

Zpráva ze zahraniční služební cesty

Jméno účastníka cesty	Bedřich Vychodil, Ladislav Cubr
Pracoviště – instituce, adresa	Odbor dlouhodobé ochrany digitálních dat 1.5
Pracoviště – zařazení	ODO
Důvod cesty	návštěva 14. konference ECDL 2010
Místo – město	Glasgow
Místo – země	Velká Británie
Datum (od-do)	5.9.-11. 9. 2010
Podrobný časový harmonogram	Bedřich Vychodil, Ladislav Cubr: 5.9 odlet Praha - Londýn, Londýn - Glasgow 6.9 Tutoriály 7.9 Hlavní konference 9-10.9 Workshopy Přílet: Glasgow – Londýn, Londýn - Praha
Spolucestující z NK	Ladislav Cubr – 14. konference ECDL 2010
Finanční zajištění	Bedřich Vychodil: IOP/NDK Cubr Ladislav: IOP / NDK
Cíle cesty	Účast na konferenci, workshopu, návštěva University of Glasgow, konzultace s odborníky
Plnění cílů cesty	úkoly splněny – viz níže
Program a další podrobnější informace	Sborník z konference - Ano, informace viz níže
Přivezené materiály	NA
Tištěné přílohy a elektronické dokumenty	
Datum předložení zprávy	6.10.2010
Podpis předkladatele zprávy	

Cestovní zpráva

Výše uvedení pracovníci Národní knihovny ČR se zúčastnili 14. evropské konference věnované výzkumu a vývoji v oblasti digitálních knihoven – „ECDL“ (The European Conference on Research and Advanced Technology for Digital Libraries), která se konala 6.-10. září 2010 v Glasgow ve Skotsku ve Velké Británii na významné místní univerzitě (Glasgow University). Jednalo se o poslední ročník konference pod tímto názvem, následný ročník se bude konat pod názvem TPDL (Theory and Practice of Digital Libraries). Důvodem je kolize názvu s evropským certifikačním programem počítačové gramotnosti, který je znám též po zkratku „ECDL“ (European Computer Driving Licence) a který je mezinárodně známější.

Na konferenci bylo představeno úhrnně několik desítek různých typů příspěvků (zejména prezentací, posterů a workshopů) věnovaných zejména současným výzkumným a vývojovým aktivitám v oblasti digitálních knihoven (okruhy jako jsou webové aplikace, systémová architektura, interoperabilita, metadatová řešení, ontologie nebo techniky vyhledávání informací), a dále též v příbuzných oblastech (digitalizace, digitální archivace).

Níže uvádíme přehled nejdůležitějších příspěvků z hlediska zaměření výše uvedených pracovníků a příbuzných oblastí ohodnocených jako relevantních.

Relevantní příspěvky

Susan Dumais z výzkumného projektu společnosti Microsoft se na úvod konference (hlavní zahajovací příspěvek) zabývala v příspěvku „Understanding and Supporting People in Interacting with Dynamic Information Environments“ problematikou rychle se měnících webových stránek (podle jejího výzkumu se změní u třetiny webových stránek obsah průměrně v rozmezí 11 týdnů). Běžné vyhledávače naleznou obvykle pouze to, co se na webu nalézá v současnosti, minulý (změněný) webový obsah je ztracen. V návaznosti na tato fakta vyvíjí Microsoft nástroj „DiffIE plug-in“, který při vyhledávání označí (barevně) veškeré informace, které se změnilly od doby, kdy jsme navštívili stránky naposledy. Bylo zjištěno, že uživatelé využívající tento plug-in znovu navštíví větší množství stránek, než by navštívili bez něj. Tato skutečnost je interpretována tak, že lidé si díky nástroji více uvědomují proměnlivost obsahu, a ten jim tak napomáhá k lepšímu vstřebávání informací.

Stephan Strodl, Petar Petrov, Michael Greifeneder a Andreas Rauber, výzkumníci známí též našemu prostředí z evropského projektu DPE, jehož partnerem byla také Národní knihovna ČR představili v příspěvku „Automating logical preservation for small institutions with Hoppla“ systém pro digitální archivaci určený menším knihovnám nebo podobným informačním institucím pro potřeby dlouhodobé ochrany spravovaných digitálních dokumentů. Jako každý reálný systém pro digitální archivaci, tak i Hoppla nabízí tzv. logickou ochranu (ochrana před riziky plynoucími ze zastarávání nebo nevhodnosti formátů souborů), tedy komplexnější řešení než je pouhá bitová ochrana. Nástroj byl testován a obsahuje migrační nástroj pro zajištění některých základních migrací. Obsahuje též nástroj pro vytváření dokumentů ve formátu PDF/A, tedy podtypu formátu od firmy Adobe, který dostojí potřebám dlouhodobé archivace, Java Open Document (<http://www.artofsolving.com/opensource/jodconverter>).

Alejandro Bia, Rafael Muñoz a Jaime Gómez v příspěvku „Estimating Digitization Costs in Digital Libraries Using DiCoMo“ představili nástroj určený pro vyhodnocení finančních nákladů spojených s digitalizačními procesy za účelem jejich ekonomicky schůdnější optimalizace. Základní tezí je myšlenka, že dobrý odhad znamená časový nebo cenový cíl, v rámci něhož může být úkol splněn za stanoveného přiměřeného časového tlaku se současně dostatečně dobrou kvalitou výsledku. Nástroj pro odhad nákladů na digitalizaci je založen na existujících modelech, statistice reálného sledování digitalizace a softwarovém řešení.

Carlo Meghini, Nicolas Spyrtos a Tsuyoshi Sugibuchi představili v příspěvku „Modelling Digital Libraries Based on Logic“ specifický datový model vyvíjený pro digitální knihovnu Europeana, který je přepracováním původního modelu.

V panelu nazvaném „Pressure to manage and share data“ se debatující soustředili na problematiku sdílení vědeckých dat. Byl ustanoven konsensus na tom, že je třeba vytvořit dlouhodobý plán pro řízení a sdílení ve vědeckém sektoru, protože vědecká oblast je velmi citlivá na otázky dostupnosti dat (publikování vědeckých výstupů apod.). Odborníci z různých institucí prezentovali svoje zkušenosti a potíže s dlouhodobým plánováním, mezi nimi nedostatečné vyjasnění etických pravidel, otázek práv a vlastnictví, strach z chyb nalezených uživateli, strach z vykrádání prací jiných, nedostatečná dokumentace nebo nedostatečné podněty finanční a další povahy. Účastníci konstatovali, že je třeba vytvořit a rozvíjet služby a podporu pro správu vědeckých dat, online průvodce, školení nebo univerzitní systémy pravidel („policies“), která mj. musí vyjasnit otázky vlastnictví dat a odpovědností vážících se k datům, řízení nebo formálního postupu pro přesunech dat. Robin Rice přednesl tezi, že vzniká nová profese – datový specialista („data professional“, ekvivalenty: data scientist, data architect, data steward, data librarian, research data specialist), který bude řešit výše uvedené problémy.

Claudio Gennaro, Giuseppe Amato, Paolo Bolettieri a Pasquale Savino v příspěvku „An approach to Content-Based Image Retrieval based on the Lucene search engine library“ představili nástroj „Lucignolo“ určený pro vyhledávání obrazových dat (<http://lucignolo.isti.cnr.it/cophirUI>). Nástroj je založen na opensourcovém vyhledávacím nástroji z projektu Lucene (<http://lucene.apache.org/>).

Mezi nejvýznamnější postery patřil poster prezentující výsledky projektu PLANETS, konkrétně „Planets Testbed“, testovací prostředí pro potřeby digitální archivace (<http://testbed.planets-project.eu/testbed/>).

Dalším významným posterem byl „Mopseus – A Digital Library Management System Focused on Preservation“, připravený řeckými vědkyněmi (Dimitris Gavrili, Christos Papatheodorou, Panos Constantopoulos, Stavros Angelis). Mopseus je archivační systém vystavený na Fedoře, který podporuje všech šest základních funkčních entit funkčního modelu podle OAIS. Vývojová stránka projektu je na <http://www.dcu.gr/index.php?p=projects&lang=en§ion=8>.

Podrobný rozpis hlavních konferenčních příspěvků (Program konference)

Monday, September 6			
08:00 to afternoon	Registration <i>BO Concourse Area</i>		
09:00 to 10:30	Tutorial 1: Exploring Perspectives on the Evaluation of Digital Libraries <i>SAW Room 422</i>	Tutorial 3: Memento and Open Annotation <i>SAW Room 423</i>	Doctoral Consortium <i>SAW Level 5</i>
11:00 to 12:30	Tutorial 1 (continued)	Tutorial 3 (continued)	Doctoral Consortium
13:30 to 15:00	Tutorial 2: Introduction to (Teaching/Learning about) Digital Libraries <i>SAW Room 422</i>	Tutorial 4: Multimedia Document Access <i>HATII Room 401</i>	Doctoral Consortium
15:30 to 17:00	Tutorial 2 (continued)	Tutorial 4 (continued)	Doctoral Consortium
from 15:30 onwards	Excursion to Auchentoshan Distillery		
Tuesday, September 7			
08:30 to	Registration		

Tuesday, September 7		
09:30	<i>BO Concourse Area</i>	
09:45 to 10:45	Keynote 1: Susan Dumais Understanding and Supporting People in Interacting with Dynamic Information Environments <i>Chair: Fabrizio Sebastiani</i> <i>BO Theatre 1</i>	
11:15 to 12:25	Session 1a: System Architectures (70mn) <i>Chair: Stefan Gradmann</i> <i>BO Theatre 1</i>	Session 1b: Metadata (60mn) <i>Chair: Traugott Koch</i> <i>BO Theatre A</i>
	Modelling Digital Libraries Based on Logic (L) <i>Carlo Meghini, Nicolas Spyrtos, Tsuyoshi Sugibuchi</i>	Uncovering Hidden Qualities - Benefits of Quality Measures for Automatically Generated Metadata (S) <i>Sascha Tönnies, Wolf-Tilo Balke</i>
	General-Purpose Digital Library Content Laboratory Systems (S) <i>Paolo Manghi, Marko Mikulicic, Leonardo Candela, Michele Artini, Alessia Bardi</i>	Query transformation in a CIDOC CRM based cultural metadata integration environment (S) <i>Manolis Gergatsoulis, Lina Bountouri, Panorea Gaitanou, Christos Papatheodorou</i>
	Component-based Authoring of Complex, Petri net-based Digital Library Infrastructure (S) <i>Yungah Park, Unmil Karadkar, Richard Furuta</i>	User-Contributed Descriptive Metadata for Libraries and Cultural Institutions (S) <i>Michael A. Zarro, Robert B. Allen</i>
14:00 to 15:30	Panel: Developing Services to Support Research Data Management and Sharing <i>Chair: Liz Lyon</i> <i>BO Theatre 1</i> Veerle van den Eynden / Sarah Jones Robin Rice / Rob Grim	
16:00 to 17:30	Session 2a: Multimedia Information Retrieval (90mn) <i>Chair: Carlo Meghini</i> <i>BO Theatre 1</i>	Session 2b: Interaction and Interactivity (80mn) <i>Chair: Maristella Agosti</i> <i>BO Theatre A</i>
	An approach to Content-Based Image Retrieval based on the Lucene search engine library (L) <i>Claudio Gennaro, Giuseppe Amato, Paolo Bolettieri, Pasquale Savino</i>	An Architecture for Supporting RFID-Enhanced Interactions in Digital Libraries (L) <i>George Buchanan, Jennifer Pearson</i>
	Evaluation Constructs for Visual Video Summaries (L) <i>Stina Westman</i>	New Evidence on the Interoperability of Information Systems within UK Higher Education Institutions (L) <i>Kathleen Menzies, Duncan Birrell, Gordon Dunsire</i>
	Visual Expression for Organizing and Accessing Music Collections in MusicWiz (L) <i>Konstantinos Meintanis, Frank M. Shipman</i>	Enhancing Digital Libraries with Social Navigation: The Case of Ensemble (S) <i>Peter Brusilovsky, Lillian Cassel, Lois Delcambre, Edward Fox, Richard Furuta, Daniel D. Garcia, Frank M. Shipman, Paul Bogen, Michael Yudelson</i>

Wednesday, September 8	
08:30 to 09:00	Registration <i>BO Concourse Area</i>
09:00 to 10:00	Keynote 2: Liina Munari Digital libraries - European perspectives and initiatives <i>Chair: Joemon Jose</i>

Wednesday, September 8		
	<i>BO Theatre 1</i>	
10:30 to 11:50	Session 3a: Digital Preservation (80mn) <i>Chair: Ingeborg Solvberg</i> <i>BO Theatre 1</i>	Session 3b: Social Web / Web 2.0 (70mn) <i>Chair: Ingo Frommholz</i> <i>BO Theatre A</i>
	Automating logical preservation for small institutions with Hoppla (L) <i>Stephan Strodl, Petar Petrov, Michael Greifeneder, Andreas Rauber</i>	Privacy-Aware Folksonomies (L) <i>Clemens Heidinger, Erik Buchmann, Matthias Huber, Klemens Böhm, Jörn Müller-Quade</i>
	Estimating Digitization Costs in Digital Libraries Using DiCoMo (L) <i>Alejandro Bia, Rafael Muñoz, Jaime Gómez</i>	Seamless web editing for personal and public access digital libraries (S) <i>David Bainbridge, Brook J. Novak</i>
	In Pursuit of an Expressive Vocabulary for Preserved New Media Art (S) <i>Andrew McHugh, Leo Konstantelos</i>	Automatic classification of social tags (S) <i>Christian Wartena</i>
11:50 to 12:40	Poster Spotlight Session <i>BO Theatre 1</i>	
14:00 to 16:00	Poster and Demo session <i>BO Concourse Area</i>	
16:30 to 17:40	Session 4a: Search in Digital Libraries (70mn) <i>Chair: Elaine Toms</i> <i>BO Theatre 1</i>	Session 4b: (Meta) Analysis of Digital Libraries (60mn) <i>Chair: Claus Peter Klas</i> <i>BO Theatre A</i>
	Exploring the Impact of Search Interface Features on Search Tasks (L) <i>Abdigani Diriye, Ann Blandford, Anastasios Tombros</i>	An Analysis of the Evolving Coverage of Computer Science Sub-Fields in the DBLP Digital Library (L) <i>Florian Reitz, Oliver Hoffmann</i>
	Relevance in Technicolor (L) <i>Ulises Cerviño Beresi, Yunhyong Kim, Mark Baillie, Ian Ruthven, Dawei Song</i>	Analysis of Computer Science Communities Based on DBLP (S) <i>Maria Biryukov, Cailing Dong</i>
	Applications of Session Analysis to Search Interface Design (S) <i>Cathal Hoare, Humphrey Sorensen</i>	Citation graph based ranking in Invenio (L) <i>Ludmila Marian, Jean-Yves Le Meur, Martin Rajman, Martin Vesely</i>
Thursday, September 9		
08:00 to 09:00	Registration <i>BO Concourse Area</i>	
09:00 to 10:20	Session 5a: Query Log Analysis (80mn) <i>Chair: Ed Fox</i> <i>BO Theatre 1</i>	Session 5b: Cooperative Work in Digital Libraries (80mn) <i>Chair: Monica Landoni</i> <i>BO Theatre A</i>
	A Search Log-Based Approach to Evaluation (L) <i>Junte Zhang, Jaap Kamps</i>	Examining Group Work: Implications for the Digital Library as Sharium (L) <i>Sandra Toze, Elaine G. Toms</i>
	Determining Time of Queries for Re-ranking Search Results (L) <i>Nattiya Kanhabua, Kjetil Nørnvåg</i>	Architecture for a Collaborative Research Environment based on Reading List Sharing (L) <i>Gabriella Kazai, Paolo Manghi, Katerina Iatropoulou, Tim Haughton, Marko Mikulicic, Antonis Lempesis, Natasa Milic-Frayling, Natalia Manola</i>
	Ranking Entities Using Web Search Query Logs (S) <i>Bodo Billerbeck, Gianluca Demartini,</i>	CritSpace: A Workspace for Critical Engagement within Cultural Heritage Digital Libraries (S)

Thursday, September 9			
	<i>Claudiu S. Firan, Tereza Iofciu, Ralf Krestel</i>	<i>Neal Audenaert, Lucchese George, Richard Furuta</i>	
11:00 to 12:30	Session 6a: Ontologies (90mn) <i>Chair: Andrew McHugh</i> <i>BO Theatre 1</i>	Session 6b: Domain-specific Digital Libraries (80mn) <i>Chair: Jose Borbinha</i> <i>BO Theatre A</i>	
	German Encyclopedia Alignment Based on Information Retrieval Techniques (L) <i>Roman Kern, Michael Granitzer</i>	A Visual Digital Library Approach for Time-Oriented Scientific Primary Data (L) <i>Jürgen Bernard, Jan Brase, Dieter Fellner, Oliver Koepler, Jörn Kohlhammer, Tobias Ruppert, Tobias Schreck, Irina Sens</i>	
	Lightweight Parsing of Classifications into Lightweight Ontologies (L) <i>Aliaksandr Autayeu, Fausto Giunchiglia, Pierre Andrews</i>	DINAH, a philological platform for the construction of multi-structured documents (L) <i>Pierre-Edouard Portier, Sylvie Calabretto</i>	
	Measuring Effectiveness of Geographic IR Systems in Digital Libraries (L) <i>Damien Palacio, Guillaume Cabanac, Christian Sallaberry, Gilles Hubert</i>	The PROBADO Project - Approach and Lessons Learned in Building a Digital Library System for Heterogeneous Non-textual Documents (S) <i>René Berndt, Ina Blümel, Michael Clausen, David Damm, Jürgen Diet, Dieter Fellner, Christian Fremerey, Reinhard Klein, Frank Krahl, Maximilian Scherer, Tobias Schreck, Irina Sens, Verena Thomas, Raoul Wessel</i>	
14:30 to 16:00	Workshop 1: Making Digital Libraries Interoperable: Challenges and Approaches <i>BO Theatre C</i>	Workshop 2: Networked Knowledge Organisation Systems and Services <i>BO Theatre A</i>	
16:30 to 18:30	Workshop 1 (continued)	Workshop 2 (continued)	
Friday, September 10			
08:00 to 09:00	Registration <i>BO Concourse Area</i>		
09:00 to 10:30	Workshop 1: Making Digital Libraries Interoperable: Challenges and Approaches <i>BO Theatre C</i>	Workshop 2: Networked Knowledge Organisation Systems and Services <i>BO Theatre A</i>	Workshop 3: Very Large Digital Libraries <i>SAW Level 5</i>
11:00 to 13:00	Workshop 1 (continued)	Workshop 2 (continued, to be finished by 13:15)	Workshop 3 (continued)
14:00 to 16:00	Workshop 1 (continued)		Workshop 3 (continued)