

Da ISO 2788 ai nuovi standard per la costruzione e l'interoperabilità dei vocabolari controllati: un'analisi comparativa

di Tiziana Calvitti ed Elisabetta Viti

1. Gli standard BS 8723:2005-2008 e ANSI/NISO Z39.19-2005

L'articolo ha lo scopo di mettere a confronto la norma internazionale ISO 2788:1986¹ e i successivi standard BS 8723:2005-2008² e ANSI/NISO Z39.19-2005³, focalizzando l'attenzione sugli elementi di novità che i due standard anglosassoni apportano alla tematica dei vocabolari controllati e alla loro interoperabilità.

TIZIANA CALVITTI, Biblioteca comunale di Terni, piazza della Repubblica, 1, 05100 Terni, e-mail tiziana.calvitti@comune.terni.it.

ELISABETTA VITI, Centro servizi biblioteca, Facoltà di lettere e filosofia, Università degli studi di Siena, via Fieravecchia 19, 53100 Siena, e-mail viti2@unisi.it; dottoranda di ricerca in Scienze bibliografiche, archivistiche, documentarie e per la conservazione e il restauro dei beni librari e archivistici, presso l'Università degli studi di Udine.

Le autrici esprimono un sincero ringraziamento alla Dr.ssa Anna Lucarelli (Biblioteca nazionale centrale di Firenze, resp. Progetto Nuovo soggetto), che ha avuto la pazienza di leggere questo contributo e di arricchirlo con preziosi suggerimenti.

Ultima consultazione siti web: 28 agosto 2009.

1 International Organization for Standardization, *ISO 2788:1986 Documentation. Guidelines for the establishment and development of monolingual thesauri*, [Geneva]: ISO, 1986. Per la traduzione italiana si veda: *ISO 2788:1993 Documentazione. Linee guida per la costruzione e lo sviluppo di thesauri monolingue*, Milano: Ente nazionale italiano di unificazione, 1993.

2 British Standards Institution, *BS 8723-1:2005 Structured vocabularies for information retrieval. Guide. Definitions, symbols and abbreviations*, London: BSI, 2005; British Standards Institution, *BS 8723-2:2005 Structured vocabularies for information retrieval. Guide. Thesauri*, London: BSI, 2005; British Standards Institution, *BS 8723-3:2007 Structured vocabularies for information retrieval. Guide. Vocabularies other than thesauri*, London: BSI, 2007; British Standards Institution, *BS 8723-4:2007 Structured vocabularies for information retrieval. Guide. Interoperability between vocabularies*, London: BSI, 2007; British Standards Institution, *DD 8723-5:2008 Structured vocabularies for information retrieval. Guide. Exchange formats and protocols for interoperability*, London: BSI, 2008.

3 American National Standard Institute – National Information Standard Organization, *ANSI/NISO Z39.19-2005. Guidelines for the construction, format and management of monolingual controlled vocabularies* <<http://www.niso.org/kst/reports/standards/>>.

L'elaborazione di BS 8723:2005-2008 e di ANSI/NISO Z39.19-2005 è testimonianza di un costante e rinnovato interesse per l'indicizzazione, ed è focalizzata sulla creazione di un linguaggio artificiale, impiegabile all'interno di una comunità sia di specialisti che di lettori. I nuovi standard propongono delle linee guida generali, indipendenti da specifici contesti culturali e documentari, finalizzate al controllo del lessico e delle relazioni semantiche esistenti tra i termini, e dimostrano come, anche alla luce delle odierne possibilità di ricerca offerte dal web, i vocabolari controllati siano ancora strumenti ottimali, costituendo l'infrastruttura, la base e l'ossatura per regolamentare e formalizzare da un lato l'attività di indicizzazione e dall'altro le ricerche dell'utente, rendendole più amichevoli, efficaci e con risultati pertinenti⁴.

A distanza di quasi venti anni dalla traduzione italiana dello standard ISO 2788:1986, grazie alle esperienze maturate in campo nazionale e internazionale, l'ottica di questi standard si è sviluppata in relazione ai profondi cambiamenti subiti dal contesto documentario, informativo ed informatico.

Pur rimanendo ancora validi i criteri base di analisi e costruzione degli strumenti thesaurali, si sono ampliate le tipologie di vocabolari controllati e cambiamenti rilevanti hanno coinvolto altri aspetti. Dall'universo documentario analogico-cartaceo si è passati a quello digitale-ipertestuale e lo scopo principale dei vocabolari controllati è diventato quello di impiegare un linguaggio codificato e formalizzato che permetta di individuare l'informazione (manifestazione) e di recuperare l'*item*, in qualsiasi formato e su qualsiasi supporto.

Nell'anno 2000 in Inghilterra, un gruppo di "appassionati" di thesauri, alcuni già membri del British Standards Institution (BSI) Committee IDT/2/2/11 (fra i quali Leonard Will, Alan Gilchrist, Ron Davies e Stella Dextre Clarke), ha deciso di avviare un progetto internazionale per la revisione dei precedenti standard sulla costruzione di thesauri e l'analisi concettuale dei documenti ISO 2788:1986 e ISO 5964:1985⁵. Poiché questi si erano sviluppati dal BS 5723:1987⁶ e dal BS 6723:1985⁷, gli esperti stabilirono di ripercorrere un analogo cammino rivedendo, per prima cosa, il British standard, così da offrire alla comunità scientifica un prodotto finito, eventualmente adottabile anche in campo internazionale. Ecco che tra il 2005 e il 2008 sono state edite separatamente le cinque sezioni in cui si suddivide lo standard BS 8723:2005-2008.

Le prime due parti dello standard inglese⁸ sono una revisione del precedente standard BS 5723:1987 (ormai ritirato) e corrispondono ampiamente ad ISO 2788:1986 che rimangono tuttora correntemente uno standard internazionale.

⁴ In proposito, si ricorda la relazione di Mary Dykstra Lynch tenuta al Seminario internazionale di studi *Verso un nuovo Soggettario?*, presso la Biblioteca nazionale centrale di Firenze, 5-6 aprile 2001 e si segnala il volume di Vanda Broughton, *Costruire thesauri: strumenti per indicizzazione e metadati semantici*, Milano: Editrice Bibliografica, 2008, p. 11-69.

⁵ International Organization for Standardization, *ISO 2788:1986 Documentation* cit.; International Organization for Standardization, *ISO 5964:1985 Documentation. Guidelines for the establishment and development of multilingual thesauri*, Geneva: ISO, 1985.

⁶ British Standards Institution, *BS 5723:1987 Guide to establishment and development of monolingual thesauri*, London: BSI, 1987.

⁷ British Standards Institution, *BS 6723:1985 Guide to establishment and development of multilingual thesauri*, London: BSI, 1985.

⁸ British Standards Institution, *BS 8723-1:2005 Structured vocabularies for information retrieval. Guide. Definitions, symbols and abbreviations* cit.; British Standards Institution, *BS 8723-2:2005 Structured vocabularies for information retrieval. Guide. Thesauri* cit.

Le successive tre parti sono completamente nuove, descrivono altri tipi di strutture di vocabolari controllati e forniscono raccomandazioni per favorire l'interoperabilità tra vocabolari a vari livelli di immagazzinamento e recupero delle informazioni⁹. La parte terza illustra tipi di vocabolari controllati, liste di intestazioni di soggetto, tassonomie, ontologie, liste di autorità dei nomi, schemi di classificazione, includendo la problematica del loro impiego anche per applicazioni come schemi di categorizzazione automatica, ausili per la ricerca a testo libero, ecc. In riferimento a queste nuove tipologie di vocabolari, lo standard non prevede vere e proprie istruzioni per l'organizzazione e strutturazione dei contenuti, così come fa invece per i thesauri, ma rinvia ad altri standard¹⁰. La quarta e la quinta parte analizzano la capacità, semantica e tecnica, di un sistema di operare in armonia con altri, focalizzandosi su quei contesti nei quali vengono impiegati più di una lingua o più di un vocabolario, che possono essere diversi per struttura e dominio disciplinare, ma nei quali l'accesso alle risorse viene espletato a partire da un unico vocabolario controllato scelto dall'utente.

Quasi contemporaneamente (nel 1999) negli Stati Uniti, la National Information Standard Organization (NISO), l'American Psychological Association (APA), l'American Society of Indexers (ASI) e l'Association of Library Collection and Technical Services (ALCTS) hanno organizzato un workshop per sondare la possibilità dello sviluppo di uno standard per i thesauri in formato elettronico. Nel luglio del 2005 viene pubblicata la quarta edizione dello standard ANSI/NISO Z39.19-2005, *Guidelines for the construction, format and management of monolingual controlled vocabularies*. Rispetto alla prima edizione¹¹, l'ottica di studio si è ampliata sempre più, non limitandosi all'analisi dei criteri di costruzione e di gestione del thesaurus, inteso come strumento impiegabile nell'indicizzazione della carta stampata e utilizzabile esclusivamente in sistemi postcoordinati. Lo standard NISO prende in considerazione gli effetti delle nuove tecnologie nella creazione di vocabolari controllati, l'impiego dei thesauri in contesti di precoordinazione, la possibilità di utilizzarli nell'indicizzazione di risorse eterogenee (quali brevetti, rapporti tecnici, musica e video) e, soprattutto, la loro utilità nel recupero dell'informazione¹².

La pubblicazione, a differenza dello standard inglese, non è suddivisa in parti edite separatamente nel tempo, ma costituisce un corpo unico diviso in capitoli e il *set* delle raccomandazioni, basato su tecniche e procedure preferite, acquisisce un carattere più normativo che sancisce l'obbligatorietà, l'auspicabilità o la possibilità di determinate

9 British Standards Institution, *BS 8723-3:2007 Structured vocabularies for information retrieval. Guide. Vocabularies other than thesauri* cit.; British Standards Institution, *BS 8723-4:2007 Structured vocabularies for information retrieval. Guide. Interoperability between vocabularies* cit.; British Standards Institution, *DD 8723-5:2008 Structured vocabularies for information retrieval. Guide. Exchange formats and protocols for interoperability* cit.

10 Per esempio: Simple Knowledge Organization System (SKOS) <<http://www.w3.org/TR/skos-reference/>>; ISO/IEC 13250-3:2007. *Information technology, Topic Maps, Part 3: XML syntax*, Geneva: ISO, 2007; Resource Description Framework RDF/XML Syntax Specification (Revised) <<http://www.w3.org/TR/rdf-syntax-grammar/>>.

11 American National Standards Institute, *Guidelines for thesaurus structure, construction and use*, New York: ANSI, 1974 (ANSI/NISO Z39.19-1974).

12 Emanuela Casson, *Dai thesauri ai vocabolari controllati: alcune novità introdotte nell'ultima edizione dello standard ANSI/NISO Z39.19-2005*, «AIDAinformazioni», 24 (2006), n. 1-2, p. 69-70.

scelte (sono frequenti espressioni come *must/must not; should/should not; may*)¹³. In entrambi gli standard anglosassoni, le raccomandazioni si riferiscono ai vocabolari controllati monolingui, con particolare riferimento alla lingua inglese¹⁴. Nella sezione sull'interoperabilità vengono considerati alcuni problemi legati ai vocabolari multilingui, come per esempio la visualizzazione di sequenze parallele che mostrano equivalenze concettuali in più di una lingua. Però, per un approfondimento sui vocabolari multilingui, gli standard rimandano espressamente a quanto previsto da ISO 5964:1985¹⁵.

Rispetto alle precedenti linee guida, anche il linguaggio usato dai redattori si arricchisce ed entrano a far parte dei nuovi standard termini legati al mondo dell'informazione e del *world wide web*. In particolare, nel glossario delle linee guida americane, non vengono riportate esclusivamente le definizioni relative agli elementi strutturali e funzionali dei thesauri, ma viene anche chiarito il significato dei vari modelli di vocabolari controllati e dei vari elementi legati all'organizzazione, alla trasmissione, alla ricerca, al ricupero delle conoscenze ed alla navigazione web, per esempio: *information storage and retrieval system, browsing*, ecc. Inoltre, si passa dalla nozione di documento, definito in ISO 2788 come "qualsiasi unità, stampata o no, catalogabile e indicizzabile", a quella di *content object*¹⁶, un'entità che contiene dati/informazioni, che a loro volta possono essere "contenitori" di dati/informazioni. Diversa, invece, l'accezione dello standard BS 8723, in cui il significato ampio di documento investe il mondo digitale e quello del web.

In entrambi gli standard anglosassoni viene analizzato l'impatto che i vocabolari controllati hanno sull'*information retrieval*. In particolare, in quello americano, è rivolta un'attenzione precisa nei confronti degli utenti, con insistenti considerazioni sull'usabilità di strumenti normalizzati, e si descrivono le varie tipologie di fruitori che potrebbero, eventualmente, interagire con le interfacce dei vocabolari (occasionalisti, esperti, professionisti del settore, ecc.). Per ognuno di loro si forniscono soluzioni *ad hoc*, che permettano e facilitino un completo ed efficace uso del sistema.

2. Le novità rispetto alle ISO 2788:1986

Le novità più significative che apportano questi standard, rispetto all'ISO 2788, riguardano l'ampliamento delle tipologie di vocabolari controllati, le forme di visualizzazione dei dati e l'interoperabilità semantica e tecnica tra thesauri diversi o tra thesauri e altre tipologie di vocabolari controllati.

2.1. Tipologie di vocabolari controllati

Come precedentemente accennato, lo standard americano NISO e l'inglese BS prendono in analisi un universo vasto di vocabolari controllati: il primo, destinando il cap. 5 alla descrizione degli scopi, della struttura e dei principi dei vocabolari; il secondo, riservan-

¹³ Emanuela Casson, *Dai thesauri ai vocabolari controllati: alcune novità introdotte nell'ultima edizione dello standard ANSI/NISO Z39.19-2005*, cit., p. 72.

¹⁴ Infatti le abbreviazioni e gli acronimi, che definiscono la relazione o la funzione del termine, sono espressi esclusivamente in lingua inglese, a differenza delle ISO 2788, che, avendo un carattere internazionale, si riferiscono anche alla lingua francese e tedesca.

¹⁵ International Organization for Standardization. *ISO 5964:1985 Documentation*. Cit.

¹⁶ Emanuela Casson, *Dai thesauri ai vocabolari controllati: alcune novità introdotte nell'ultima edizione dello standard ANSI/NISO Z39.19-2005*, cit., p. 72.

do l'intera parte 3, *Vocabularies other than thesauri*, alla descrizione dettagliata delle varie tipologie di vocabolari che possono trovare un impiego nell'ambito dell'indicizzazione.

L'impianto espositivo del NISO procede alla descrizione delle caratteristiche e delle funzioni delle quattro tipologie di vocabolari analizzate: liste, *synonym rings*, tassonomie e thesauri. I vocabolari vengono presentati in ordine di complessità crescente, procedendo dalle semplici liste (dette anche *pick list*) ordinate alfabeticamente o con altro principio logico, ai *synonym rings*, utili nel contesto del recupero dell'informazione ma non impiegabili nell'indicizzazione, fino alle strutture più complesse: tassonomie e thesauri. Conclude il capitolo una breve descrizione dell'impiego di metadati e schemi di metadati nei vocabolari controllati.

Anche lo standard inglese introduce la trattazione delle varie tipologie di strumenti di controllo terminologico, corredandola con una serie di definizioni utili alla contestualizzazione dell'argomento. Tra gli strumenti analizzati troviamo gli schemi di classificazione, ove si affrontano vantaggi e svantaggi degli schemi enumerativi a fronte di quelli analitico-sintetici, l'impiego delle faccette, i criteri di selezione e ordinamento delle classi, la costruzione delle gerarchie, l'uso delle vedette e delle notazioni di classe. Un esempio particolare di vocabolario controllato è quello dei *business classification schemes*, «modello concettuale delle relazioni gerarchiche di funzioni, attività e transazioni»¹⁷ per l'organizzazione di record gestionali di enti, organizzazioni ed aziende dove i documenti trovano una sistematizzazione per classi-funzioni o per soggetti-funzioni che ne facilita la disposizione e la conservazione.

Si analizzano, a seguire, le tassonomie: di facile uso per gli utenti inesperti, divenute già molto popolari tra i professionisti dell'informazione e gli abituali utilizzatori della rete, sono meno prescrittive di thesauri e classificazioni. Sono definite come vocabolari strutturati secondo principi classificatori con caratteristiche thesaurali, progettate per la navigazione con media elettronici, particolarmente utili per la mappatura di altri schemi e usate proficuamente nei siti web e nei portali, con l'impiego di metadati. Delle intestazioni per soggetto si sottolineano, in particolare, le caratteristiche comuni con i thesauri e le classificazioni. I nuovi schemi in via di sviluppo, infatti, tendono ad accomunare le *subject headings* ad un thesaurus, mediante l'esplicitazione delle gerarchie e tramite collegamenti non solo fra termine e termine, ma anche fra termine e intestazioni di soggetto. Le cosiddette ontologie, originariamente sviluppate nell'ambito degli studi sull'interazione uomo-macchina e sull'intelligenza artificiale, sono riferite a specifici domini della conoscenza e contengono particolari relazioni plasmabili sulla comprensibilità dei legami tra i concetti. Le liste d'autorità di nomi comuni e propri sono organizzate sempre più in analogia con strutture thesaurali, con ammissione di relazioni gerarchiche (corrispondenti alla relazione gerarchica esemplificativa del thesaurus), di equivalenza e di varianti storiche (che permettono di censire usi e trasformazioni occorse nel tempo riguardo alla forma di un nome di organizzazione, luogo ecc.).

2.2. Presentazione e *displaying*

La trattazione delle tematiche legate al *displaying* dei vocabolari controllati segue, nei tre standard, una linea comune dettata dai medesimi oggetti in esame, seppur

¹⁷ Fiorella Foscarini, *L'approccio alla gestione dei documenti basato sulle funzioni. L'analisi delle funzioni come elemento centrale nella progettazione e gestione dei sistemi documentali*, Firenze, 17 giugno 2008 <<http://www.rinascimento-digitale.it/formazione/corso-conservazione-giugno2008/foscarini-3-approcciofunzionale.pdf>>.

con approfondimenti e riferimenti notevolmente differenti, conseguenza dei contesti cronologici di compilazione.

La norma ISO 2788 dedica un apposito capitolo alle modalità di presentazione dei termini e delle relazioni nei thesauri. La presentazione si riferisce ad un contesto di stampa tradizionale e si limita a tre forme basilari di visualizzazione: alfabetica, sistematica (corredata da apposito indice alfabetico, complemento obbligato per un corretto recupero dell'informazione in un contesto cartaceo) e grafica, anch'essa munita di apposita sezione alfabetica.

Il recente standard inglese BS 8723 dichiara esplicitamente il suo parallelismo con la norma internazionale ISO, trattando i vari tipi di *display* come sezioni interne, successive alla descrizione delle diverse forme di vocabolari¹⁸. Esordisce con la descrizione del *display* informatico, illustrando un'ampia varietà di visualizzazioni e mantenendo poi, nel corso della trattazione, un costante riferimento al contesto digitale.

Lo standard americano NISO, nella sua quarta edizione, definisce requisiti e raccomandazioni specifici per il *displaying* di vocabolari controllati, dedicando un capitolo all'argomento. L'attenzione si sofferma sull'illustrazione delle caratteristiche necessarie per una corretta ed efficace visualizzazione, come per esempio: la rappresentazione delle relazioni tra i termini e le scelte tipografiche adottate¹⁹; gli specifici tipi di visualizzazione adottabili; le possibili forme (cartacea, elettronica, ipertestuale) di distribuzione dello strumento; infine, la documentazione, allegabile al vocabolario, predisposta per facilitare la fruizione da parte degli utenti, anche mediante l'adozione di tecnologie che permettano la visualizzazione e la navigazione per accedere ai *content object*.

La norma ISO dopo aver introdotto le varie forme di presentazione, apre la trattazione con quella che è considerata la tipologia base di ogni interfaccia vocabolario-utente, ossia la presentazione alfabetica e illustra come variante interna della visualizzazione alfabetica anche il *display* gerarchico. Segue la presentazione sistematica che descrive l'impiego di indici ruotati, ossia un «indice alfabetico di stringhe di soggetto (o di termini composti da più parole), nelle quali ciascun descrittore (o ciascuna parola) è a turno evidenziato, pur rimanendo nella propria posizione entro la stringa»²⁰.

Lo standard inglese, nella sezione dedicata ai thesauri e alla descrizione delle singole tipologie di *display*, esordisce con la visualizzazione del singolo termine, ripercorrendo, nella descrizione della sequenza consigliata (relazioni tra termini e note), quella proposta dalla norma ISO. Nella presentazione alfabetica, i termini possono convenzionalmente mostrare uno o più livelli di sovraordinazione o subordinazione, prevedendo l'impiego della numerazione del livello di gerarchia (BT1, BT2, ecc.) o espedienti grafici quali l'indentazione. Nella sezione dedicata alla presentazione gerarchica, lo standard inglese analizza l'impiego dell'analisi a faccette.

Il NISO, nella parte dedicata ai vari tipi di presentazione, descrive, più dettagliatamente rispetto agli altri due standard, le varie opzioni di visualizzazione alfabeti-

18 La parte 2, *Thesauri*, dedica il paragrafo 10 alla presentazione e al *layout* dei suddetti thesauri; la parte terza, *Vocabularies other than thesauri*, ad ogni tipologia di vocabolario descritta - schemi di classificazione, tassonomie, ecc. - fa seguire l'esposizione di tipi di *display* consigliati.

19 Prestando attenzione anche al conflitto che può insorgere tra simboli convenzionali, impiegati per la lettura del vocabolario, e comandi-macchina, utilizzati dai *software* di visualizzazione.

20 Glossario di organizzazione della conoscenza <<http://www.iskoi.org/doc/glossario.htm>>.

ca, spaziando dalla semplice elencazione alfabetica, con impiego di re-indirizzamento automatico da errori o variazioni ortografiche, alla visualizzazione piatta, in cui un termine ha un solo primo livello di BT e NT (a quest'ultima, si consiglia di integrare un sistema notazionale che permetta all'utente di un sistema informatico di indicizzazione di espandere la ricerca e localizzare facilmente il termine di interesse); segue nella descrizione la lista permutata o ruotata visualizzabile in ambiente elettronico in modalità KWIC (*Keyword in context*), in cui la parola è evidenziata all'interno della stringa, pur rimanendo nella sua posizione, oppure KWOC (*Keyword out context*), quando la parola è evidenziata assumendo sempre la prima posizione nella stringa. Infine, viene descritta la modalità di presentazione dettagliata del singolo termine, che esplica il massimo delle sue potenzialità in contesto elettronico con la possibilità per l'utente di navigare interattivamente nel vocabolario, scegliendo quando visualizzare nella sua completezza un termine e le sue note, muovendosi tra le sue relazioni ed eventualmente spostandosi, previa configurazione del software, dal termine al *content object* al quale fa riferimento.

Lo standard NISO cede successivamente il passo all'illustrazione di tipologie differenti di *display* gerarchici: multilivello, struttura ad albero, struttura *top term*, gerarchica *two way*, per categoria sovraordinata e mediante impiego di faccette. Rispetto allo standard BS e alla norma ISO, lo standard americano descrive, in maniera dettagliata, la struttura ad albero, nella quale ad ogni termine è assegnata una notazione di classificazione e che esplica il massimo della sua forza all'interno di sistemi informatici. Tale struttura può essere costruita anche dall'alto verso il basso e deve prevedere un sistema notazionale-classificatorio complesso ed efficace, che sia ospitale a tutti i livelli e sia il più chiaro possibile per gli utenti poco esperti che potrebbero incontrare difficoltà nell'individuazione e nell'interpretazione della notazione. L'ultima delle presentazioni consigliate dallo standard NISO è quella per faccette, intese qui come principio di ordinamento e di raggruppamento dei termini, che rende presentabile in maniera più rigorosa la struttura sottostante: le faccette sono visualizzate come termini apicali della gerarchia. Lo stesso principio sottende la presentazione del *top term*. L'ultimo *display* analizzato è quello grafico, che visualizza i concetti e le loro relazioni mediante contenitori, disegni e frecce e risulta particolarmente *user friendly*, poiché comunica le relazioni tra termini in maniera più incisiva rispetto alla visualizzazione lineare. Lo standard sottolinea l'esistenza di numerosi prodotti commerciali che provvedono alla realizzazione grafica dei termini e delle loro relazioni, strumenti informatici che permettono di costruire collegamenti interattivi con il contenuto testuale finale, capaci di generare mappe concettuali utilizzando termini del vocabolario controllato o parole del linguaggio naturale. Il formato grafico, ottimale in ambito digitale, su supporto a stampa riduce notevolmente le sue potenzialità, presentandosi in maniera statica e perdendo la sua dinamicità. È questa la tipologia di visualizzazione cui fa riferimento la norma ISO; lo standard inglese BS invece menziona brevemente tale *display* nella parte introduttiva analizzando i vantaggi dell'impiego di un *database* per la conservazione delle informazioni.

Mentre lo standard NISO conclude la rassegna dei *display* con la presentazione grafica, la norma internazionale ISO e lo standard inglese BS si dedicano estesamente alla presentazione sistematica o classificata. La norma ISO introduce l'argomento con l'illustrazione della struttura e dei principi organizzativi del thesaurus sistematico-classificato, ponendo l'attenzione sui vantaggi e gli svantaggi dell'adozione di un sistema di organizzazione per settori o discipline e per faccette. In particolare, il primo tipo di suddivisione viene considerata adatta per i thesauri monodisciplinari e presenta il beneficio dell'amichevolezza nei confronti degli utenti, poiché, generalmente,

corrisponde al loro modo di pensare e di organizzare i concetti; dall'altro lato, tale impostazione comporta la difficoltosa interscambiabilità di vocabolari specialistici e uno sforzo di revisione non indifferente qualora emergano nuove discipline o nuovi campi specialistici. La suddivisione per faccette offre come vantaggio una maggiore interscambiabilità tra vocabolari differenti e una minore necessità di revisione dell'impianto complessivo, pur sparpagliando concetti altrimenti riuniti in gruppi disciplinari affini, ed offre una minore amichevolezza del thesaurus che può rimanere ostico all'utente finale. I due criteri possono essere adottati anche contemporaneamente, prevedendo una prima divisione disciplinare ed un'ulteriore classificazione interna per faccette. In tale contesto le etichette di nodo, criteri logici di divisione di una categoria, possono fungere da indicatori di faccetta per rendere manifesta la base logica sulla quale è stata organizzata la gerarchia, oppure per collocare termini di faccette differenti in prossimità del concetto con il quale vengono solitamente associati.

Lo standard inglese ripercorre l'impostazione della presentazione classificata dell'ISO, dedicando apposite sezioni ai criteri misti di organizzazione e di visualizzazione. Il criterio di ordinamento per argomento o disciplina è facilmente comprensibile dall'utente finale anche non esperto, tuttavia esso sottolinea come termini abitualmente utilizzati in numerosi campi, possano collocarsi in più posizioni all'interno della stessa disciplina, pertanto suggerisce di optare per un posizionamento multiplo del termine, laddove richiesto, oppure di creare una lista di termini generica, separata dal vocabolario, da usare all'occorrenza. L'ordinamento per faccette "puro", così come presentato dallo standard internazionale ISO, non compare nel BS che, invece, dedica largo spazio all'ordinamento misto disciplinare e per faccette, introducendo il concetto di "localizzazione principale" sviluppato nell'ottica dell'utilizzabilità dello strumento da parte degli utenti. Si tratta di un artificio che, in realtà, mostra la sua migliore applicazione nel formato a stampa, poiché il contesto digitale annulla la replicazione dei termini essendo questi sempre unici e collegati alle diverse gerarchie da rinvii ipertestuali: nell'ottica del risparmio dello spazio e della facilità di consultazione del vocabolario, ogni termine che abbia a trovarsi contemporaneamente in più posizioni sarà primariamente inserito nella gerarchia principale di appartenenza - ossia dove l'utente si aspetta di trovarlo - e nella posizione che dovrebbe occupare altrove vi sarà un rinvio alla localizzazione principale. Un sistema di notazioni agganciato ai termini permetterà di collocare il termine in più gerarchie, mostrandolo nella sua posizione principale e sostituendolo con la notazione riferita alla sua localizzazione principale nelle altre posizioni. L'utente sarà così facilitato nella consultazione di uno strumento ordinato con metodo rigoroso, quale quello classificato, pur senza conoscere le norme tecniche che sottostanno alla costruzione della struttura. La nozione di localizzazione principale è un espediente molto conveniente per restringere l'ampiezza dei thesauri e dei vocabolari a stampa, ma aggiunge gradini alla completa ricerca di un termine. Il formato elettronico rende il processo di visualizzazione molto più semplice, perché permette agli utenti di scegliere quale porzione di gerarchia rendere visibile, ampliando o restringendo la visualizzazione e mostrando tanti dettagli quanti ne vengono richiesti, navigando fino alla visualizzazione del record o *content object* finale: questo è fondamentalmente il tipo di *display* denominato classificato esteso. Non è un *display* realizzabile in contesto analogico, dove potrebbe essere sostituito da un sistema notazionale che tuttavia sacrificerebbe la leggibilità dello strumento al risparmio di spazio. Le stesse considerazioni vengono applicate nel BS alla visualizzazione delle relazioni poligerarchiche, poco efficaci in un vocabolario stampato, perfettamente integrabili in un ordinamento classificato e sistematico elettronico.

Lo standard americano, a differenza delle altre due norme che concludono la loro illustrazione dei *display* con la descrizione delle tipologie viste precedentemente, prosegue l'analisi dell'argomento affrontando i formati di visualizzazione: a stampa, elettronici statici (solitamente in PDF), dinamici (modalità *on-line*), come parte di un sistema integrato di recupero dell'informazione o applicazione web; infine, sistema misto statico/dinamico, ossia applicazione software da installare sui computer degli utenti finali. Per quanto attiene le considerazioni sui formati a stampa, si sottolinea la necessità di evitare le consultazioni di sequenze multiple, al fine di salvaguardare il tempo dell'utente; si propone come soluzione quella di aggiungere la presentazione gerarchica completa che, sulla carta significa consumo di spazio; a questo proposito si analizza nell'ambito della relazione di *use* la problematica della giustapposizione di termini: quando un termine non preferito segue o precede immediatamente il termine di accesso preferito potrebbe, per l'economia dello spazio, essere soppresso. Viene consigliato di licenziare comunque una versione elettronica del vocabolario, anche statica, in cui proporre la lista completa dei termini e produrre la versione a stampa priva dei termini superflui, in alternativa ogni vocabolario a stampa dovrebbe includere più indici che ne facilitino la consultazione.

Successivamente, viene introdotta la parte più innovativa relativa alla visualizzazione elettronica e all'ipertestualità; questa parte si basa su numerosi studi sull'interazione uomo-macchina e sullo sviluppo di interfacce amichevoli, senza dimenticare le problematiche legate all'accessibilità di persone con disabilità: standard di usabilità e accessibilità devono essere applicati rigorosamente nella progettazione e nella realizzazione del *display*.

L'ambiente elettronico permette la ricerca più semplice e intuitiva per l'utente mediante le parole chiave²¹ e risponde agli stessi obiettivi degli indici ruotati nel formato a stampa; mette l'utente in grado di decidere quando visualizzare il dettaglio del termine e selezionare la porzione di gerarchia da visualizzare. Tra le varie opzioni, NISO presenta anche la *pick list*: molto usata nei siti web è detta anche menù a cascata, permette all'utente di selezionare, posizionandosi sul punto di interesse, la terminologia adatta, conducendo ad un ristretto set di termini. Nell'ambito dell'indicizzazione semantica e dei vocabolari controllati si predilige fornire le *pick list* in abbinamento ai vocabolari organizzati per faccette. È indubbio che la navigazione ipertestuale offra dei vantaggi inimmaginabili negli altri contesti: facilità di navigazione delle gerarchie, collegamenti a tutti i livelli di dettaglio del termine, spostamento da una modalità di *display* all'altra, navigazione profonda fino al *contet object* finale.

La parte finale delle considerazioni è dedicata alla proposta di allegare sempre al vocabolario, qualunque sia il suo formato, una guida all'uso, rivolta a tutti i fruitori²². Un'appendice finale, che descrive in maniera sintetica caratteristiche e uso delle opzioni di visualizzazione dei vocabolari²³, proponendo per ogni tipologia un'elencazione delle caratteristiche propriamente dette e dei possibili usi pratici, conclude l'argomento.

21 Il termine digitato viene ricercato all'interno di tutti i descrittori siano essi semplici o composti (attraverso il troncamento, la ricerca "inizia con" e "contiene").

22 Nella versione a stampa dovrebbe trovare spazio nell'introduzione, nella versione elettronica potrebbe concretizzarsi in un documento PDF che l'utente può scaricare, stampare e conservare, in applicazioni software dinamiche dovrebbe prevedersi un apposito collegamento di *help*.

23 *Appendix C: Characteristic and Uses of Controlled Vocabulary Display Options*, p. 137-141.

2.3. L'interoperabilità

L'altra novità, su cui tanto interesse si sta concentrando, è la possibilità di realizzare un colloquio, semantico e tecnico, tra vocabolari controllati. Questi strumenti, pur osservando i criteri di costruzione e gestione precisati e normalizzati dagli standard di riferimento, possono presentare differenze notevoli, avere variabili domini disciplinari, prevedere una terminologia settoriale, diverse modalità di espressione formale dei singoli concetti²⁴ e adottare sistemi informatici differenti.

Come sottolinea lo standard NISO, la necessità di interoperabilità viene avvertita a vari livelli:

- dagli utenti che necessitano di eseguire una ricerca, usando una chiave d'interrogazione tratta da un vocabolario preferito, all'interno di contenitori multipli di risorse indicizzate anche con differenti thesauri o schemi di classificazione;
- dagli indicizzatori, che vogliono indicizzare i contenuti di un dominio disciplinare, usando la terminologia controllata di un altro ambito;
- dai produttori e distributori di contenuti che vogliono fondere due o più database indicizzati usando differenti vocabolari controllati;
- dalla comunità degli utenti, indicizzatori, produttori di contenuti, che vogliono una fusione di due o più vocabolari controllati, per formarne uno nuovo che racchiuda tutti i concetti e i termini contenuti negli originali;
- dalla comunità degli utenti multilingui che vogliono poter fruire di una modalità d'indicizzazione, di ricerca e di recupero tramite l'accesso linguistico preferito, in modo da reperire informazioni espresse in più di una lingua, a partire da un unico accesso.

Sul tema dell'interoperabilità, i due standard NISO e BS, nati a distanza di poco tempo l'uno dall'altro, fondamentalmente si integrano e ci forniscono raccomandazioni pratiche e indicazioni metodologiche per far colloquiare strumenti differenti di controllo semantico, così da agevolare il recupero di informazioni, indicizzate anche con differenti vocabolari, a partire da una medesima *query*. In entrambi gli standard, l'interoperabilità semantica viene affrontata separando il piano dell'indicizzazione da quello della ricerca.

L'indicizzazione, intesa dunque come assegnazione di termini per rappresentare il contenuto concettuale di entità portatrici di significato (*content objects*), frequentemente richiede la consultazione di un vocabolario controllato, di tipo generale o specializzato, più o meno orientato a quello che viene definito un particolare dominio²⁵.

Per quanto riguarda il versante della ricerca, l'interoperabilità diventa più complessa. La ricerca incrociata tra sistemi e domini, dove più vocabolari controllati sono

24 Nei diversi tipi di vocabolario i concetti sono rappresentati e normalizzati diversamente: nei thesauri con i termini preferiti; negli schemi di classificazione con notazioni; nelle tassonomie con etichette di categoria, notazioni, identificatori; negli schemi di intestazioni per soggetto con termini o stringhe precoordinate; nelle liste di autorità e nelle ontologie con termini o identificatori.

25 Gli indicizzatori e gli specialisti che hanno acquisito familiarità con uno specifico vocabolario controllato, possono scegliere di usarlo anche per indicizzare altri domini di contenuto. Questo processo può essere abbastanza proficuo se i nuovi domini condividono aspetti cospicui con il dominio originale, per il quale il vocabolario è stato sviluppato. Spesso succede che in determinati contesti organizzativi e istituzionali, dove si opera favorendo la cooperazione e l'integrazione, siano fusi più database i cui contenuti informativi siano stati indicizzati con differenti vocabolari controllati. In casi come questi, una delle soluzioni adottabili potrebbe essere quella di allestire un nuovo vocabolario controllato (cercando di integrare le due strutture "sorgenti" di controllo terminologico) e di indicizzare, nuovamente, i contenuti, usando i termini proposti dal nuovo strumento.

in uso, può intaccare sia la precisione che il richiamo. Per preparare differenti vocabolari controllati compatibili per la ricerca sarà necessario porre l'attenzione sui concetti e/o sui termini di mappatura usando l'interazione uomo-macchina.

Lo standard BS focalizza il tema dell'interoperabilità sul processo di mappatura, in quanto operazione indispensabile perché differenti lingue o linguaggi siano connotati e resi ricercabili. Spesso, professionisti e utenti comuni adottano una terminologia differente per esprimere il medesimo concetto, usando criteri di ricerca e di navigazione diversi. In questo caso specifico, ad esempio, potrebbe essere necessario creare mappature tra terminologia professionale, usata in una comunità scientifica specifica, ed equivalenti termini del linguaggio comune, presenti nella rete. Spesso in diversi contesti specialistico-settoriali è possibile trovare termini identici ma con significati differenti (per es. al termine "operazioni" sarà attribuito un significato diverso all'interno di un thesaurus di medicina o di un thesaurus di ambito militare), oppure termini che sono trattati come sinonimi in un vocabolario controllato di carattere generale, possono essere tutti preferiti all'interno di in un vocabolario controllato specialistico-settoriale; oppure ancora, in contesti disciplinari affini, i termini possono presentare delle differenze d'ambito d'uso.

Lo standard BS propone tre modelli di interoperabilità tra vocabolari in cui i termini vengono mappati dall'uomo con l'aiuto della macchina e con modalità diverse, come illustrato nella seguente sintesi:

Modello 1: unità strutturale (*structural unity*). Si ha quando i vocabolari che si vogliono mappare condividono la stessa struttura, ossia le relazioni gerarchiche e associative tra concetti sono sovrapponibili.

Modello 2: vocabolari non sovrapponibili (*non-equivalent pairs*). Si stabiliscono collegamenti fra vocabolari che non condividono la stessa struttura, ma stanno tutti sullo stesso livello. La mappatura dovrebbe essere definita tra alcuni o tutti i termini, notazioni o categorie di più di un vocabolario. Lo scopo della mappatura è quello di aiutare gli utenti a trovare informazioni in una collezione che è stata indicizzata con uno dei vocabolari, partendo da una fase di ricerca che usa termini o notazioni di un vocabolario differente.

Modello 3: struttura cardine (*backbone structure*). Si ha quando, durante l'operazione di mappatura fra due o più vocabolari, uno di questi viene preso come modello di riferimento.

Solitamente la mappatura dovrebbe essere stabilita attraverso la comparazione sistematica di due vocabolari, trovando per ogni termine del vocabolario di partenza un termine equivalente o una combinazione di termini nel vocabolario d'arrivo. La mappatura è difficile quando i due vocabolari sono di diverso tipo e, in particolare, quando uno dei due vocabolari contiene concetti complessi, espressi con termini composti, che, nell'altro, trovano una corrispondenza concettuale e terminologica con la combinazione di componenti separate. Lo standard inglese delinea quattro livelli di equivalenza tra termini:

- 1) equivalenza esatta, i termini del vocabolario di partenza corrispondono esattamente con quelli del vocabolario d'arrivo;
- 2) equivalenza inesatta, i termini dei due vocabolari sono sovrapponibili nell'ambito disciplinare ma non coincidono del tutto;
- 3) equivalenza parziale, in uno dei due vocabolari il termine esprime un concetto più ampio e generale;
- 4) non equivalenza, uno dei due vocabolari non contiene il termine presente nell'altro.

Lo standard conclude la parte delle varie tipologie trattando la tematica dell'equivalenza fra lingue, all'interno di un thesaurus multilingue, e del rapporto fra vocabolari costruiti con un'architettura differente.

Nel primo caso, il principio indicato come cardine è quello in base al quale tutte le lingue di un thesaurus multilingue hanno uno *status* identico e ogni concetto dovrebbe essere rappresentato in ogni lingua. Le equivalenze tra vocabolari dovrebbero essere stabilite tra i corrispondenti termini preferiti nelle varie lingue. Le singole relazioni gerarchiche e associative potrebbero essere rappresentate come relazioni di equivalenza tra i corrispondenti termini preferiti di ogni lingua. L'equivalenza tra termini non preferiti può essere utile quando i termini non sono proprio sinonimi o quando sono quasi sinonimi. Il BS analizza inoltre le varie soluzioni che si possono impiegare per stabilire una relazione di equivalenza fra vocabolari strutturati diversamente.

In questi anni l'attività di ricerca ha cercato di sviluppare, sia sul piano teorico che pratico, dei metodi che rendano i vocabolari controllati capaci di essere usati tra più *database* o tra più sistemi e per renderli, inoltre, condivisibili tra indicizzatori e ricercatori. Nell'appendice D lo standard NISO prende in esame i vari metodi per attuare l'interoperabilità²⁶:

- 1) derivazione/creazione di un modello: in tale approccio viene scelto un vocabolario controllato già esistente come modello per sviluppare o derivare altri vocabolari controllati che potranno essere più o meno interoperabili. Questo serve per stabilire un impianto di base che garantisca una coerenza nella selezione dei termini nella struttura gerarchica. Ad esempio molti utenti dell'AAT²⁷ usano solo una porzione del thesaurus per indicizzare la propria collezione e a questo gruppo di termini aggiungono la propria terminologia, che si sviluppa sulla base della garanzia bibliografica locale;
- 2) traduzione/adattamento: si attua sviluppando nuovi vocabolari controllati, sulla base della traduzione in altre lingue di vocabolari controllati già esistenti;
- 3) vocabolario controllato satellite: l'approccio satellitare per costruire un nuovo vocabolario controllato è simile alla derivazione/creazione di un modello. In questo caso, il vocabolario controllato sorgente funge e diventa un meta-vocabolario che ingloba al suo interno vocabolari specializzati;
- 4) nodo o *leaf linking*: la base è costituita da un vocabolario "madre" la cui struttura gerarchica è dotata di "nodi", che permettono di "linkare" a famiglie di vocabolari. Questo approccio garantisce che la nuova famiglia dei vocabolari controllati sia composta dai termini della struttura, dalla struttura del termine, dall'ambito d'uso e dalle linee guida e permetta vari livelli di specificità, senza sommergere il vocabolario controllato originale con dettagli di cui la maggior parte degli utenti non necessita;
- 5) commutazione (*switching*): realizzato sviluppando un linguaggio o un sistema di comunicazione che operi a livello intermedio, in modo da consentire il passaggio da un termine al suo equivalente in diversi vocabolari;
- 6) mappatura automatizzata (*co-occurrence mapping*) e mappatura intellettuale (*direct mapping*): nel primo caso, i termini oggetto di mappatura sono scelti da campi di metadati all'interno di database sottostanti, da parole chiave non controllate, da altri metadati assegnati al contenuto o da parole scelte dalle stesse informazioni in *full text*; nel secondo caso, sviluppando intellettualmente un sistema di equivalenze tra termini di diversi vocabolari controllati, tra termini e indici di classificazione, ecc.;
- 7) collegamento attraverso una lista unificata temporanea (*linking through a temporary union list*): differenti vocabolari controllati sono collegati attraverso la mappatura informatica di termini che non sono concettualmente equivalenti ma che sono

²⁶ Appendix D: *Methods for Achieving Interoperability*, p. 142-146.

²⁷ Art & Architecture Thesaurus Online <http://www.getty.edu/research/conducting_research/vocabularies/aat/>.

in qualche modo strettamente collegati. La mappatura è fatta dinamicamente, in risposta a una domanda specifica dell'utente, e i risultati sono visualizzati in una lista temporanea. Non sono prodotti né una tabella di mappatura né un nuovo vocabolario controllato. Alcuni sistemi, inoltre, usano l'incapsulamento di un *semantic network*, generato da uno o più *database* scelti per mappare la terminologia richiesta dall'utente dentro le parole che occorrono nelle varie informazioni;

8) collegamento tra più server di vocabolari controllati: un approccio supplementare all'interoperabilità implica la trasmissione di una domanda ad uno o più server di vocabolari controllati. Le domande degli utenti sono inoltrate ad un server centrale che le invia, tramite un protocollo di trasmissione, ai singoli server che supportano i vocabolari controllati, questi restituiscono le risposte e l'unità centrale mette insieme i vari risultati e li presenta all'utente in una forma consolidata.

2.3.1 I formati di scambio e i protocolli per l'interoperabilità

La quinta parte dello standard BS è la più in *progress* e per questo è una versione provvisoria soggetta a sperimentazione, analisi e cambiamento. In accordo con i suggerimenti di modifica ricevuti, il responsabile del BSI Committee deciderà di convertirla in *British Standard*. La suddetta parte detta alcune linee guida sullo scambio dei dati di vocabolari controllati tra differenti applicazioni coinvolte nel ricupero dell'informazione.

In un ambiente di rete, un vocabolario server solitamente deve inviare informazioni a una molteplicità di client che possono essere motori di ricerca, sistemi di indicizzazione, o altri sistemi di gestione di vocabolari. Per supportare queste funzioni la configurazione del formato di scambio e del protocollo deve avere una struttura tale da poter permettere l'interazione tra l'applicazione di un computer server per un vocabolario strutturato e un'altra applicazione client, applicando quel vocabolario strutturato al sistema di ricupero dell'informazione.

L'interoperabilità tra applicazioni differenti è favorita dalle seguenti componenti: infrastruttura di rete; formati per lo scambio di dati; protocolli, includendo protocolli di comunicazione base tra computer (per esempio quelli applicati specificatamente ai vocabolari); archivi di vocabolari, metadati e altri strumenti comuni; solida architettura di dati.

Lo standard propone e predilige come formato per lo scambio dei dati di un intero thesaurus o di una sua parte, il linguaggio XML (*Extensible Markup Language*). Come è noto, questo "linguaggio marcatore", permette, appunto, tramite *tag*, di estendere/creare ulteriori linguaggi marcatori. È un linguaggio autodescrittivo, cioè la sua definizione è contenuta all'interno del linguaggio stesso (i *tag* e il loro contenuto), e per la rappresentazione dei caratteri di tutte le lingue del mondo si avvale della modalità UNICODE²⁸.

Lo standard inglese illustra inoltre altri formati di scambio particolarmente diffusi. Il MARC (*Machine Readable Cataloguing*) è uno standard ampiamente usato per lo scambio di informazioni bibliografiche tra biblioteche. Sebbene sia più frequentemente usato per lo scambio di record bibliografici, offre un formato per le liste di autorità che può essere usato anche per trasmettere i dati di un thesaurus. SKOS (*Sim-*

²⁸ Il formato corrisponde al modello *UML* (*Unified Modelling Language*) ed è architettato per trasmettere i dati dei thesauri rispondenti fedelmente o parzialmente allo standard *BS 8723:2* e i dati dei thesauri multilingui che seguono le indicazioni del *BS 8723:4*; per le altre tipologie di vocabolari controllati non vengono fornite indicazioni.

ple Knowledge Organization Systems), è un formato particolarmente indicato per l'applicazione a sviluppi del *semantic web* ed è codificato usando XML e RDF (*Resource Description Framework*). Esso è applicabile oltre che ai thesauri anche agli schemi di classificazione, tassonomie e altri tipi di vocabolario. *Zthes*, in origine sviluppato come un'applicazione del protocollo Z39.50, si è molto diffuso e, ad oggi, può essere applicato anche ai protocolli SRW/SRU (*Search/Retrieve Web service, Search/Retrieve via URL*).

I protocolli più usati e proposti dal BS per facilitare l'interoperabilità tecnica e le operazioni di indicizzazione e ricerca tra thesauri, sono i seguenti: *ADL Thesaurus Protocol (Alexandria Digital Library)*, è un protocollo per l'accesso ai thesauri monolingui, supporta l'interrogazione e la navigazione in rete del thesaurus; *SKOS API*, applicazione del *Web Service API (Application programming Interface)*, strutturata per permettere l'accesso ai thesauri e a altri sistemi di organizzazione della conoscenza attraverso il web. Definisce una serie di operazioni per accedere e interrogare il thesaurus; *SRW/SRU (Search and Retrieve via the Web)*, per il quale il *SOAP protocol o URLs* forniscono un meccanismo di accesso e di ricupero "Search and Retrieve via URLs"; *Z39.50*, più formalmente conosciuto come *ANSI/NISO Z39.50, Information retrieval (Z39.50): Application service definition and protocol specification*, strutturato per interrogare database generici, non soltanto vocabolari.

La sezione viene conclusa con una serie di esempi di applicazione dello schema XML, che illustrano le diverse funzioni dei modelli di dati.

3. Il nuovo standard ISO 25964

Nel 2007 è stata fatta la proposta di revisione, fusione e ampliamento degli standard sulla costruzione dei thesauri e sull'analisi concettuale ISO 2788:1986 e ISO 5964:1985, per rispondere alle nuove esigenze tecnico-scientifiche di interoperabilità e di integrazione tra vocabolari controllati diversi ed in particolare per suggerire criteri specifici di codifica dei requisiti funzionali dei software per la gestione e la visualizzazione elettronica, delle procedure di mappatura tra thesauri e altri tipi di vocabolari controllati, delle architetture e dei formati per lo scambio di dati tra vocabolari controllati.

È stato istituito un gruppo di lavoro chiamato *WG8 (Structured Vocabularies)*²⁹, diretto da Stella Dextre Clarke³⁰ e coordinato da una sottocommissione tecnica l'ISO TC 46/SC9. Il gruppo ha dato vita al progetto ISO NP 25964, decidendo di usare lo standard BS 8723:2005-2008 come base di riferimento per la revisione dei precedenti standard ISO.

A differenza dello standard inglese, organizzato in cinque parti, il nuovo ISO 25964, *Information and documentation-Thesauri and interoperability with other vocabularies*, si basa esclusivamente su due sezioni:

Part. 1 *Thesauri for information retrieval*, focalizzata sullo sviluppo e il mantenimento dei thesauri, monolingui e multilingui, includendo formati e protocolli per lo scambio di dati. Dello standard BS 8723 acquisisce interamente le sezioni 1 e 2 e parzialmente le sezioni 4 e 5;

Part. 2 *Interoperability with other vocabularies*: incentrata sull'interoperabilità tra differenti thesauri e tra thesauri e altri tipi di strumenti strutturati, come schemi di classificazione, tassonomie e ontologie. Fornisce una guida sulla mappatura dei dati tra

29 Formato da rappresentati di dodici paesi: Canada, Danimarca, Finlandia, Francia, Germania, Nuova Zelanda, Sud Africa, Spagna, Svezia, Regno Unito, Ucraina, USA.

30 Stella G. Dextre Clarke, *ISO 2788 + ISO 5964 + Much Energy = ISO 25964*, «Asis&T bulletin», (oct.-nov. 2008) <http://www.asis.org/Bulletin/Oct-o8/Oct-Nov08_DextreClarke.html>.

strutture diverse e l'architettura dei modelli e protocolli per l'interoperabilità. Dello standard BS acquisisce tutta la sezione 3 e parte della sezione 4³¹.

I capitoli 1-14 della prima parte corrispondono ampiamente al contenuto delle ISO 2788:1986 e delle ISO 5964:1985. I capitoli restanti e l'intera parte seconda, coprono tematiche scientifiche che nei precedenti standard non sono state trattate.

Nel dicembre del 2008 è stato distribuito alla comunità scientifica il *draft* della prima parte delle ISO/CD 25964-1 per eventuali revisioni e suggerimenti. Non resta che attendere ed auspicare, in tempi brevi, la versione finale del nuovo standard internazionale.

31 Stella G. Dextre Clarke, *Standards for multilingual thesauri*, in: *Metadata Search and Retrieval Seminar 2: "Speaking in Tongues"*, London, 3rd October 2008, <http://www.mileproject.eu/asset_arena/document/KE/MILE_LONDON_S&R_OCT_3_2008_DEXTRE_CLARKE.PPT>.

From ISO 2788 to the new standards for the construction and interoperability of structured vocabularies: a comparative analysis

by Tiziana Calvitti and Elisabetta Viti

The article aims at comparing the international norm ISO 2788:1986 and the subsequent standards BS 8723:2005-2008 and ANSI/NISO Z39.19-2005, concentrating on the new elements that the two British and American standards bring to the theme of structured vocabularies and their interoperability. The preparation of BS 8723:2005-2008 and ANSI/NISO Z39.19-2005 testifies to a constant and renewed interest for indexing. The new standards demonstrate how, even in the light of modern possibilities for research offered by the WEB, structured vocabularies are still excellent tools, forming the infrastructure, the basis and the bone structure for regulating and formalizing on the one hand indexing activity and on the other user research.

The most important new points of the British BS and American ANSI/NISO standards with respect to the ISO 2788, regard the increase in types of structured vocabularies analysed: as well as the thesauri, examinations are made of lists, *synonym rings*, taxonomies, lists of authorities, ontologies and classification tables. Another new element is the presentation of forms of data visualization (so-called *displaying*), with special attention being paid to the transformations made by web technologies and hypertexts. Lastly, recent standards lay special emphasis on the subjects of semantic and technical interoperability between different thesauri or between thesauri and other types of structured vocabularies. In this regard English and American standards offer models and methods of mapping of structured vocabularies in order to activate a dialogue between different tools. The exchange formats and interoperability protocols are the object of analysis of the fifth part of the BS standard, not by chance a topic *in progress* discussed in a provisional version that is still in the process of being experimented, analysed and changed, while awaiting validation as a *British Standard*.

The article concludes with a rapid examination of the draft of the new ISO 25964 standard, born out of a proposal advanced in 2007 for the revision, fusion

TIZIANA CALVITTI, Biblioteca comunale di Terni, piazza della Repubblica, 1, 05100 Terni, e-mail tiziana.calvitti@comune.terni.it.

ELISABETTA VITI, Centro servizi biblioteca, Facoltà di lettere e filosofia, Università degli studi di Siena, via Fieravecchia 19, 53100 Siena, e-mail viti2@unisi.it; studying for a research doctorate in library, archive, documentary sciences and in the preservation and restoration of book and archivi heritage, at the University of Udine.

and extension of ISO 2788:1986 and ISO 5964:1985. The WG8 work group (*Structured Vocabularies*), coordinated by a technical sub-committee ISO TC 46/SC9 produced the ISO NP 25964 project, deciding to use the BS 8723:2005-2008 standard as a reference base for the revision of the previous ISO standards. The draft of the first part of the ISO/CD 25964-1 was distributed to the scientific community in December 2008.

The future standard will have to meet new requirements of interoperability and integration between different structured vocabularies and will have to offer specific criteria for coding the functional software requisites for electronic control and visualization, procedures for mapping between thesauri and other types of structured vocabularies, and architecture and formats for data exchange.