di una impostazione influenzata dall'eredità del passato e dalla non completa assimilazione del principio di decentralizzazione. La presenza di differenti culture e il diverso modo di percepire i concetti impongono di considerare non solo l'interoperabilità tecnologica, ma anche quella "semantica". Nel seguito descriveremo brevemente questi due aspetti e le tecnologie che li supportano.

4.3.1 L'interoperabilità tecnologica

Le informazioni poste sul web sono leggibili da parte di chiunque e l'interoperabilità tecnologica viene assicurata da un uso adeguato delle tecnologie web, *in primis* quelle funzionali alla strutturazione dell'informazione: XML, Namespaces, XML Schema.

XML – Extensible Markup Language⁷ è nato per supportare la realizzazione delle nuove applicazioni web, in cui i dati costituiscono un elemento

⁷ <http://www.w3.org/XML/>; <http://www.w3.org/TR/2000/REC-xml-20001006>; <http://www.w3.org/TR/2003/REC-xml-events-20031014/>; *SIGNORE 2001*.

essenziale (*data-centric web applications*) ed è stato quindi il primo passo per assegnare una semantica ai tag e supportare le transazioni sul web, permettendo lo scambio di informazioni tra database diversi. XML, che consente di avere viste diverse degli stessi dati e di personalizzare le informazioni mediante opportuni agenti, è indipendente dalla piattaforma e dal linguaggio, e non a caso è stato definito “ASCII del 2000”. La sintassi XML è semplice, la sua elaborazione automatica è poco complessa e un documento XML resta comprensibile alla lettura diretta. XML ricopre un ruolo centrale nell’architettura del web e ogni nuovo linguaggio utilizzato per definire un nuovo standard deve essere descritto in XML. La struttura formale del documento viene espressa nella DTD (Document Type Definition), che può essere inclusa nel documento o referenziata come risorsa esterna. Un documento XML si dice “well formed” quando rispetta le regole di scrittura; viene detto “validato” quando è coerente con la struttura definita nella DTD.

XML Namespaces – La sempre crescente popolarità di XML determina un proliferare di applicazioni, e la conseguente possibilità che esse debbano essere utilizzate insieme. Ne consegue che è necessario poter distinguere tra i vari namespace. Un XML namespace⁸ è un “insieme di nomi, caratterizzati da un URI di riferimento, utilizzati come element type e attribute name”. Ogni nameset ha associato un prefisso che lo identifica, e quindi i tag sono individuati univocamente specificando il prefisso e il nome “locale”.

XML Schema Definition – La DTD viene espressa con una sua propria sintassi, e quindi richiede editor, parser e processor *ad hoc*. Inoltre, è difficile estenderla, non contempla datatype e deve supportare tutti gli elementi e attributi descritti dai namespace inclusi. Queste limitazioni sono superate da XMLSchema⁹, che ha le stesse funzionalità delle DTD, ma offre alcuni significativi vantaggi, in quanto è espresso con la sintassi XML e include datatype, inheritance, regole di combinazione degli schema. XMLSchema fornisce anche un miglior supporto dei namespace, offre la possibilità di agganciare documentazione e informazioni semantiche e permette di rappresentare vincoli sui possibili valori, tipi complessi e gerarchie di tipi.

4.3.2 L’interoperabilità semantica

I metadati – Nel navigare sul web si seguono dei link, che portano a quella che formalmente viene detta “risorsa” (resource) identificata univo-

⁸ <http://www.w3.org/TR/REC-xml-names/>.

⁹ <http://www.w3.org/TR/xmlschema-0/>; <http://www.w3.org/TR/xmlschema-1/>; <http://www.w3.org/TR/xmlschema-2/>.

camente da un URI. Le informazioni sulla risorsa¹⁰ vengono generalmente dette “metadati”. Si può quindi dire che “i metadati sono informazioni, comprensibili dalla macchina, relative a una risorsa web o a qualche altra cosa”. Si noti quindi che secondo questa definizione i metadati possono descrivere non solo risorse web, ma anche qualsiasi altra entità che possa essere identificata in maniera univoca con un meccanismo analogo a quello utilizzato per individuare le risorse web (quindi mediante un URI). Di conseguenza, i metadati possono descrivere, per esempio, un dipinto, un reperto archeologico, un sito, etc. Il punto chiave è il fatto che i metadati sono comprensibili dalla macchina (*machine understandable*), e quindi costituiscono un tipo di informazione che può essere utilizzata in maniera opportuna dai software agent.

Va sottolineato che “i metadati sono dati”, e quindi possono essere “memorizzati come dati”, e essere “descritti da altri metadati”, e così via. Il numero di metalivelli da specificare dipende dalle caratteristiche delle applicazioni e dalle specifiche esigenze.

Resource Description Framework – L’uso efficace dei metadati richiede che vengano stabilite delle convenzioni per la “semantica”, la “sintassi” e la “struttura”. Le “singole comunità” interessate alla descrizione delle loro risorse specifiche definiscono la “semantica” dei metadati pertinenti alle loro esigenze. La “sintassi”, cioè l’organizzazione sistematica dei data element per l’elaborazione automatica, facilita lo scambio e l’utilizzo dei metadati tra applicazioni diverse. La “struttura” può essere vista come un vincolo formale sulla sintassi, per una rappresentazione consistente della semantica.

Resource Description Framework (RDF)¹¹ è lo strumento base per la codifica, lo scambio e il riutilizzo di metadati strutturati, e consente l’interoperabilità tra applicazioni che si scambiano sul web informazioni machine-understandable. RDF permette di definire la semantica dei tag XML, e fornisce un modello per descrivere le risorse¹², che hanno delle proprietà (o anche attributi o caratteristiche). Il data model RDF, che consente di rappresentare statement RDF in modo sintatticamente neutro, è molto semplice ed è basato su tre tipi di oggetti: *Resources* (sempre individuate da un URI), *Properties* (un aspetto specifico, identificato da un “nome”, che assume un “valore”), *Statements* (una tripla composta da un “soggetto”, un “predicato” e un “oggetto”, ovvero una tripla composta da una “risorsa”, una “proprietà” e un “valore”).

¹⁰ Nel linguaggio corrente una risorsa viene anche detta “documento” o “oggetto”, per mettere in evidenza il fatto che sia leggibile da un essere umano, o da una macchina.

¹¹ <http://www.w3.org/RDF/>; <http://www.w3.org/TR/NOTE-rdf-simple-intro>; <http://www.w3.org/TR/1999/PR-rdf-schema-19990303/>; MILLER 1998; LASSILA, SWICK 1999.

¹² RDF definisce una risorsa come un qualsiasi oggetto che sia identificabile univocamente mediante un URI.

RDF supporta l'utilizzo di convenzioni che rendono più agevole l'interoperabilità tra insiemi separati di metadati, e consente di pubblicare vocabolari machine readable, ma anche leggibili da utenti umani, definiti dalle singole comunità disciplinari, favorendo enormemente il riuso e l'estensione della semantica tra comunità diverse.

Graficamente, le relazioni tra *Resource*, *Property* e *Value* vengono rappresentate mediante grafi etichettati orientati, in cui le risorse vengono identificate come nodi (graficamente delle ellissi), le proprietà come archi orientati etichettati, e i valori corrispondenti a sequenze di caratteri come rettangoli. Un insieme di proprietà che fanno riferimento alla stessa risorsa viene detto descrizione (*description*).

RDF identifica univocamente le proprietà mediante il meccanismo dei namespace XML (XMLns)¹³, che forniscono un metodo per identificare in maniera non ambigua la semantica e le convenzioni che regolano l'utilizzo delle proprietà, identificando l'authority che gestisce il vocabolario. Uno degli esempi più noti è la Dublin Core Metadata Initiative (DC)¹⁴. Si può utilizzare un namespace XML per identificare in maniera non ambigua lo schema per il vocabolario Dublin Core puntando alla risorsa Dublin Core che ne definisce la semantica. La descrizione di una risorsa può utilizzare le proprietà definite nel namespace Dublin Core, o alcune di esse, eventualmente aggiungendo altre proprietà che rispondano ad esigenze specifiche, semplicemente puntando al namespace che ne definisce la semantica.

È importante sottolineare che siamo così in presenza di una architettura decentralizzata e *peer-to-peer*, in cui le varie comunità disciplinari hanno a priori pari dignità, e possono conquistarsi la loro autorevolezza.

Web Ontology Language – RDF e RDF Schema sono due componenti importanti, ma non risolvono tutti i problemi. Infatti applicazioni sofisticate richiedono di poter ragionare sui dati. Il Semantic Web deve quindi essere supportato da ontologie, che definiscono i concetti e le relazioni utilizzate per descrivere e rappresentare uno specifico dominio di conoscenza. In altri termini, occorre disporre di un linguaggio che consenta di definire la terminologia usata, le caratteristiche logiche e i vincoli delle proprietà, l'equivalenza dei termini, le cardinalità delle associazioni, etc. Un'ulteriore complessità deriva dal fatto che il web è intrinsecamente distribuito, e di conseguenza applicazioni diverse possono usare ontologie diverse, o le stesse ontologie, ma espresse in lingue diverse. Il W3C, sfruttando anche i risultati di altri progetti, quali DAML e OIL, ha definito un linguaggio,

¹³ Namespaces in XML – World Wide Web Consortium Recommendation, 14 January 1999 (<http://www.w3.org/TR/REC-xml-names/>).

¹⁴ <http://dublincore.org/>.

denominato OWL¹⁵, che permette di esportare le ontologie in modo interoperabile. OWL offre tre sottolinguaggi, di crescente potere espressivo: *OWL Lite* (per rappresentare classificazioni gerarchiche e vincoli semplici), *OW DL* (per una maggiore potenza espressiva, garantendo comunque che tutte le conclusioni siano computabili e concluse in un tempo finito), *OWL Full* (che offre la massima potenza espressiva, ma non fornisce garanzie sui tempi di computazione, e difficilmente sarà supportato nella sua interezza da software che implementano il ragionamento). Ognuno di questi linguaggi è un'estensione del precedente, sia in termini di ciò che può essere espresso che in termini della validità delle conclusioni.

O.S.

5. SYMONTOX E OPAL

Il Master in Economia, Gestione e Marketing dei Turismi e dei Beni Culturali, erogato dalla Facoltà di Scienze Politiche dell'Università Luiss "Guido Carli", ha fornito l'occasione per portare avanti uno studio di grande respiro sull'applicazione di metodologie di KM all'analisi di materiali archeologici. Infatti è stato siglato un accordo di collaborazione fra la Luiss e due istituti del CNR: l'Istituto di Studi sulle Civiltà Italiane e del Mediterraneo Antico (ISCIMA) e l'Istituto per l'Analisi dei Sistemi ed Informatica (IASI). Nell'ambito di questo accordo si è svolto il periodo di stage della Dott.ssa Tatiana Bommarà, volto all'analisi ontologica di una classe di materiali di produzione etrusca, che in precedenza è stato oggetto di indagini statistiche: le urne volterrane del Museo Guarnacci. Questo lavoro ha prodotto come risultato finale, oltre alla stessa analisi, una tesi di Master dal titolo: "Ricerche ontologiche e semantiche applicate ai Beni Culturali: le urne volterrane del Museo Guarnacci".

Dato il successo della prima esperienza, è stato deciso di ripeterla ogni anno avvalendosi del contributo di stagisti provenienti dal Master Luiss, in modo da portare avanti il lavoro completando l'analisi delle urne volterrane, per poi passare ad altre classi di materiali.

Lo strumento messo a disposizione dallo IASI/CNR consiste in un software prototipale per la gestione di ontologie di dominio: SymOntoX (*Symbolic Ontology XML-based management system*). Un'ontologia può essere definita come «an explicit, agreed and shared definition of a portion of reality by means of a conceptual model. This model may exist in someone's head or be embedded in a software or information system, in an object or in a process. The task of an ontology builder is to identify the model

¹⁵ Web Ontology Language (<http://www.w3.org/2004/OWL/>).

and make it explicit. This allows the model to be accessed by, or communicated to, a wider range of potential users, be they people, organisations or software agents» (MISSIKOFF 2003). Rispetto ad un thesaurus, orientato alla descrizione di termini, un'ontologia descrive concetti e le relazioni che li interconnettono, mentre una Base di Conoscenza (Knowledge Base o KB) oltre ai concetti include anche oggetti del mondo reale (USHOLD, GRUNINGEN 1996).

SymOntoX permette la gestione di ontologie in qualunque dominio applicativo, essendo basato sul metamodello OPAL (*Object, Process and Actor Language*) per la rappresentazione della conoscenza (MISSIKOFF, TAGLINO 2002). Secondo il metamodello OPAL (Fig. 2), un'ontologia viene costruita definendo una serie di concetti e stabilendo fra loro delle relazioni semantiche. La definizione dei concetti principali di un dominio (modellazione concettuale) avviene mediante l'utilizzo di una serie di categorie di concetti e di relazioni semantiche predefinite (idee di modellazione). Ai concetti caratteristici del dominio considerato saranno infine associati degli oggetti del mondo reale. Il livello destinato alla specifica delle idee di modellazione viene detto "meta-livello", il livello che contiene i concetti principali di un dominio viene invece definito "livello intensionale", mentre il livello in cui vengono raccolti gli oggetti del mondo reale che corrispondono ai concetti principali di dominio è chiamato "livello estensionale".

I due meccanismi principali utilizzati per classificare i materiali sono rappresentati dalla "specificità strutturale" e dalla "organizzazione gerarchica". La specificità strutturale di un tipo consiste nel fornirne la lista delle proprietà e, per ognuna di esse, la tipizzazione (ed eventualmente il range) dei valori che possono assumere. Ad esempio:

$$\langle \text{type-name} \rangle : = (\langle \text{prop-name} \rangle : \langle \text{value-type} \rangle \\ \langle \text{prop-name} \rangle : \langle \text{value-type} \rangle ; \langle \text{value-range} \rangle)$$

Per tipizzazione dei valori si intende il definire se un attributo può avere come valore, ad esempio, un numero intero (*integer*), una stringa di caratteri (*string*), ecc., mentre il *range* è la gamma di valori ammessi per un certo attributo.

L'organizzazione gerarchica viene invece definita utilizzando la relazione ISA, che consiste in una gerarchia di generalizzazione/specializzazione fra tipi. Ad esempio possiamo dichiarare che "brocca ISA vaso": intuitivamente questa dichiarazione stabilisce che tutte le caratteristiche del tipo "vaso" sono presenti anche nel tipo "brocca". In aggiunta, il tipo "brocca" può avere delle caratteristiche che non sono necessariamente presenti in tutti i vasi (per es. l'ansa). Questo principio viene anche definito "ereditarietà" in quanto il tipo brocca "eredita" tutte le caratteristiche del tipo vaso. A livello estensionale la gerarchia ISA si trasforma in una relazione di inclusione fra classi. Nell'esempio

The Three Levels of Knowledge

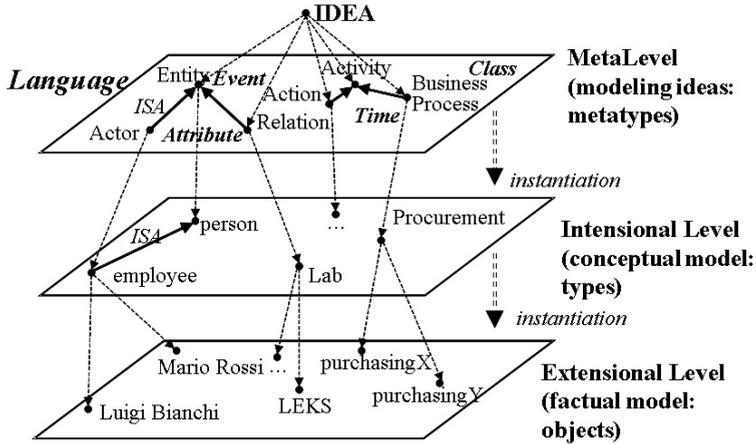


Fig. 2 – Il metamodello OPAL.

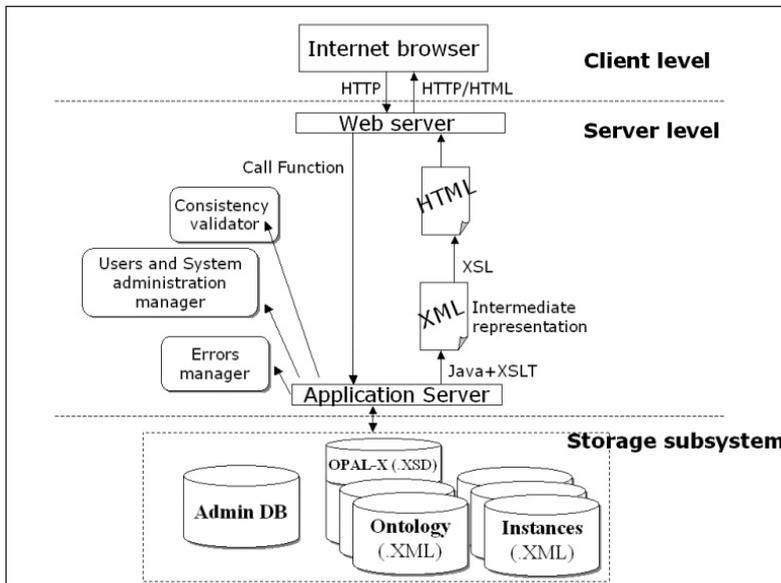


Fig. 3 – L'architettura di SymOntoX.

fornito, l'insieme degli oggetti definibili come "vaso" contiene l'insieme degli oggetti definibili come "brocca".

La definizione di un sottotipo (ad es. da "vaso" a "brocca") si realizza utilizzando il meccanismo della specializzazione che può essere di due tipi, per specificazione e per restrizione:

- Specializzazione per specificazione: la specializzazione avviene mediante l'aggiunta di nuove proprietà a quelle già definite nel tipo più generale.
- Specializzazione per restrizione: questo meccanismo permette di rifinire una o più proprietà del tipo più generale operando sui due strumenti di tipizzazione: l'elencazione dei valori ammessi in caso di variabili discrete e/o la gamma di valori ammessi in caso di variabili continue.

SymOntoX (Fig. 3) permette la gestione di più ontologie di dominio e supporta varie tipologie di utenti con diversi diritti di accesso, supporta il multilinguismo e, essendo basato su tecnologia XML, può essere utilizzato via Internet. Il software è sviluppato secondo un'architettura a tre livelli corrispondenti ad un'interfaccia utente, un motore di gestione delle ontologie, e un database.

6. STATO DELL'ARTE ED ESPERIENZE PRECEDENTI IN AMBITO ARCHEOLOGICO

Il panorama scientifico attuale offre un'abbondanza di fonti per quanto concerne aspetti teorici legati alla natura, creazione e gestione di ontologie. Al contrario è riscontrabile una preoccupante carenza di casi studio reali (e non semplici prototipi esemplificativi) nel campo dei beni culturali in generale, e dell'archeologia in particolare. Il valore del presente contributo risiede proprio in questo aspetto: infatti, oltre all'analisi delle urne volterrane (cfr. *infra* § 7), OPAL e SymOntoX sono già stati utilizzati in precedenza per supportare la classificazione di un'altra classe di materiali. A scopo esemplificativo, per facilitare la comprensione della metodologia seguita e dello strumento utilizzato sarà infatti illustrata sinteticamente, nei principali passi operativi seguiti, l'analisi tassonomica svolta in una precedente applicazione.

Il caso studio è rappresentato dalla modellazione concettuale della classe delle fibule (e di alcune sue specializzazioni), partendo dall'analisi di materiali provenienti da scavi e ricognizioni della necropoli villanoviana dei Quattro Fontanili a Veio. Un interesse particolare è dato dal fatto che questi materiali sono stati in precedenza oggetto di altre analisi (CLOSE BROOKS 1965; TOMS 1986; GUIDI 1993), offrendo così la possibilità di utilissimi confronti. La classificazione delle fibule è stata costruita sulla base del lavoro di J. SUNDWALL del 1943 e viene sviluppata secondo una gerarchia di attributi principali che genera, di conseguenza, una gerarchia di tipi progressivamente più specializ-

zati. La gerarchia degli attributi principali è la seguente: 1) forma dell'arco, 2) forma della staffa, 3) decorazione.

Seguendo questa logica, per il nostro esempio sono stati utilizzati sette tipi; questi sono, in ordine gerarchico:

- 1) Fibula;
- 2) Fib_arc_ing (fibula ad arco ingrossato);
- 3) Fib_arc_sang (fibula ad arco a sanguisuga);
- 4) Fib_ing_simm (fibula ad arco ingrossato e staffa simmetrica);
- 5) Fib_ing_asimm (fibula ad arco ingrossato e staffa asimmetrica);
- 6) Fib_sang_simm (fibula ad arco a sanguisuga e staffa simmetrica);
- 7) Fib_sang_asimm (fibula ad arco a sanguisuga e staffa asimmetrica).

Secondo i Dizionari Terminologici dell'Istituto Centrale per il Catalogo e la Documentazione (ICCD) il termine fibula indica un «oggetto di ornamento utilizzato per fissare capi di abbigliamento. È costituito da un arco, un ago generalmente collegato all'arco per mezzo di una molla, ed una staffa in cui è inserita la punta dell'ago» (BARTOLONI *et al.* 1980). Già a partire da questo livello, l'utilizzo di SymOntoX e della metodologia OPAL ha prodotto una serie di osservazioni che hanno portato ad una parziale ridefinizione dell'organizzazione gerarchica della classe delle fibule. Infatti se possiamo essere sostanzialmente d'accordo nell'individuare un primo livello la cui valenza distintiva è basata sulla funzione, quella appunto di fissare capi di abbigliamento, nella definizione del secondo livello emerge un primo problema. Nell'organizzazione gerarchica canonica, si ha infatti un secondo livello formato da 1) fibule ad arco serpeggiante, 2) fibule ad arco rivestito e 3) fibule ad arco. Da una prima osservazione intuitiva, si può riscontrare che nei casi 1 e 2 il concetto di arco viene rifinito dagli attributi, rispettivamente “serpeggiante” e “rivestito”, mentre nel caso 3 il concetto di arco non viene rifinito affatto.

In un contesto logico-matematico, ciò assegna automaticamente l'entità “fibula ad arco” ad un livello più generalizzato. Inoltre, da un punto di vista morfologico, definire il corpo di una fibula serpeggiante come “arco” risulta alquanto fuorviante: infatti la definizione di arco che si può trovare in dizionari ed enciclopedie oscilla da “linea curva parte di una circonferenza” a “struttura curva con la quale si supera uno spazio”. Il concetto di curvatura appare dunque costantemente associato al termine “arco”, mentre è altresì ben lontano dalla forma del corpo di una fibula serpeggiante, dove è semmai l'ago ad avere una forma “ad arco”. Tale osservazione viene inoltre rafforzata dal fatto che nei depositi archeologici le fibule serpeggianti mostrano una valenza sociale ben definita essendo rinvenute esclusivamente in tombe maschili rispetto a quelle ad arco che sono invece caratteristiche

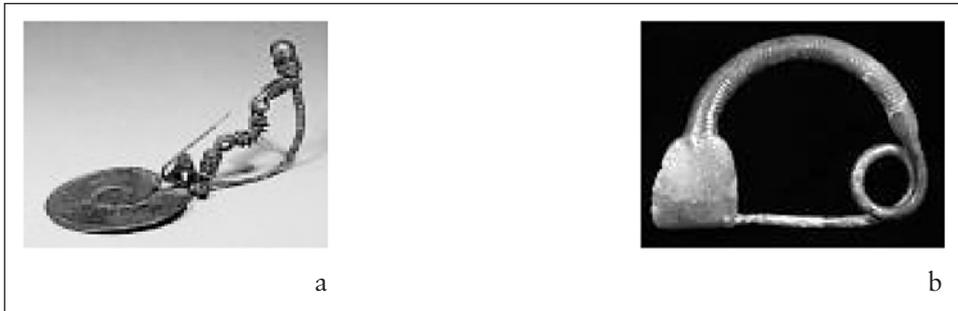


Fig. 4 – a) fibula serpeggiante; b) fibula ad arco.

di sepolture femminili (GUIDI 1993). Alla luce di quanto detto finora viene proposta un'organizzazione del secondo livello, definito “forma generica” in cui troviamo i concetti di “fibula serpeggiante” (Fig. 4, a) e “fibula ad arco” (Fig. 4, b), che a sua volta viene rifinita, nel terzo livello, in “fibula ad arco rivestito” e “fibula ad arco decorato”.

Nel terzo livello viene dunque definita la “tecnica decorativa” delle fibule ad arco; qui l'elemento distintivo è rappresentato dalla decorazione che viene scelta per ornare il corpo della fibula. In un caso (fibule ad arco rivestito) vengono utilizzati dei dischi che vengono infilati nell'arco e che possono essere di vari materiali: ambra, bronzo e osso sono i più comuni; nell'altro (fibule ad arco decorato) la decorazione viene invece realizzata attraverso l'incisione, o altre tecniche, di vari motivi decorativi.

A questo punto è emerso un nuovo problema derivante dall'inquadramento delle fibule cosiddette “inornate”; tali oggetti sono infatti sprovvisti di decorazione per una scelta consapevole e non per scarsità di mezzi dato il livello di manifattura generalmente piuttosto raffinato. Ancora una volta un'accurata modellazione concettuale ha fornito la soluzione per rappresentare in modo formale e rigoroso anche questo aspetto: nella definizione delle tecniche decorative, dopo aver indicato i valori espliciti (rivestimento, incisione o altro), è stato fornito anche il numero di scelte possibili per questo parametro, e cioè da 0 a 1; questo indica che al massimo è possibile selezionare una sola scelta, ma anche che è possibile non associare alcun valore a questa proprietà, definendo quindi in questo modo le “fibule inornate”.

Applicando dunque la metodologia OPAL, e dopo le opportune verifiche di consistenza, il risultato dell'analisi tassonomica assume la struttura mostrata in Fig. 5. Mentre l'ontologia, dopo le opportune verifiche sintattiche e semantiche, assume la seguente forma:

```
fibula:=  
  (ago:(dritto,curvo,ad_u),  
   forma_generica:(serpeggiante,ad_arco),  
   staffa:(simmetrica,asimmetrica,allungata,a_disco))  
&  
fibula_ad_arco:= ISA fibula  
  (ago:(dritto),  
   forma_generica:(ad_arco),  
   forma_specifica:(ad_arco_ingrossato,a_sanguisuga,  
    a_losanga,a_navicella),  
   tecnica_decorativa:{(rivestimento,incisione)}0..1)  
&  
fibula_ad_arco_decorato:= ISA fibula_ad_arco  
  (staffa:(simmetrica,asimmetrica,allungata),  
   tecnica_decorativa:(incisione),  
   sintassi_decorativa:(continua,a_fasce,a_banda_centrale,  
    incisione_profonda),  
   motivi_decorativi:{(fasce_lisce,fasce_lin_or,fasce_lin_ver,...)})
```

La Fig. 6 mostra invece la schermata di SymOntoX dedicata alla definizione della gerarchia di specializzazione e in cui è possibile rilevare l'implementazione dei costrutti utilizzati nell'analisi dei materiali.

O.M.

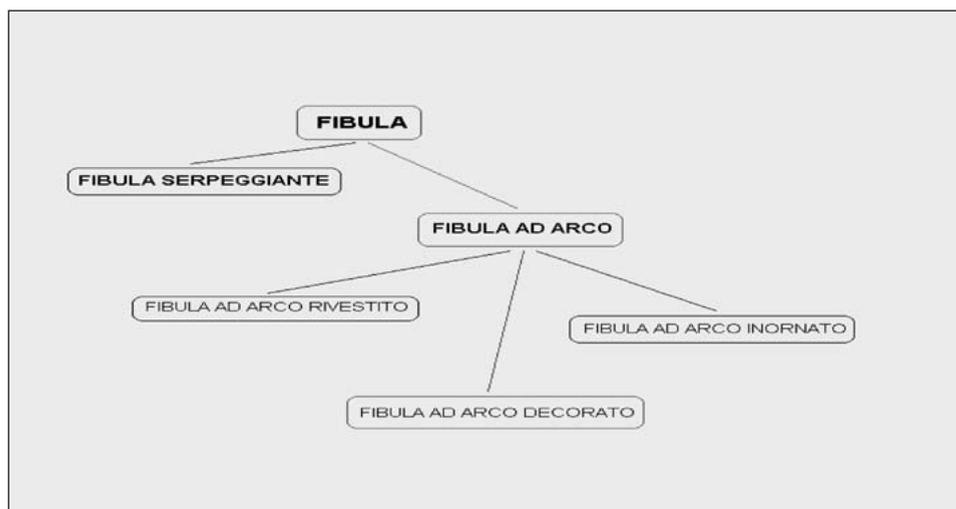


Fig. 5 – Risultato dell'analisi tassonomica.



Fig. 6 – Schermata di SymOntoX dedicata alla definizione della gerarchia di specializzazione.

7. UN'ONTOLOGIA PER LA CLASSIFICAZIONE DELLE URNE ETRUSCHE

Il progetto di ricerca avviato nel 2004 dall'ISCIMA, in collaborazione con l'Università Luiss “Guido Carli” e con lo IASI del CNR, è inteso a sviluppare alcuni aspetti metodologici e applicativi connessi con la classificazione dei materiali archeologici, con la formalizzazione e la trasmissione della conoscenza e con l'attività di formazione. Si tratta di aspetti a cui il settore informatico dell'Istituto ha rivolto una particolare attenzione fin dalla sua nascita, avvenuta più di un ventennio fa grazie a una felice intuizione di Mauro Cristofani.

La classificazione automatica dei dati archeologici è stata affrontata nel corso del tempo attraverso approcci differenziati, strettamente connessi anche con l'adozione di diverse modalità di codifica delle informazioni (ORLANDI 1993); in ogni caso, nella classificazione è stata riconosciuta un'operazione intellettuale, cioè un processo di esplicazione concettuale, basilare per la formazione di concetti scientifici (HEMPEL 1965, 139). Si è così partiti dall'applicazione di tecniche statistiche a dati codificati sotto forma di matrice numerica del tipo unità di rilevazione/variabile, per passare a procedure di analisi esplorativa di dati testuali e per giungere infine a una rappresentazione formalizzata di dati descrittivi semi-strutturati, tramite l'uso di linguaggi di marcatura. Quest'ultimo approccio ben si ricollega alle problematiche di

diffusione del sapere scientifico, attraverso l'uso di standard di codifica e le potenzialità dell'ICT.

Per quanto attiene all'attività di formazione svolta presso il nostro Istituto, essa costituisce il punto di raccordo tra una lunga militanza di oltre quindici anni in diverse sedi universitarie per l'insegnamento dell'informatica applicata all'archeologia o più in generale ai beni culturali e la possibilità di offrire un tipo di didattica orientata a un'attività di tutorato più specifico e personalizzato. Anche perché, nel corso degli anni, il mondo universitario è rimasto sordo al riconoscimento stabile di materie di insegnamento, per le quali è evidente la necessità di ricorrere a ricercatori formati nel settore degli studi umanistici (cfr. in particolare ORLANDI, MORDENTI 2003).

7.1 *La scelta del caso di studio*

L'esperienza di collaborazione scientifica tra istituzioni diverse è nata sotto l'egida unificatrice dell'informatica, con un indirizzo applicativo rivolto al tema: "Ricerche semantiche e ontologiche applicate ai beni culturali: la classificazione dei materiali archeologici".

La scelta, come caso di studio, delle urne funerarie lapidee prodotte in epoca ellenistica a Volterra e nel territorio circostante è legata a una delle prime applicazioni informatiche promosse dall'ISCIMA nell'ambito di una linea di ricerca dal titolo "Automatizzazione di corpora etruschi", nata agli inizi degli anni Ottanta e intesa alla sperimentazione di nuove metodologie d'indagine per lo studio di classi di materiali fra loro omogenei, afferenti alla cultura etrusca. La classificazione dei materiali archeologici, quali gli specchi bronzei (MOSCATI 1984, 1986) e le urne funerarie lapidee (cfr. in sintesi MOSCATI 1997a), è stata inizialmente realizzata attraverso un approccio metodologico di tipo quantitativo, mediante l'utilizzazione di tecniche di analisi statistica.

Nel caso del Progetto Volaterrae, il campione selezionato e sottoposto ad analisi raccoglieva le urne conservate nel Museo Guarnacci, nel Museo Archeologico di Firenze e in altri musei e collezioni italiani e stranieri. Lo studio ha preso avvio dall'analisi di un complesso di circa 1200 casse, descritte in base a variabili quantitative e qualitative: oltre a informazioni estrinseche (luogo e stato di conservazione, luogo di rinvenimento e, laddove noto, contesto funerario di pertinenza), sono stati privilegiati gli aspetti tipologici e morfologici, più idonei per natura ad un trattamento di tipo quantitativo, insieme alle informazioni concernenti le scene figurate rappresentate sulle casse, al fine di verificare l'esistenza di un rapporto fra morfologia e iconografia (MOSCATI 1995, 1997b).

I risultati raggiunti nel corso del tempo possono essere distinti in due diversi livelli di indagine: il primo, basato sull'applicazione di tecniche di analisi descrittiva, ha consentito di ottenere dati quantitativi, intesi a spie-

gare il comportamento singolo o reciproco delle variabili. Il secondo livello, basato sull'applicazione di tecniche di analisi multivariata, ha dato vita a un processo epistemologico, che ha permesso di dedurre nuove informazioni rispetto a quelle di partenza e di individuare al loro interno una struttura formale, mettendo in moto, attraverso le successive azioni di convalida, processi interpretativi dinamici. In ogni caso, l'analisi statistica ha offerto la base per giudizi quantitativamente determinati, connessi a una formalizzazione delle operazioni di classificazione mediante l'uso di meccanismi di correlazione e di associazione su cui si basano le analisi statistiche.

7.2 L'approccio statistico alla classificazione dei dati archeologici

Già nel corso del Progetto Volaterrae era stata ravvisata, nell'ambito di una stimolante partecipazione al progetto "Tecniche di analisi dei dati nello studio dell'immagine iconografica", coordinato da Sergio Camiz e finanziato dal CNR, la necessità di ribaltare la procedura adottata, per evitare una definizione tipologica a priori di alcuni elementi caratterizzanti le casse delle urne e per verificare il peso discriminante delle singole varianti tipologiche individuate. Le cornici decorate a basso rilievo che delimitano superiormente le casse sono state pertanto descritte in base alle singole modanature che le compongono, lasciando come elemento fisso della descrizione la fascia con dentelli, che compare pressoché sempre al centro della sequenza.

Questo nuovo tipo di classificazione descrittiva, oltre ad offrire alcuni spunti interpretativi circa il procedimento utilizzato dagli scalpellini nella decorazione a basso rilievo delle cornici e degli zoccoli delle casse, ha dato avvio anche a un nuovo tipo di approccio, che vede il superamento del sistema di codifica numerica più tradizionale, basato sul criterio di presenza/assenza delle singole variabili, per giungere, attraverso un processo di formalizzazione del linguaggio, a una descrizione di tipo testuale degli aspetti iconografici, con particolare riferimento alla sequenza delle modanature, agli schemi compositivi delle scene rappresentate sulla cassa e infine alla sintassi relativa alla disposizione dei personaggi all'interno del campo figurato (MOSCATI 2004). Lo scopo è di individuare nel testo, tramite procedure di analisi esplorativa di dati testuali (LEBART, SALEM 1994), occorrenze simultanee di forme lessicali e di "segmenti ripetuti" (sequenze di forme che si ripetono nello stesso ordine) che indichino, ad esempio, la presenza di schemi simili, anche in rappresentazioni diverse, tratti da uno stesso repertorio di modelli ma poi combinati in modi diversi (cfr. in particolare CAMIZ, ROVA 1996; CAMIZ, ROVA, TULLI 1998 e da ultimo CAMIZ 2004).

Per quanto attiene alla classificazione delle cornici e degli zoccoli, essa è stata realizzata basandosi sulla lettura in ordine sequenziale, dall'alto verso il basso, delle singole modanature: un tipo di procedura già utilizzato per descrivere questi elementi strutturali nelle schede del *Corpus delle Urne Etrusche di*

età ellenistica (CUE: CRISTOFANI *et al.* 1975; CRISTOFANI 1977; CATENI 1984). Questo “riavvicinamento” al testo descrittivo coincide sostanzialmente con un nuovo indirizzo di ricerca del settore informatico dell’ISCIMA, inaugurato alla metà degli anni Novanta e indirizzato alla rappresentazione dei documenti archeologici tramite linguaggi di marcatura (SGML prima e XML poi) come formato di archiviazione e trasmissione delle informazioni in rete (MOSCATI, MARIOTTI, LIMATA 1999 e da ultimo MOSCATI c.s. a, b).

In genere, l’attività di ricerca in questo settore ha consentito di mettere a punto una metodologia di indagine informatizzata, mirata al recupero e alla sistematizzazione dei dati contenuti all’interno di documenti archeologici non strutturati o semi-strutturati di natura diversa, ma sempre connessi a indagini condotte sul campo (diari di scavo manoscritti, rapporti di scavo editi, documenti di archivio dell’Ottocento contenenti itinerari archeologici). Poiché l’adozione di linguaggi di marcatura si è dimostrata una scelta appropriata per la rappresentazione formalizzata, l’elaborazione automatica e il recupero delle informazioni contenute nei testi esaminati, nell’ambito dello stage formativo degli studenti del Master si è avviata anche la marcatura XML dei volumi editi del *corpus* delle urne volterrane, come base per successive analisi informatizzate (Fig. 7).

7.3 L’approccio ontologico alla classificazione dei dati archeologici

Il passaggio dalle problematiche connesse con la rappresentazione di documenti testuali attraverso l’uso di marcatori al dibattito sul Semantic Web e alla successiva definizione di ontologie nel dominio dell’archeologia è stato senza dubbio sequenziale. Infatti, se la marcatura individua porzioni di testo secondo uno schema gerarchico che ne descrive la struttura, attraverso un modello di tipo sintattico, i metodi e i linguaggi del Semantic Web tendono a descrivere la semantica dei documenti e le ontologie a evidenziare il reticolo di interrelazioni tra classi e proprietà delle classi. Questa distinzione, però, non è sempre così rigida, in quanto l’aspetto strutturale di un testo comporta anche valenze semantiche, che le soluzioni proposte dal Semantic Web aiutano a rendere esplicite, tramite regole di interpretazione processabili (PARODI, FERRARA 2002).

Nel caso delle urne, è stato studiato un modello formale di organizzazione della conoscenza (BOMMARA 2004), che integra tecniche di rappresentazione e di analisi ontologica, intese a offrire un contributo alla classificazione tassonomica di questi oggetti e alla costruzione di un insieme di regole logiche che ne verifichino la coerenza. Questa esperienza ha consentito di approfondire ulteriormente le problematiche teoriche connesse con la rappresentazione della conoscenza archeologica e della realtà storica che si è chiamati a descrivere e interpretare; infatti «the nature of representations of the world necessarily shapes our archaeological constructs, as it shapes those of any science» (GARDIN, PEEBLES 1992, 386).

```

<URNA>
  <INFO>
    <NUMERO_Catalogo>26</NUMERO_Catalogo>.
    <OGGETTO>Urna</OGGETTO> in <MATERIALE>tuofo</MATERIALE>;
    <LUOGO_CONSERVAZIONE>Volterra, Museo Guarnacci</LUOGO_CONSERVAZIONE>;
    <NUMERO_Inventario> inv. 538</NUMERO_Inventario>.
    <DIMENSIONI>
      <COPERCHIO>Cop.:
        <ALT>0,36</ALT>;<LUNG>0,57</LUNG>x<LARG>0,23</LARG>.
      </COPERCHIO>
      <CASSA>Cassa:
        <ALT>0,45</ALT>;<LUNG>0,57</LUNG>x<LARG>0,23</LARG>.
      </CASSA>
    </DIMENSIONI>
  </INFO>
  <DESCRIZIONE>
    <COPERCHIO classe="antropomorfo" tipo="femminile">Sul coperchio recumbente femminile vestita di <abito>chitone
    manicato e mantello che copre le gambe, il dorso, la nuca, il braccio e la spalla s., formando una grossa piega sotto l'addome
    </abito>. <capo>Sulla testa <ornamenti>un diadema decorato nei bordi da una doppia fila di perline</ornamenti></capo>; <collo>al
    collo <ornamenti>un torques</ornamenti></collo>. <abito>Il chitone presenta sulla spalla s. <ornamenti>un fermaglio a tre dischi
    </ornamenti> ed è cinto in vita da <ornamenti>una cintura decorata da placche rettangolari</ornamenti></abito>. <capo>
    <acconciatura>I capelli, con scriminatura centrale, si dispongono attorno al viso in ciocche ondulate</acconciatura></capo>.
    <bracciodx>Il braccio d. è disteso lungo il corpo e <manodx>la mano, che poggia sul ginocchio d. piegato, regge <oggetto_mano>un
    flabello con la punta rivolta verso l'alto e lungo manico. Il flabello è decorato internamente da un motivo perlinato a forma di foglia e da
    piume incise</oggetto_mano></manodx></bracciodx>. <bracciosx>Il braccio s. poggia su <bracciosx><cuscino>due cuscini col
    bordo ricamato da linee ondulate incise e nappine agli angoli</cuscino>, <bracciosx><manosx>la mano con le dita disposte nel
    <gesto>gesto apotropaiico</gesto>, tiene <oggetto_mano>un oggetto non identificabile</oggetto_mano></manosx></bracciosx>.
    </COPERCHIO>
    .....
    .....
    .....
  </DESCRIZIONE>
  <CONSERVAZIONE>Figura sul cop. mutila della prima falange dell'indice della mano s.; manca l'attributo nella mano s. Il manico
  del flabello ha una lacuna nella parte mediana</CONSERVAZIONE>.
  <BIBLIOGRAFIA>Bibl.: B.-K. III, p. 225, tav. CLII, 1; Ducati, tav. 275, fig. 666; Giglioli, tav. CCCCVIII, 2; Gambetti, p. 21, n. 1, p.
  71; Nielsen, pp. 305-7, 309, 314; Fiumi 1976, fig. 38</BIBLIOGRAFIA>.
  <RINVENIMENTO>Acquisita dal Museo tra il 1875 e il 1877</RINVENIMENTO>.
</URNA>

```



Fig. 7 – Esempio di marcatura XML di una scheda del CUE (Volterra, Museo Guarnacci, inv. n. 538).

Il tema centrale della sperimentazione è costituito dalla modellazione concettuale della classe “urna” attraverso la costruzione di un’ontologia che descriva il significato dei termini e rappresenti le relazioni che tra essi intercorrono, con lo scopo di aumentare le capacità espressive della rappresentazione stessa. Quest’ultima non segue più uno schema rigidamente arborescente di tipo gerarchico, ma prende in considerazione una complessa rete di relazioni, a cui è associata una rappresentazione degli oggetti di carattere strutturale e funzionale.

La prima fase del progetto è stata dedicata all’“estrazione dei concetti”, cioè alla definizione terminologica dei concetti della realtà da descrivere e rappresentare. Quindi, avendo ben presente la distinzione semantica tra “scomposizione” (definizione delle parti che compongono un oggetto) e “specializzazione” (cfr. *supra* § 5), si è passati all’“estrazione delle relazioni”, cioè alla definizione delle relazioni, gerarchiche e non, esistenti tra classi, e

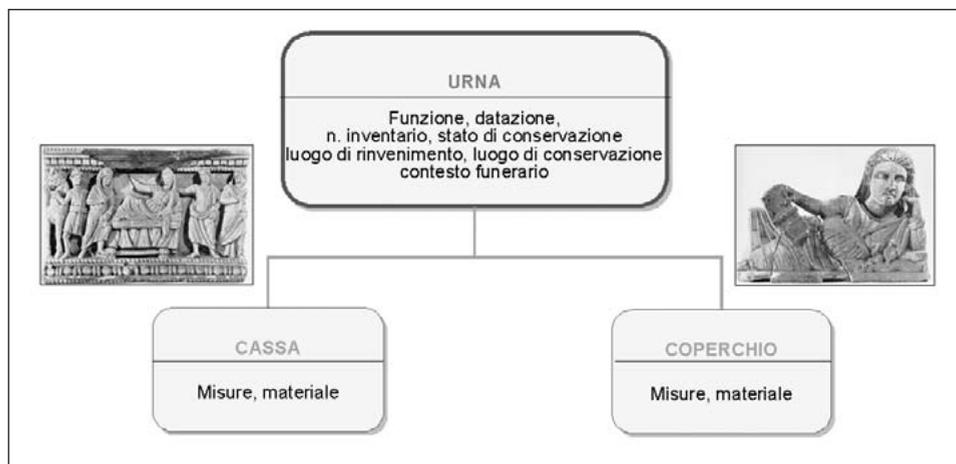


Fig. 8 – Scomposizione dell’oggetto urna in “cassa” e “coperchio”.

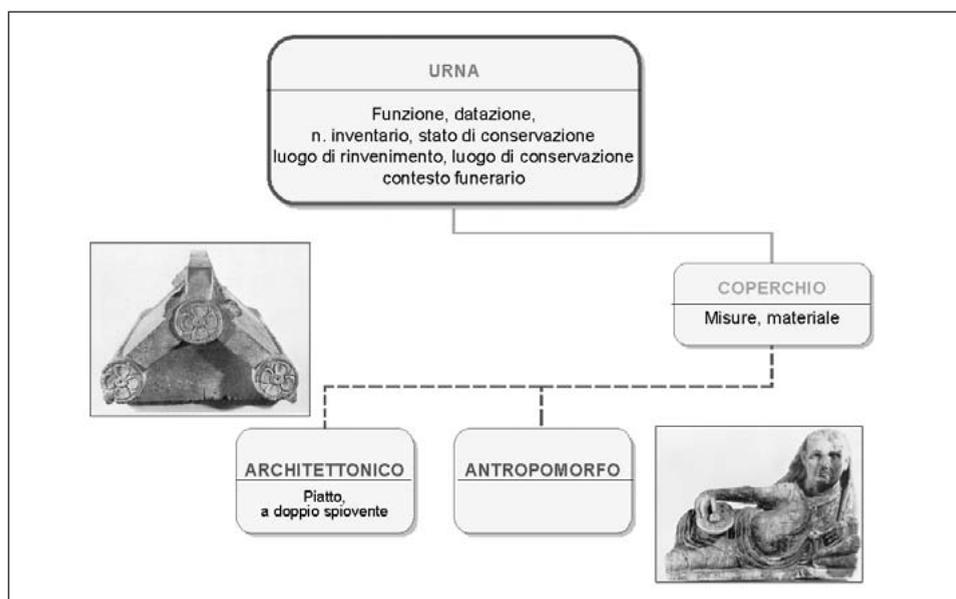


Fig. 9 – Specializzazione del coperchio dell’urna in “architettonico” e “antropomorfo”.

relativi attributi, e infine alla costruzione dell’ontologia vera e propria e alla sua gestione mediante l’uso del software prototipale SymOntoX e della metodologia OPAL.

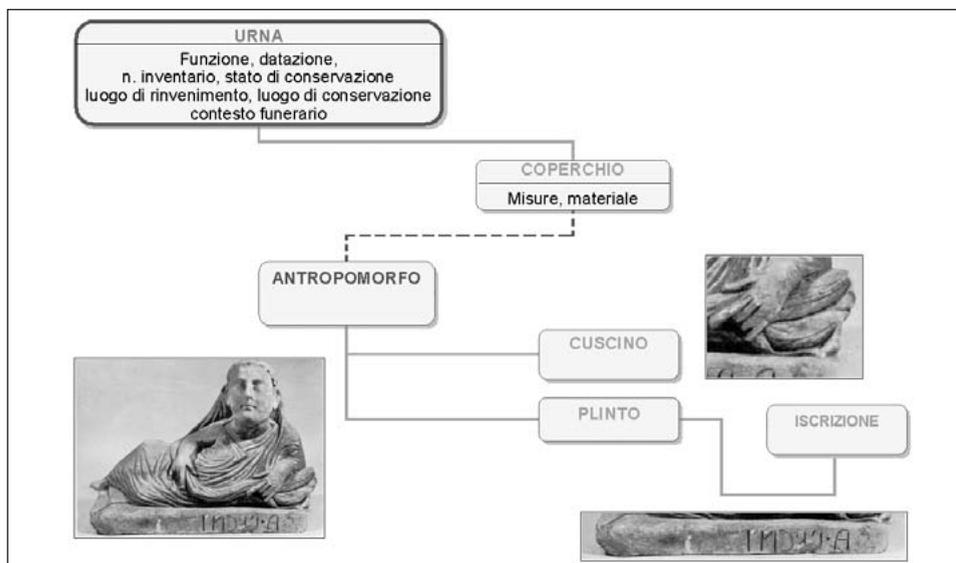


Fig. 10 – Rappresentazione di alcune delle parti in cui è stato scomposto il coperchio antropomorfo.

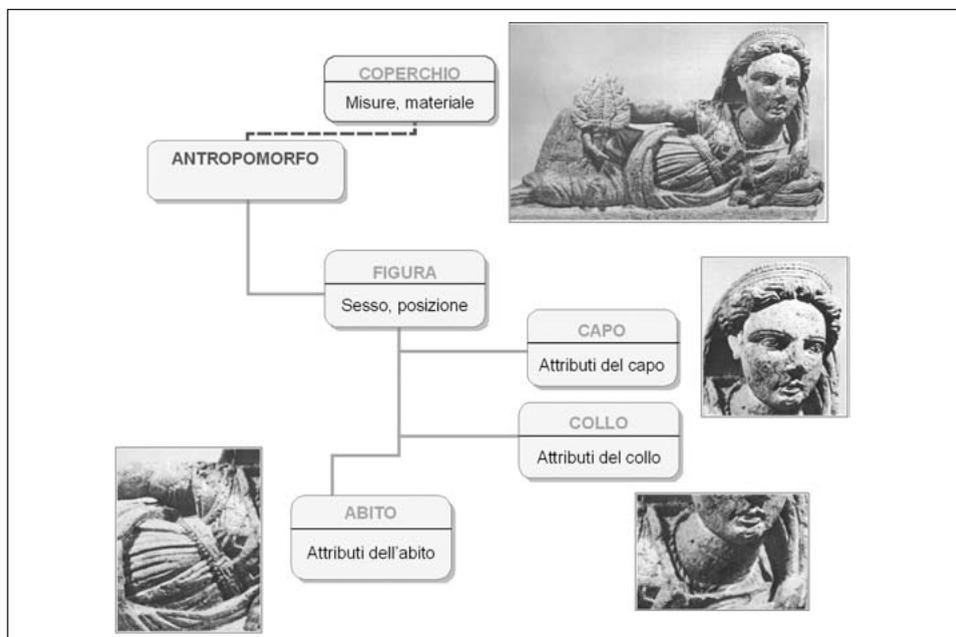


Fig. 11 – Attributi di alcune delle parti in cui è stata scomposta la figura del defunto.

Partendo dalle caratteristiche strutturali e considerando una relazione di tipo *part of* (scomposizione), l'oggetto urna è stato suddiviso in "cassa" e "coperchio", e proprio su quest'ultimo si è concentrata l'attenzione nello sviluppo dell'ontologia (Fig. 8). "Coperchio" si specializza in "architettonico" e "antropomorfo" (Fig. 9); a sua volta "antropomorfo", secondo una relazione di tipo *part of*, è stato suddiviso da un punto di vista strutturale in "cuscino", "plinto" (che può presentare un'iscrizione) e "figura" del defunto (Fig. 10). L'analisi della figura, di cui si indicano il sesso e la posizione, ha portato a un'ulteriore scomposizione in "capo", "collo", "abito", "braccio", "mano" e a una specializzazione di alcune di queste parti, in base a una serie di attributi che ne arricchisce la descrizione e la conseguente classificazione (Fig. 11).

In conclusione, il progetto avviato costituisce, da un punto di vista archeologico, uno strumento di approfondimento della questione legata alla distinzione tra una classificazione estensionale e intensionale, che risiede nella scelta di procedure basate sul raggruppamento di oggetti o sull'articolazione di concetti. L'uso di ontologie arricchisce e specializza senz'altro le operazioni di classificazione e propone forme originali di fruizione dei beni studiati, che possono essere organizzati in classi e sottoclassi – che sono una specializzazione delle classi e ne ereditano gli attributi – collegate le une alle altre attraverso relazioni tra parti descrittive e dimensioni interpretative.

Da un punto di vista informatico, il progetto si orienta verso l'interoperabilità tecnologica e semantica, che va assumendo sempre maggiore importanza nell'evoluzione del web verso il Semantic Web ed in cui la famiglia di tecnologie XML svolge un ruolo essenziale nei vari livelli architetturali. Infatti, nel processo di valorizzazione e fruizione dei beni culturali, le potenzialità dell'Information Technology a supporto delle operazioni conoscitive sono legate alle possibilità di proporre modalità inedite e più complesse di descrizione e rappresentazione del manufatto che consentano di muoversi su livelli differenziati, che dai dati giungono, attraverso i metadati, alla definizione di schemi ontologici, relativi a un dominio ben definito, entro cui si muove la ricerca applicata.

P.M.

ORESTE SIGNORE

Istituto di Scienza e Tecnologia dell'Informazione
CNR – Pisa

OLEG MISSIKOFF

Centro di Ricerca sui Sistemi Informativi
Università Luiss "Guido Carli"

PAOLA MOSCATI

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

BIBLIOGRAFIA

- BERNERS-LEE T. 1997, *Metadata Architecture* (<http://www.w3.org/DesignIssues/Metadata.html>).
- BERNERS-LEE T. 1998, *Semantic Web Road Map* (<http://www.w3.org/DesignIssues/Semantic.html>).
- BERNERS-LEE T. 1999, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*, San Francisco, Harper.
- BERNERS-LEE T., HENDLER J., LASSILA O. 2001, *The Semantic Web*, «Scientific American» (<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>).
- BOMMARA T. 2004, *Ricerche ontologiche e semantiche applicate ai Beni Culturali: le urne volterrane del Museo Guarnacci* (tesi di Master Universitario in Economia, gestione e marketing dei turismi e dei beni culturali, Facoltà di Scienze Politiche, Università Luiss “Guido Carli”, a.a. 2003-2004).
- CAMIZ S. 2004, *On the coding of archaeological finds*, in P. MOSCATI (ed.), *New Frontiers of Archaeological Research. Languages, Communication, Information Technology*, «Archeologia e Calcolatori», 15, 201-218.
- CAMIZ S., ROVA E. 1996, *Metodi di analisi per lo studio di un gruppo di sigilli cilindrici vicinorientali e di altre immagini strutturate*, in P. MOSCATI (ed.), *III Convegno Internazionale di Archeologia e Informatica (Roma 1995)*, «Archeologia e Calcolatori», 7, 647-659.
- CAMIZ S., ROVA E., TULLI V. 1998, *Exploratory analysis of images engraved on ancient Near-Eastern seals based on a distance among strings*, «Statistica», 58, 4, 669-689.
- CATENI G. (ed.) 1986, *Corpus delle urne etrusche di età ellenistica. Urne volterrane 2. Il Museo Guarnacci*, 2, Pisa, Pacini.
- CLOSE BROOKS J. 1965, *Proposta per una suddivisione in fasi della necropoli veiente di Quattro Fontanili*, «Notizie degli Scavi», 53-62.
- CRISTOFANI M. et al. (eds.) 1975, *Corpus delle urne etrusche di età ellenistica 1. Urne volterrane 1. I complessi tombali*, Firenze, Centro Di.
- CRISTOFANI M. (ed.) 1977, *Corpus delle urne etrusche di età ellenistica 2. Urne volterrane 2. Il Museo Guarnacci*, 1, Firenze, Centro Di.
- DIGICULT 2003, *DigiCULT. Towards a Semantic Web for Heritage Resources*. Thematic Issue 3, May (http://www.digicult.info/downloads/ti3_high.pdf).
- DOERR M., HUNTER J., LAGOZE C. 2003, *Towards a Core Ontology for Information Integration*, «Journal of Digital Information», 4, 1 (<http://jodi.ecs.soton.ac.uk/Articles/v04/i01/Doerr/>).
- GARDIN J.-C., PEEBLES C.S. (eds.) 1992, *Representations in Archaeology*, Bloomington, Indiana University Press.
- GUIDI A. 1993, *La necropoli veiente dei Quattro Fontanili nel quadro della fase recente della prima età del ferro italiana*, Firenze, Leo S. Olschki.
- HEMPEL C.G. 1965, *Aspects of Scientific Explanation*, New York, The Free Press.
- LASSILA O., SWICK R. 1999, *Resource Description Framework (RDF) Model and Syntax Specification*, W3C Recommendation 22 February 1999 (<http://www.w3.org/TR/REC-rdf-syntax/>).
- LEBART L., SALEM A. 1994, *Statistique Textuelle*, Paris, Dunod.
- MILLER E. 1998, *An Introduction to the Resource Description Framework*, «D-Lib Magazine», May (<http://www.dlib.org/dlib/may98/miller/05miller.html>).
- MISSIKOFF O. 2003, *Ontologies as a reference framework for the management of knowledge in the archaeological domain*, in *Proceedings of the CAA2003 Conference Enter the Past (Vienna 2003)*, BAR International Series 1227, Oxford, Archaeopress.
- MISSIKOFF M., TAGLINO F. 2002, *Business and Enterprise Ontology Management with SymOntox*, New York, Springer-Verlag.

- MOSCATI P. 1984, *Ricerche matematico-statistiche sugli specchi etruschi*, Contributi del Centro Linceo Interdisciplinare “Beniamino Segre”, 66, Roma, Accademia Nazionale dei Lincei.
- MOSCATI P. 1986, *Analisi statistiche multivariate sugli specchi etruschi*, Contributi del Centro Linceo Interdisciplinare “Beniamino Segre”, 74, Roma, Accademia Nazionale dei Lincei.
- MOSCATI P. 1995, *Méthodes quantitatives et problèmes iconographiques*, in *Aplicaciones Informáticas en Arqueología: Teorías y sistemas (Saint-Germain-en-Laye 1991)*, 1, Bilbao, 317-329.
- MOSCATI P. 1997a, *Ricerche informatiche sulle urne volterrane*, in AA.VV., *Atti del XIX Convegno di Studi Etruschi e Italici (Volterra 1995)*, Firenze, Leo S. Olschki, 339-345.
- MOSCATI P. 1997b, *Un gruppo di urne volterrane con rappresentazione del “viaggio agli inferi in carpentum”*, in AA.VV., *Etrusca et Italica. Scritti in ricordo di M. Pallottino*, Pisa-Roma, 403-423.
- MOSCATI P. 2004, *Per la descrizione computerizzata delle urne volterrane. Problemi di formalizzazione*, in M. FANO SANTI (ed.), *Studi di archeologia in onore di Gustavo Traversari*, Roma, L’Erma di Bretschneider, 647-655.
- MOSCATI P. c.s. a, *Linguaggi di marcatura per la conservazione e la valorizzazione dell’informazione archeologica*, in AA.VV., *Digital Repositories for Cultural Heritage (Roma, Accademia Nazionale dei Lincei, 2003)*, in corso di stampa.
- MOSCATI P. c.s. b, *Esperienze di codifica elettronica di documenti archeologici*, in XML per i Beni Culturali. *Esperienze e prospettive per il trattamento di dati strutturati e semistrutturati (Pisa, Scuola Normale Superiore, 2004)*, in corso di stampa.
- MOSCATI P., MARIOTTI S., LIMATA B. 1999, *Il “Progetto Caere”: un esempio di informatizzazione dei diari di scavo*, «Archeologia e Calcolatori», 10, 165-188.
- NONAKA I., TAKEUCHI H. 1995, *The Knowledge-Creating Company*, Oxford, Oxford University Press.
- PARODI M., FERRARA A. 2002, XML, *Semantic Web e rappresentazione della conoscenza*, «Mondo digitale», 3, 42-51.
- ORLANDI T. (ed.) 1993, *Discipline umanistiche e informatica 1. Il problema dell’integrazione. Atti del Seminario (Roma 1991)*, Contributi del Centro Linceo Interdisciplinare “Beniamino Segre”, 87, Roma, Accademia Nazionale dei Lincei.
- ORLANDI T. (ed.) 1997, *Discipline umanistiche e informatica 2. Il problema della formalizzazione. Atti dei Seminari (Roma 1994)*, Contributi del Centro Linceo Interdisciplinare “Beniamino Segre”, 96, Roma, Accademia Nazionale dei Lincei.
- ORLANDI T., MORDENTI R. 2003, *Lo status accademico dell’Informatica umanistica*, «Archeologia e Calcolatori», 14, 7-32.
- SENGE P. 1990, *The Fifth Discipline: The Art and Practice of the Learning Organization*, New York, Doubleday.
- SIGNORE O. 2001, *Il ruolo centrale di XML nell’evoluzione del Web*, in XML Day Milan Conference Proceedings (Milan 2001) (<http://www.w3c.it/papers/>).
- SIGNORE O. 2003, *Strutturare la conoscenza: XML, RDF, Semantic Web*, in 1st Clinical Knowledge 2003 (Udine 2003) (<http://www.w3c.it/papers/ck2003.pdf>; <http://www.w3c.it/talks/ck2003/>).
- SIGNORE O. 2004, *Representing Knowledge in Semantic Cultural Web*, in EVA 2004 Jerusalem Conference on the Digitisation of Cultural Heritage (Jerusalem 2004) (<http://www.w3c.it/talks/eva2004Jerusalem/>).
- SIGNORE O. 2005, *Ontology Driven Access to Museum Information*, in CIDOC2005 Documentation & Users, Proceedings of the CIDOC Annual Conference (Zagreb 2005) (<http://www.w3c.it/papers/cidoc2005.pdf>; slides: <http://www.w3c.it/talks/2005/cidoc2005/>).
- SUNDWALL J. 1943, *Die Alteren Italischen Fibeln*, Berlin.
- TOMS J. 1986, *The relative chronology of the Villanovan Cemetery of Quattro Fontanili at Veii*, Napoli, AION, 7.

USCHOLD M., GRUNINGER M. 1996, *Ontologies: Principles, Methods and Applications*, «The Knowledge Engineering Review», 11, 2, 93-155.

WORKSHOPSW 2004 = *I Workshop "Rappresentazione della conoscenza nel Semantic Web culturale"*, Roma, Progetto Minerva (<http://www.w3c.it/events/minerva20040706/>).

ABSTRACT

Despite the fact that an increasing number of researchers in the cultural heritage sector is recognising the advantages that could derive from the use of knowledge management methodologies and tools, a lack of awareness of the basic principles of this discipline is still rather evident. Key concepts like "knowledge representation", metadata, conceptual modeling, syntactic or semantic interoperability, ontologies, can prove difficult to understand (and even more difficult to apply) for researchers with a background in the humanities.

This contribution, therefore, aims at clarifying the theoretical reference framework through the concrete analysis of archaeological materials. In fact, while it seems easier to borrow definitions and theoretical concepts or to artificially create even very complex conceptual models (e.g. the CIDOC CRM, which has recently been recognised as an ISO standard), it is a lot harder to implement such principles onto real world objects analysis.

According to this assessment, and to the need of going from theoretical to practical aspects, the paper is structured in three parts: the first offers a theoretical base that makes available, even for non-experts, the tools for addressing more operational aspects; the second describes, through practical examples, both the knowledge representation model and the software tool used for analysing a class of materials, the Etruscan urns, as shown in the third part. The final objective is, therefore, to provide a point of reference for facilitating the approach towards KM (Knowledge Management) and help clarifying the key elements of a discipline that is obtaining a growing success but, so far, still showing a high level of entropy.