

Report dalla 7th European Conference on Digital Libraries, ECDL 2003, Trondheim, 17-22 Agosto 2003

Antonella Zane

Centro di Ateneo per le Biblioteche – Università degli Studi di Padova
antonella.zane@unipd.it

ECDL , la più importante conferenza europea sulle *Digital Library* (DL), ha aperto i lavori della sua settima edizione (2003) a Trondheim in Norvegia il 17 agosto 2003 con un'inaspettata partecipazione di pubblico (360 iscritti provenienti da 35 Paesi) e un ricco programma che prevedeva: una giornata dedicata ai *tutorial* (link al programma nel sito <http://www.ecdl2003.org/>), tre giorni di conferenza vera e propria (dal 18 al 20) e due giornate di *workshop* (di cui sono stati pubblicati dei brevi rapporti in Dlib-Magazine (<http://www.dlib.org/dlib/september03/09inbrief.html>)).

Le giornate di conferenza hanno seguito un programma alquanto articolato che comprendeva tre sessioni plenarie, tre dibattiti (panel), 14 sessioni tematiche, una breve presentazione plenaria dei poster e delle dimostrazioni seguite dalla loro esibizione nelle sale della sede del convegno.

Sia i lavori che i poster presentati al convegno hanno superato una forte selezione operata dal Comitato di Programma che ha accettato solo 47 lavori su 162 ricevuti e 52 tra poster e dimostrazioni su 72 ricevuti.

Cinque i contributi italiani presenti nel programma di cui due (una comunicazione orale e un poster) sono stati presentati dall'Università di Padova: "Annotations: Enriching a Digital Library" di Maristella Agosti e Nicola Ferro (Dipartimento di ingegneria dell'informazione) e "LVDL: the first digital thematic library on the Lagoon of Venice" di Stefania Biagioni, Carlo Carlesi, Lisa Locatello, Monica Ortolan, Antonella Zane (Biblioteca Vallisneri e ISTI – CNR di Pisa).

Gli studi presentati nell'ambito delle numerose sessioni tematiche coprono gran parte delle aree di ricerca d'interesse per le biblioteche digitali:

- concetti generali, metodi, standard, aspetti economici e strategie
- costruzione, gestione e integrazione delle collezioni
- architettura e interoperabilità
- *knowledge management, discovery e retrieval*
- interazione con l'utente

In questa relazione viene riportata la sintesi di una selezione delle comunicazioni orali previste dal programma. Gli interventi presi in esame sono contrassegnati da un numero tra parentesi quadre, in ordine crescente. Una copia cartacea integrale degli Atti del Convegno è disponibile presso la sede del CAB.

PROGRAMMA DEL CONVEGNO

SESSIONI PLENARIE

- Digital Libraries: What Should We Expect from Search Engines [1]
- Stewardship in the digital age [2]
- Information retrieval research and digital libraries [3]

PANEL

- The Future of Academic Publishing
- Digital? Libraries?, all these question-marks [4]
- Digital preservation: Are Metadata Really Crucial?

SESSIONI TEMATICHE

Uses, users and user interaction

- Users and Uses of Online Digital Libraries in France. Houssein Assadi, Thomas Beauvisage, Catherine Lupovici and Thierry Cloarec. [5]
- In search for patterns of user interaction for digital libraries. Jela Steinerová. [6]
- Detecting Research Trends in Digital Library Readership. Johan Bollen, Rick Luce, Somesekhar Vemulapalli and Weining Xu.
- Evaluating the Changes in Knowledge and Attitudes of Digital Library Users. Gemma Madle, Patty Kostkova, Jane Mani-Saada and Julius R Weinberg. [7]

Metadata applications

- Towards a Role-Based Metadata Scheme for Educational Digital Libraries : A Case Study in Singapore. Dian Melati Md Ismail, Ming Yin, Yin-Leng Theng, Dion Hoe-Lian Goh and Ee-Peng Lim.
- Incorporating Educational Vocabulary in Learning Object Metadata Schemas. Jian Qin and Carol Jean Godby.
- Findings from the Mellon Metadata Harvesting Initiative. Martin Halbert, Joanne Kaczmarek and Kat Hagedorn. [8]
- Semantic Browsing. Alexander Faaborg and Carl Lagoze. [9]
- Metadata Editing by Schema. Hussein Suleman.

Annotation and recommendation

- Annotations: Enriching a Digital Library. Maristella Agosti and Nicola Ferro. [10]
- Identifying Useful Passages in Documents based on Annotation Patterns. Frank Shipman, Morgan Price, Catherine C. Marshall and Gene Golovchinsky. [11]
- Others Also Use: A Robust Recommender System for Scientific Libraries. Andreas Geyer-Schulz, Andreas Neumann and Anke Thede. [12]

Automatic classification and indexing

- Cross-Lingual Text Categorization. Nuria Bel, Cornelis H.A. Koster and Marta Villegas.
- Automatic multi-label subject indexing in a multilingual environment. Boris Lauser and Andreas Hotho.
- Automatic Induction of Rules for Classification and Interpretation of Cultural Heritage Material. S. Ferilli, F. Esposito, T.M.A. Basile and N. Di Mauro.
- An Integrated Digital Library Server with OAI and Self-Organizing Capabilities. Hyunki Kim, Chee-Yoong Choo and Su-Shing Chen

Web technologies

- YAPI: Yet Another Path Index for XML searching. Giuseppe Amato, Franca Debole, Pavel Zezula and Fausto Rabitti.
- Structure-Aware Query for Digital Libraries: Use Cases and Challenges for the Humanities Christopher York, Clifford Wulfman and Greg Crane
- Combining DAML+OIL, XSLT and probabilistic logics for uncertain schema mappings in MIND. Henrik Nottelmann and Norbert Fuhr.
- Digitometric Services for Open Archives Environments. Tim Brody, Simon Kampa, Stevan Harnad, Les Carr and Steve Hitchcock. [13]

Topical crawling. Subject gateways

- Search Engine-Crawler Symbiosis: Adapting to Community Interests. Gautam Pant, Shannon Bradshaw and Filippo Menczer.
- Topical Crawling for Business Intelligence. Gautam Pant and Filippo Menczer.
- SozioNet: Networking Social Science Resources. Wolfgang Meier, Natascha Schumann, Sue Heise and Rudi Schmiede. [14]

- VASCODA: A German Scientific Portal for Cross-Searching Distributed Scientific Digital Resource Collections. Heike Neuroth and Tamara Pianos. [15]

Architectures and systems

- Scenario-Based Generation of Digital Library Services. Rohit Kelapure, Marcos André Gonçalves and Edward A. Fox.
- An Evaluation of Document Prefetching in a Distributed Digital Library. Jochen Hollman, Anders Ardö and Per Stenström.
- An Architecture for Online Information Integration on Concurrent Resource Access on a Z39.50 Environment. Michalis Sfakakis and Sarantos Kapidakis.

Knowledge organization: Concepts

- The ADEPT Concept-based Digital Learning Environment. T. R. Smith, D. Ancona, O. Buchel, M. Freeston, W. Heller, R. Nottrott, T. Tierney and A. Ushakov.
- A User Evaluation of Hierarchical Phrase Browsing. Katrina D. Edgar, David M. Nichols, Gordon W. Paynter, Kirsten Thomson and Ian H. Witten.
- Visual Semantic Modeling of Digital Libraries. Qinwei Zhu, Marcos André Gonçalves, Rao Shen, Lillian Cassell and Edward A. Fox. [16]

Collection building and management

- Connecting Interface Metaphors to Support Creation of Path-Based Collections. Unmil P. Karadkar, Andruid Kerne, Richard Furuta, Luis Francisco-Revilla, Frank Shipman and Jin Wang.
- Managing Change in a Digital Library System with many Interface Languages. David Bainbridge, Katrina D. Edgar, John R. McPherson and Ian H. Witten.
- A Service for Supporting Virtual Views of Large Heterogeneous Digital Libraries. Leonardo Candela, Donatella Castelli and Pasquale Pagano.

Knowledge organization: Authorities and works

- A Framework for Unified Authority Files: A Case Study of Corporate Body Names in the FAO Catalog. James Weinheimer and Kafkas Caprazli.
- geoXwalk – A Gazetteer Server and Service for UK Academia. James S. Reid.
- Utilizing Temporal Information in Topic Detection and Tracking. Juha Makkonen and Helena Ahonen-Myka.
- Automatic Conversion from MARC to FRBR. Christian Mönch and Trond Aalberg.

Information retrieval in different application areas

- Musescape: A Tool for Changing Music Collections into Libraries. George Tzanetakis.
- A Digital GeoLibrary: Integrating Keywords And Place Names. Mathew Weaver, Lois Delcambre, Leonard Shapiro, Jason Brewster, Afrem Gutema and Timothy Tolle.
- Document-Centered Collaboration for Scholars in the Humanities - The COLLATE System. Ingo Frommholz, Holger Brocks, Ulrich Thiel, Erich Neuhold, Luigi Iannone, Giovanni Semeraro, Margherita Berardi and Michelangelo Ceci.

Digital preservation

- DSpace as an Open Archival Information System: Current Status and Future Directions. Robert Tansley, Mick Bass and MacKenzie Smith.
- Preserving the Fabric of Our Lives: A Survey of Web Preservation Initiatives. Michael Day.
- Implementating Preservation Strategies for Complex Multimedia Objects. Jane Hunter and Sharmin Choudhury.

Indexing and searching of special document and collection information

- Distributed IR for Digital Libraries. Ray R. Larson.
- Reference Directed Indexing: Redeeming Relevance for Subject Search in Citation Indexes. Shannon Bradshaw.
- Space-Efficient Support for Temporal Text Indexing in a Document Archive Context. Kjetil Nørvag.
- Clustering Top-Ranking Sentences for Information Access. Anastasios Tombros, Joemon Jose and Ian Ruthven.

[1] La prima sessione plenaria è stata condotta da John M. Lervik, CEO nonché co-fondatore di Fast Search & Transfer ASA (FAST), una compagnia nata nel 1997 dall'impegno di un gruppo di ingegneri e ricercatori della NTNU (Norwegian University of Science) di Trondheim. Il motore di ricerca FAST è alla base di numerosi servizi web sviluppati da importanti società commerciali quali Dell, Freeserve, IBM, Reed Elsevier, Reuters, T-Online (Deutsche Telekom), Thomas Publishing, Virgilio (Telecom Italia).

Con un intervento dal titolo "Digital Libraries: What Should We Expect from Search Engines" Lervik ha illustrato le potenzialità dei motori di ricerca di terza generazione non solo nel loro tradizionale compito di ricerca delle informazioni ma anche nella gestione delle informazioni e dei metadati 2) nelle funzioni di personalizzazione 3) come potenti strumenti di navigazione e *information discovery*.

Dopo un breve confronto tra le prestazioni di un motore di ricerca e quella di un database, Lervik si è soffermato sul funzionamento e le caratteristiche del motore di ricerca di terza generazione. Grazie ad analisi linguistiche e strutturali sofisticate nonché all'analisi ortografica, morfologica, sintattica e semantica è possibile aumentare sia la rilevanza (*relevance*) delle pagine recuperate sia il richiamo (*recall*).

Per il ranking dei documenti viene utilizzato il CASFQ che valuta, per ciascun documento, i seguenti parametri: *Completeness, Authority, Statistics, Quality, Freshness*.

La visualizzazione dei risultati può essere arricchita con strumenti di *information discovery* quali ad es. *Live Analytics*, introdotto recentemente nel prodotto FAST. Si tratta di uno strumento per il *Dynamic Drill-Down*, cioè la navigazione tra i risultati di una ricerca, che permette di scorrere alcuni indici (titolo della rivista, autore, sostanze chimiche, descrittori semantici quali ad esempio i Mesh di Medline).

Il commento finale di Lervik è quello di usare i database relazionali per le transazioni e i motori di ricerca per il *retrieval*. Questa è la strada che ha scelto anche Elsevier quando, al fine di migliorare le prestazioni nel recupero dei documenti, ha deciso di trasferire tutti i suoi dati dal database relazionale al motore di ricerca.

[2] Clifford Lynch è direttore della Coalition for Networked Information (CNI) dal luglio 1997. Alla CNI, sponsorizzata dall'Association of Research Libraries e Educause, partecipano come membri duecento organizzazioni che si occupano dell'uso dell'IT e dell'informazione in rete al fine di migliorare e supportare la produzione accademica e intellettuale. Lynch si è anche occupato, con diversi incarichi, di conservazione del digitale e proprietà intellettuale.

Il suo intervento, intitolato "Stewardship in the digital age", ha riguardato il problema della conservazione del nostro patrimonio culturale in formato digitale e la necessità di intervenire in tempi rapidi e con strategie adatte per garantirne la conservazione non tanto a lungo termine quanto nell'immediato futuro. Tale obiettivo deve essere perseguito a tre livelli: istituzionale (utilizzando il sistema della replicazione a grande scala dei dati); delle comunità o nazioni (grazie alla duplicazione, back-up e restauro delle collezioni culturali che possono coinvolgere anche popolazioni di Paesi confinanti); individuale (attraverso l'utilizzo di sistemi sicuri per la conservazione e l'accesso dei dati personali -quali ad esempio foto, e-mail video - che potrebbero fornire nell'immediato futuro nuove possibilità di espansione del mercato dell' IT).

[3] Karen Sparck Jones è docente di Computers and Information al Computer Laboratory, dell'Università di Cambridge. Le sue ricerche nell'ambito dell'*Information Retrieval* (IR) hanno ottenuto numerosi riconoscimenti tra i quali, nel 2002, quello dell'American Society for Information Science and Technology. Recentemente le sue attività di ricerca si sono concentrate sul *retrieval* dei documenti parlati (*spoken document*) e sul *summarizing*. Partendo dalla constatazione che l'indicizzazione e la ricerca automatica di documenti, da tempo un'area di ricerca per le *computing applications*, non ha interessato per molti anni i servizi bibliografici tradizionali, K. S. Jones ha voluto illustrare, nel suo intervento "Information retrieval research and digital libraries" quali sono le potenzialità della tecnologia per l'IR nel nuovo contesto delle biblioteche digitali, proponendo una panoramica sulla storia della ricerca sulle tecniche di IR.

Negli anni '60 l'obiettivo principale degli studi di IR era quello di automatizzare le strategie esistenti e di condurre analisi sulla frequenza delle parole. I test venivano effettuati su piccole collezioni di documenti. Negli anni successivi ('70-'80) gli studiosi svilupparono le idee del '60 arrivando all'elaborazione di modelli teorici che poterono supportare anche con evidenze sperimentali basate su collezioni più grandi. I risultati di queste ricerche tuttavia non trovavano ancora una applicazione nell'ambito di contesti operativi reali quali ad es. sistemi bibliotecari automatizzati. Le esigenze delle biblioteche non coincidevano con quelle della ricerca nell'IR tranne che per il linguaggio naturale e il full text. Gli importanti sviluppi che si sono avuti, a partire dagli anni '90, nel campo dell'*Information Technology* (collegamento delle macchine in reti, gli oggetti multimediali, il Web) e del *Natural*

Language Processing hanno avuto il duplice positivo effetto di stimolare la ricerca IR e il rapporto tra ricerca e mondo reale. In questi anni sono stati avviati programmi di valutazione quali DARPA, NIST, ARDA e test con molti utenti su collezioni di ampie dimensioni (TREC). Gli studi attuali di IR hanno come oggetto la sperimentazione con lingue nuove (es. il cinese), nuove tipologie di documenti e nuovi media (audio, video etc). Altre funzionalità avanzate riguardano inoltre il *summarizing* e il *question answering systems*. Il primo, che consiste nel riassumere automaticamente dei testi, è stato applicato anche nel campo dell'informazione giornalistica (es. il notiziario "Columbia Newsblaster: Summarizing all the News on the Web" riassume le notizie prendendole da decine di articoli pubblicati in tutto il mondo). Il secondo prevede che a una domanda qualsiasi in linguaggio naturale si ottenga una risposta in linguaggio naturale. Dopo aver auspicato una maggior collaborazione e comunicazione tra biblioteche e ricercatori di IR, la Sparen ha invitato i bibliotecari ad utilizzare con maggior tempestività i risultati della ricerca nel campo dell'IR e ad utilizzare il più possibile i metodi statistici per la gestione dell'informazione.

[4] Il Panel "Digital? Libraries?, all these question marks" presieduto da Stefan Gradmann dell'Università di Amburgo si poneva l'obiettivo di indagare quanto appropriato e attuale ancora sia il termine "digital library" e quali relazioni esistano tra il settore di Ricerca & Sviluppo delle biblioteche digitali e le biblioteche reali. Il tema è stato affrontato da quattro diversi relatori, ciascuno interprete di una particolare visione: le biblioteche, il mondo digitale, US/National Science Foundation, Europa/Comunità europea.

Dal punto di vista delle biblioteche (Jela Steinerová, Università di Bratislava) questo termine identifica un sistema complesso che fornisce supporto agli utenti nel trovare informazioni e accesso intellettuale alle informazioni in contesti molto speciali. Il concetto di DL introduce anche delle domande, secondo la Steinerová, "metaforiche":

- può il bibliotecario amare i suoi utenti nel mondo digitale?
- può la DL integrare il mondo fisico e cognitivo?
- il materiale digitale è reale (fisico) o solo cognitivo?

La ricerca sulle DL introduce quindi aspetti importanti per le biblioteche: l'approccio centrato sull'utente, l'organizzazione della conoscenza a diversi livelli (individuale, comunità, disciplinare), l'introduzione di nuovi servizi etc. In conclusione, secondo la Steinerová, il termine DL racchiude in sé tradizione, creatività e nuovi problemi (accessibilità, conservazione etc.)

Quanto al versante digitale (Dieter Fellner Università Tecnica di Braunschweig), va notato che ad oggi la maggior parte delle DL contiene ancora solo documenti di testo, rappresentando quindi una minima parte delle informazioni disponibili. E' importante che le biblioteche comincino a trattare anche nuovi media acquisendo nuove competenze nella classificazione e indicizzazione dei documenti non testuali (quali ad es. i video, gli audio etc.) nonché nella *summarization* (basata su tecniche di *sentence compression*).

La UE (Vittorio Casarosa CNR Pisa) ha affrontato questo tema finanziando alcuni progetti importanti quali ad esempio la costituzione di una rete di eccellenza di centri europei di ricerca sulle DL chiamata Delos. Le DL dovrebbero permettere alla gente di accedere alla conoscenza umana per mezzo di strumenti connessi a Internet. Questa visione tuttavia comporta una serie di problemi che riguardano: i contenuti (come mescolare oggetti diversi nella stessa DL); la conservazione (a breve e lungo termine); la gestione (come individuare e disegnare l'architettura); l'usabilità (visualizzazione); gli *stakeholders* (università, editori, industria del sw, industria delle telecomunicazioni, le *memory-based organisations* cioè biblioteche, musei, archivi). Sono necessarie inoltre politiche chiare che indichino cosa digitalizzare del cartaceo, cosa conservare dell'elettronico, l'adozione di nuove tecnologie, la realizzazione di nuovi servizi. Si tratta cioè di passare dalla "Digital Library" al "Digital Information REALM'S (Resources embedded in Archivi, Library, Museum)"

Dan Atkins dell'Università del Michigan ha illustrato alcuni temi presenti nello studio finanziato dall'americana National Science Foundation "Science and Engineering Infrastructure for the 21st Century: the role of the National Science Foundation". La crescita esponenziale della potenza dei

computer, della ampiezza di banda per la comunicazione e della capacità di immagazzinare dati non può non avere un forte riflesso sulle infrastrutture di ricerca, offrendo nuove opportunità e sfide per le forze sociali e tecnologiche. In questa cornice, l'evoluzione che coinvolge le biblioteche può essere rappresentata come segue:

biblioteche tradizionali → biblioteche ibride → biblioteche digitali → collezioni digitali di comunità (*digital collectives*) → cyberinfrastrutture avanzate (CI) → comunità virtuali per l'insegnamento e la ricerca (*e-learning, grid teams, e-science*) → università del mondo.

Questa rivoluzione (vedi "Preparing for the Revolution: Information Technology and the Future of the Research University" la cui versione elettronica è disponibile gratuitamente alla pagina <http://books.nap.edu/catalog/10545.html>) è accompagnata da importanti tendenze quali il software libero, gli archivi aperti e i consorzi. Per il mondo delle biblioteche digitali ciò implica la necessità di impegnarsi su nuovi fronti quali quello della Ricerca & Sviluppo delle DL e dello sviluppo delle risorse umane.

[5] **Houssein Assadi** ha presentato i risultati di un progetto, chiamato BibUsages, nato dalla collaborazione tra la Bibliothèque Nationale de France e la France Telecom R&D.

L'obiettivo del progetto era quello di studiare la popolazione di utenti remoti che consulta le biblioteche digitali in Internet, prendendo come campione gli utenti di Gallica, la biblioteca digitale della biblioteca nazionale francese. Le caratteristiche e i comportamenti di ricerca e navigazione di questa particolare fascia di utenti remoti differiscono alquanto sia da quelle dell'utente medio di Internet che da quelle dell'utente delle biblioteche tradizionali: si tratta in media di ultraquarantenni che utilizzano Gallica per ricerche personali, che posseggono – in percentuale di gran lunga superiore rispetto alla popolazione media – un accesso a Internet a banda larga e che navigano preferenzialmente dalle pagine della biblioteca digitale a quelle dei motori di ricerca e dei siti commerciali (es. Amazon).

[6] **Jela Steinerová** ha riportato i primi risultati di uno studio effettuato tra gli utenti delle biblioteche accademiche e di ricerca della Slovacchia, che mira a identificare il comportamento degli utenti nella ricerca delle informazioni (preferenza nella scelta delle fonti, punti di accesso etc.) al fine di poter derivare modelli di interazione tra utenti, risorse informative e stili informativi. È stato sviluppato a questo scopo un questionario disegnato su un "modello concettuale" nel quale vengono richiesti dati sull'utente (età, sesso, lingua, disciplina etc.) e dati sul loro comportamento di ricerca. Dalla prima analisi dei dati emerge che gli utenti considerano la ricerca delle informazioni come un'attività individuale. Tuttavia i risultati evidenziano che essi necessitano di formazione e supporto da parte del bibliotecario per imparare le strategie di ricerca, per la formulazione delle interrogazioni, per una panoramica sull'ampio spettro di fonti informative disponibili.

[7] **Gemma Madle** ha presentato una nuova metodologia per valutare le biblioteche digitali e l'impatto delle informazioni mediche contenute nei siti Web sull'utente. Questo metodo si basa non su dati tecnici di usabilità bensì sulla valutazione dell'aumento di conoscenze che deriva dalla consultazione dei siti stessi. A tal scopo è stato realizzato un sito web divulgativo (nell'ambito della National electronic Library for Health) che si pone l'obiettivo di correggere il comportamento del pubblico nei confronti dei farmaci e delle prescrizioni mediche. Sono stati quindi elaborati due questionari da far compilare all'utente prima e dopo l'uso la navigazione nel sito. I risultati hanno dimostrato le potenzialità delle risorse digitali nel modificare il livello di conoscenze e l'attitudine degli utenti.

[8] **Martin Halbert** ha illustrato i principali risultati raggiunti da 4 progetti americani finanziati dalla Andrew W. Mellon Foundation nel 2001 per sperimentare l'applicazione del *metadata harvesting* utilizzando l'OAI-PMH (Open Archive Protocol for Metadata Harvesting).

Le organizzazioni coinvolte in questa sperimentazione sono:

- l'Università dell'Illinois che ha costruito un portale Web dedicato all'eredità culturale (Cultural Heritage)
- l'Università del Michigan che ha sviluppato un servizio chiamato OAIster (<http://www.oaister.org/>) che permette agli utenti di accedere non solo ai metadati ma anche alla risorsa digitale vera e propria
- l'Università Emory di Atlanta (progetto MetaArchive) che ha sperimentato il protocollo su collezioni disciplinari in biblioteche di piccole-medie dimensioni
- l'organizzazione non-profit SOLINET di Atlanta che ha progettato AmericanSouth.org, un portale accademico per lo studio della storia e della cultura dell'America del Sud. Quest'ultimo progetto, assieme al progetto Metarchive dell'Università di Emory, ha dato vita alla MetaScholar Initiative (<http://metascholar.org/>).

I risultati di questa indagine evidenziano che l'OAI-PMH può essere considerato sostanzialmente un buon protocollo per la velocità e la facilità con cui permette di effettuare l'*harvesting* di metadati da più archivi. La funzione di ricerca tuttavia può essere compromessa dalla inconsistenza dei metadati forniti e dai seguenti ulteriori punti critici:

Metadati diversi: pur utilizzando tutti il Dublin Core, archivi diversi possono applicare schema Dublin Core diversi rendendo necessarie, da parte dei *service provider*, alcune attività di normalizzazione. Ulteriori difficoltà possono sorgere inoltre nell'utilizzo di formati di metadati diversi, ad es. quando metadati espressi in EAD (Encoded Archival Description), un formato utilizzato per le collezioni, vengono convertiti in UDC (Unqualified Dublin Core)

Record duplicati: questa eventualità occorre soprattutto quando vengono presi in considerazione non solo archivi originali (es. ArXiv.org) ma anche aggregatori di archivi (es. CiteBase).

Granularità o specificità degli oggetti digitali: riguarda la copresenza, in uno stesso archivio, di records che descrivono un singolo documento bibliografico - ad es. un libro - con record che descrivono collezioni. Questo potrebbe confondere l'utente che si può aspettare di accedere a tutte le poesie contenute in un libro così come riesce ad accedere ai singoli oggetti (es. manoscritti) di una collezione.

Restrizioni di accesso alle risorse: queste note sono incluse nel DC Rights element che tuttavia manca di una standardizzazione. E' opinione di alcuni inoltre che la presenza di metadati di oggetti ad accesso limitato in un archivio ad accesso aperto possa confondere l'utente che cerca di accedere a quella determinata risorsa

Barriere nell'adozione del protocollo: malgrado molte biblioteche ed archivi abbiano già aderito all'OAI come *data provider*, ad oggi una piccolissima percentuale di questi ha usato, implementato o adottato l'OAI-PMH. La ragione di ciò andrebbe ricercata non tanto in una mancanza di interesse quanto in una mancanza di fondi di cui soffrono tali istituzioni. Secondo gli autori quindi andrebbe incoraggiata l'adozione di un programma con finanziamenti centrali per la diffusione di questa pratica.

La ricetta elaborata dalla Metascholar Initiative, per una migliore efficacia dei servizi basati sul *metadata harvesting*, si chiama "Metadata Gardening". Con questo termine si intende il coordinamento dell'intero ciclo di *production, dissemination, reprocessing and extended uses* dei metadati. Uno degli obiettivi di questo progetto globale è anche quello di vedere se possono esistere sistemi di ricerca federati che possono interrogare simultaneamente sia collezioni di metadati OAI che motori di ricerca Web.

[9] Il lavoro di **Alexander Faaborg e Carl Lagoze** si inserisce nell'ambito del Semantic Web, iniziativa del W3C nata per facilitare l'organizzazione e la classificazione delle risorse web in un modo leggibile dalle macchine e dagli esseri umani e che ha prodotto l'RDF (Resource Description Framework). L'apporto degli autori consiste nella implementazione di due applicazioni software: 1) Site Annotator, una interfaccia utente che permette all'utente di creare al di sopra della pagina web, un livello semantico personalizzato con annotazioni che utilizzano RDF; 2) Web Task Pane, un browser Web che modifica Internet Explorer permettendo agli utenti di navigare tra scaffali virtuali organizzati

per concetti semantici creati dall'utente. Il lavoro si basa sul concetto che l'utilità del web come risorsa informativa può essere arricchita se gli utenti stessi possono mettere in relazione dei documenti e annotarli. Questo sviluppo trova forti potenzialità di applicazione nel campo della didattica. Gli strumenti qui presentati affiancano il risultato di altri studi sviluppati negli ultimi anni e che vengono descritti al sito <http://annotation.semanticweb.org/tools>.

[10] **Maristella Agosti and Nicola Ferro** hanno presentato uno studio sulle Annotazioni condotto in seno al progetto di ricerca ECD (Enhanced Content Delivery) lanciato dal CNR italiano nel 2002 ed al quale hanno aderito come partner università, centri di ricerca, aziende italiane e l'ICCU. L'obiettivo generale di questo progetto è quello di sviluppare metodi e tecnologie in grado di offrire agli utenti finali contenuti arricchiti, cioè materiali digitali trasformati e organizzati resi disponibili attraverso servizi di biblioteca digitale. ECD prende come riferimento il Digital Library Service System OpenDLib, un *software toolkit* sviluppato dall'Istituto di Scienze e Tecnologia dell'Informazione del CNR di Pisa, che permette di progettare servizi di biblioteca digitale sulla base dei requisiti di una data comunità di utenti. Il lavoro di Agosti e Ferro ha previsto da una parte l'analisi della semantica del concetto di annotazione e, dall'altra, lo sviluppo di un prototipo di servizio di annotazione, chiamato Annotation Manager, che estende la biblioteca digitale OpenDLib.

[11] L'analisi condotta da **Frank Shipman et al.** su un campione di studenti in legge, ha l'obiettivo di comprendere la relazione tra un determinato tipo di nota e la funzione (evidenziare un passaggio, riassumere un concetto etc.) al fine di poter disegnare un software per ebook in grado di guidare l'accesso alle parti annotate del testo ritenute più importanti. Questo studio ha permesso agli autori di progettare uno "scanner di segni" (*mark parser*) in grado di identificare, classificare e ordinare le annotazioni. Questo sistema è stato integrato in Xlibris, un prototipo di interfaccia per la lettura al computer (*active reading machine*).

[12] I *recommendation services* sono stati invece l'oggetto della presentazione del lavoro di **Andreas Geyer-Schulz**. Si tratta di servizi di ausilio alla ricerca per mezzo di suggerimenti / raccomandazioni (es link a riferimenti bibliografici o bibliografie), che possono ridurre i tempi di ricerca e valutazione dei documenti da parte degli utenti. Questa tipologia di servizi può essere integrata sia nelle biblioteche digitali che nei sistemi informativi automatizzati tradizionali (es. OPAC). Lo studio della letteratura ha evidenziato che la totalità della ricerca effettuata finora si è focalizzata sui *recommendation services* nell'ambito di biblioteche digitali. Il progetto presentato dagli autori invece ha permesso l'integrazione di un *recommender system* nel sistema di automazione in funzione all'Università di Karlsruhe, rendendo disponibile questo servizio (chiamato Empfehlungen – Others Also Use...) dall'OPAC.

[13] Nell'ambito della sessione dedicata alle tecnologie Web cito solo il lavoro di **Tim Brody et al.** dedicato ai servizi digitometri riservati agli archivi di e-prints che utilizzano l'OAI-PMH. Celestial è un software in grado di creare una copia degli archivi OAI agendo come sostituto in caso di inagibilità del data provider originale. Citebase Search, sviluppato nell'ambito del JISC/NSF Open Citation Project conclusosi a dicembre 2002, è un database citazionale in grado di estrarre le citazioni bibliografiche contenute nei documenti depositati negli OA. I servizi qui descritti dimostrano, secondo gli autori, che l'adesione degli autori all'OAI può portare un enorme beneficio alla letteratura scientifica aumentandone la visibilità e quindi l'impatto sulla ricerca.

[14] Il progetto SozioNet, finanziato nel 2002 dal Ministero Federale per l'Istruzione, la Scienza, la Ricerca e la Tecnologia (BMBF), costituisce una parte di un portale nazionale dedicato alle scienze sociali e sviluppato dalla German Infoconnex Initiative. Ispirandosi al successo di altre iniziative quali

MathNet e SOSIG, SozioNet mira a fornire accesso a tutte le risorse relative alle scienze sociali che si trovano disponibili nel Web gratuitamente. Il progetto, presentato da **Wolfgang Meier**, prevede 1) che gli studiosi e le 13 istituzioni accademiche e di ricerca che partecipano al progetto adottino, per creare le loro risorse Web, standard di metadati comuni; 2) l'implementazione di un software (MetaWizard) per supportare la generazione dei metadati 3) la realizzazione di una infrastruttura per l'*harvesting* dei metadati dagli archivi delle istituzioni partecipanti e da risorse Web selezionate e di un modulo per la ricerca specializzata. Il linguaggio scelto per esprimere i metadati è RDF, esteso da DAML (DARPA Agent Markup Language). DAML verrà presto superato dal Web Ontology Language (OWL) del W3C disponibile al momento sotto forma di bozza di lavoro (<http://www.w3.org/TR/owl-ref/>).

[15] Il nuovissimo portale tedesco per l'informazione scientifica Vascoda è stato presentato da **Tamara Pianos**. Tra i servizi, Vascoda presenta un motore di ricerca che permette la ricerca in collezioni multidisciplinari distribuite di materiale digitale e digitalizzato, mirando ad essere un *one stop shopping* dell'informazione scientifica. A differenza di Google, Vascoda indicizzerà solo materiale di interesse accademico selezionato da esperti. Vengono presi in considerazione full text, banche dati, motori di ricerca disciplinari. Vascoda, a cui partecipano più di 30 partner tra biblioteche e centri informativi, è finanziato dal Ministero Federale per l'Istruzione, la Scienza, la Ricerca e la Tecnologia (BMBF) e dalla Società tedesca per la ricerca (DFG). Nasce dalla cooperazione del Catalogo dei periodici elettronici tedesco (EZB) con la Virtual Library e l'Information Alliance e andrà a costituire un nucleo della nascente Biblioteca Digitale Tedesca.. La prima *release* (agosto 2003) includerà solo funzioni di ricerca semplice (tecnologia HTTP e SOAP) mentre con la seconda *release* saranno possibili ricerche avanzate e opzioni di *browsing*. L'accesso per soggetto è uno dei servizi chiave dei portali pertanto uno dei servizi di Vascoda sarà quello di permettere il *browsing* per soggetto tra le collezioni distribuite di metadati grazie a una mappatura dei diversi sistemi di classificazione locali in un sistema di classificazione comune (DDC).

[16] **Lilian Cassel** ha illustrato 5Sgraph, uno strumento grafico innovativo che permette anche a utenti non esperti di programmazione e informatica (es. i "bibliotecari digitali") di costruire biblioteche digitali senza conoscere i fondamenti teorici e la sintassi dei linguaggi descrittivi. 5Sgraph si basa su un metamodello che descrive le biblioteche digitali utilizzando la teoria delle 5S (*Streams, Structures, Spaces, Scenarios, Societies*), elaborata da Gonçalves, Fox *et al.* per la descrizione delle biblioteche digitali, ed utilizza come linguaggio di descrizione il 5SL (*5S description language*).