# Table of Contents

## Section: Preservation strategies

When Should I Make Preservation Copies of Myself?  
*Chuck Cartledge, Michael Nelson*  
Page 1

An Argument for Archiving Facebook as a Heterogeneous Personal Store  
*Catherine Marshall, Frank Shipman*  
Page 11

Implementing Digital Preservation Strategy: Developing Content Collection Profiles at the British Library  
*Michael Day, Ann MacDonald, Akiko Kimura, Maureen Pennock*  
Page 21

The Archival Acid Test: Evaluating Archive Performance on Advanced HTML and JavaScript  
*Mat Kelly, Michael L. Nelson, Michele C. Weigle*  
Page 25

## Section: Recommendation and indexing

Recommendation Based on Deduced Social Networks in an Educational Digital Library  
*Monika Akbar, Clifford Shaffer, Weiguo Fan, Edward Fox*  
Page 29

Dynamic Taxonomy Composition via Keyqueries  
*Tim Gollub, Matthias Hagen, Michael Völske, Benno Stein*  
Page 39

Personalized PageRank for Making Recommendations in Digital Cultural Heritage Collections  
*Arantxa Otegi, Eneko Agirre, Paul Clough*  
Page 49

Making Research Data Findable in Digital Libraries: A Layered Model for User-Oriented Indexing of Survey Data  
*Tanja Friedrich, Andreas Oskar Kempf*  
Page 53

## Session: Publication impacts

Characterizing Scholar Popularity: A Case Study in the Computer Science Research Community  
*Glauber Dias Gonçalves, Flavio Vinicius Diniz de Figueiredo, Jussara Marques de Almeida, Marcos Andre Goncalves*  
Page 57

Community-based Endogamy as an Influence Indicator  
*Thiago H. P. Silva, Mirella M. Moro, Ana Paula C. Silva, Wagner Meira Jr., Alberto Laender*  
Page 67

Disambiguating Publication Venue Titles using Association Rules  
*Denilson Pereira, Eduardo Silva, Ahmed Esmeein*  
Page 77

## Session: Personal DL Design

From User Needs to Opportunities in Personal Information Management:  
A Case Study on Organisational Strategies in Cross-Media Information Spaces  
*Sandra Trullemans, Beat Signer*  
Page 87

PerCon: A Personal Digital Library for Heterogeneous Data  
*Su Inn Park, Frank Shipman*  
Page 97

Social Information Behaviour in Physical Libraries: Implications for the design of digital libraries  
*Annika Hinze, Claire Timpany, Nicholas Vanderschatz, Hayat Alqurashi*  
Page 107
Session: Building systems
Towards Building a Scholarly Big Data Platform: Challenges, Lessons and Opportunities
117
Bridging the Gap Between Real World Repositories and Scalable Preservation Environments
Bolette Ammitzbøll Jurik, Asger Askov Blekinge, Rune Bruun Ferneke-Nielsen, Per Moldrup-Dalum
127
Using ACM DL paper metadata as an auxiliary source for building educational collections
Yinlin Chen, Edward Fox
137
Crowd-sourcing Web Knowledge for Metadata Extraction,
Zhaohui Wu, Wenyi Huang, Chen Liang, C. Lee Giles
141

Session: Browsing and searching
Lend me some sugar: Borrowing rates of neighbouring books as evidence for browsing
Dana McKay, Wally Smith, Shanton Chang
145
Improving the visibility of geospatial data on the Web
Javier Lacasta, Francisco Javier López-Pellicer, Walter Renteria-Agualimpia, Javier Nogueras-Iso
155
The Anatomy of a Search and Mining System for Digital Humanities
Martyn Harris, Mark Levene, Dell Zhang, Dan Levene
165
Increasing the visibility of library records via a consortial search engine
Ônne Mets, Silvia Gstrein, Veronika Gründhammer
169

Session: Item identification
Combining Domain-Specific Heuristics for Author Name Disambiguation
Alan Filipe Santana, Marcos André Gonçalves, Anderson A. Ferreira, Alberto Laender
173
Detecting and Modeling Local Text Reuse
David A. Smith, Ryan Cordell, Elizabeth Maddock Dillon, John Wilkerson, Nick Stramp
183
Identifying the Same Records across multiple Ukiyo-e Image Databases Using Textual Data in Different Languages
Bilgisaikhan Batjargal, Takeo Kuyama, Fuminori Kimura, Akira Maeda
193
Reducing Computational Effort for Plagiarism Detection by using Citation Characteristics to Limit Retrieval Space
Norman Meuschke, Bela Gipp
197

Session: Quality Data and Metadata
Quality Assessment of Collaborative Content With Minimal Information
Daniel Hasan Dalip, Harlley Lima, Marcos Gonçalves, Marco Cristo, Pável Calado
201
Fast Image-based Chinese Calligraphic Character Retrieval on Large Scale Data
Pengcheng Gao, Jiangqin Wu, Yuan Lin, Yang Xia, Tianjiao Mao, Baogang Wei
211
Human and Machine Error Analysis on Dependency Parsing of Ancient Greek Texts
Saeed Majidi, Gregory Crane
221
The Feasibility of Investing of Manual Correction of Metadata for a Large-Scale Digital Library
225
### Session: Topics, evolution, and relationships

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Framework for Analyzing Semantic Change of Words across Time</td>
<td>229</td>
</tr>
<tr>
<td>Adam Jatowt, Kevin Duh</td>
<td></td>
</tr>
<tr>
<td>Representing Topics Labels for Exploring Digital Libraries</td>
<td>239</td>
</tr>
<tr>
<td>Nikolaos Aletras, Timothy Baldwin, Jey Han Lau, Mark Stevenson</td>
<td></td>
</tr>
<tr>
<td>Topical Establishment Leveraging Literature Evolution</td>
<td>249</td>
</tr>
<tr>
<td>Han Xu, Eric Martin, Ashesh Mahidadia</td>
<td></td>
</tr>
<tr>
<td>A Method to Support Analysis of Personal Relationship through Place Names</td>
<td>253</td>
</tr>
<tr>
<td>Fuminori Kimura, Akira Maeda</td>
<td></td>
</tr>
</tbody>
</table>

### Session: Knowledge Infrastructure and Repositories

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management</td>
<td>257</td>
</tr>
<tr>
<td>Christine L. Borgman, Peter T. Darch, Ashley E. Sands, Jillian C. Wallis, Sharon Traveek</td>
<td></td>
</tr>
<tr>
<td>CED2AR: The Comprehensive Extensible Data Documentation and Access Repository</td>
<td>267</td>
</tr>
<tr>
<td>Carl Lagoze, Lars Vilhuber, Jeremy Williams, Benjamin Perry, William C. Block</td>
<td></td>
</tr>
<tr>
<td>Big Brother is Watching You -- But in a Good Way</td>
<td>277</td>
</tr>
<tr>
<td>Carlin St Pierre, David Bainbridge, Bill Rogers</td>
<td></td>
</tr>
<tr>
<td>A comparative analysis of the HSS &amp; HEP data submission workflows</td>
<td>281</td>
</tr>
<tr>
<td>Suenje Dallmeier Tiessen, Artemis Lavasa, Laura Rueda, Patricia Heterich, Rachael Kotarski, Elizabeth Newbold</td>
<td></td>
</tr>
<tr>
<td>An Open Cultural Digital Content Infrastructure</td>
<td>285</td>
</tr>
<tr>
<td>Ioanna Ourania Stathopoulou, Haris Georgiadis, Vangelis Banos, Panagiotis Stathopoulou, Nikos Houssos, Evi Sachini</td>
<td></td>
</tr>
</tbody>
</table>

### Session: Data transformation and description

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Converting the Zeri photo archive in Linked Open Data: formalizing the conceptual model</td>
<td>289</td>
</tr>
<tr>
<td>Ciro Gonano, Francesca Mambelli, Silvio Peroni, Francesca Tomasi, Fabio Vitali</td>
<td></td>
</tr>
<tr>
<td>Comic2CEBX: A System for Automatic Comic Content Adaptation</td>
<td>299</td>
</tr>
<tr>
<td>Luyuan Li, Yongtao Wang, Liangcai Gao, Zhi Tang, Ching Y. Suen</td>
<td></td>
</tr>
<tr>
<td>Explorations in Linked Data practice for early music corpora</td>
<td>309</td>
</tr>
<tr>
<td>Tim Crawford, Ben Fields, David Lewis, Kevin R. Page</td>
<td></td>
</tr>
<tr>
<td>Creating lightweight ontologies for dataset description: Practical applications in a cross-domain research data management workflow</td>
<td>313</td>
</tr>
<tr>
<td>João Aguiar Castro, João Rocha Da Silva, Cristina Ribeiro</td>
<td></td>
</tr>
<tr>
<td>A preliminary evaluation of HathiTrust metadata: Assessing the sufficiency of legacy records</td>
<td>317</td>
</tr>
<tr>
<td>Katrina Fenlon, Colleen Fallaw, Timothy W. Cole, Myung-Ja Han</td>
<td></td>
</tr>
</tbody>
</table>

### Session: Web archives and memory

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not All Mementos Are Created Equal: Measuring The Impact Of Missing Resources</td>
<td>321</td>
</tr>
<tr>
<td>Justin F. Brunelle, Mat Kelly, Hany Salahelddeen, Michele C. Weigle, Michael L. Nelson</td>
<td></td>
</tr>
<tr>
<td>Finding Pages on the Unarchived Web</td>
<td>331</td>
</tr>
<tr>
<td>Hugo Huurdeman, Anat Ben-David, Jaap Kamps, Thaer Samar, Arjen P. de Vries</td>
<td></td>
</tr>
<tr>
<td>What Triggers Human Remembering of Events? A Large-Scale Analysis of Catalysts for Collective Memory in Wikipedia</td>
<td>341</td>
</tr>
</tbody>
</table>
Session: Citation, citation, citation
Towards a Stratified Learning Approach to Predict Future Citation Counts
Tanmoy Chakraborty, Suhasnau Kumar, Pawan Goyal, Niloy Ganguly, Animesh Mukherjee
Full-Text based Context-Rich Heterogeneous Network Mining Approach for Citation Recommendation
Xiaozhong Liu, Yingying Yu, Chun Guo, Yizhou Sun, Liangcai Gao
RefSeer: A Citation Recommendation System
Wenyi Huang, Zhaohui Wu, Prasenjit Mitra, C. Lee Giles
Do Altmetrics Follow the Crowd or Does the Crowd Follow Altmetrics?
Hamed Alhoori, Richard Furuta

Session: Education and collaboration
Towards Automatic Identification of Core Concepts in Educational Resources
Md Sultan, Steven Bethard, Tamara Sumner
Using Affective Embodied Agents in Information Literacy Education
Yanru Guo, Dion Hoe-Lian Goh, Brendan Luyt
Bend Me Shape Me: A Practical Experience of Repurposing Research Data
Dana McKay
Research networks in data repositories
Mark R. Costa, Jian Qin, Jun Wang

Demonstrations
TagTick: A Tool for Annotation Tagging over Solr indexes
Michele Artini, Claudio Atzori, Alessia Bardi, Sandro La Bruzzo, Paolo Manghi
Keeping your Aggregative Infrastructure Under Control
Michele Artini, Claudio Atzori, Paolo Manghi
Epimenides: An Information System offering Automated Reasoning for the Needs of Digital Preservation
Yannis Kargakis, Yannis Tzitzikas
Extraction of Evolution Descriptions from the Web
Helge Holzmann, Thomas Risse
Rexplore: Unveiling the Dynamics of Scholarly Data
Francesco Osborne, Enrico Motta
Explore The Stacks: A System for Exploration in Digital Libraries
Mark Michael Hall
Great War Stories Told by the People - Crowdsourced Cultural Heritage in Digital Museums
Ingo Frommholz, David Graves, Haiming Liu, Ashwin Kumar, Gordon Brady
When Catalogs Collide: A Mashup Up of the Bibliographic Records from New Zealand's National Bibliography and the HathiTrust
Steffan Safey, David Bainbridge
LODE: Linking and Enhancing High Quality RDF Repositories to the Web of Data
Timo Sztyler, Jakob Huber, Jan Noessner, Jaimie Murdock, Colin Allen, Mathias Niepert
The SCAPE preservation lifecycle
Kresimir Duretec, Michael Kraxner, Artur Kulmukhametov, Markus Plangg, Christoph Becker, Luis Faria
SNAC: The Social Networks and Archival Context Project - Towards an Archival Authority Cooperative
Ray Larson, Daniel Pitti, Adrian Turner
### Demonstrations (continued)

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visualized Related Topics (VRT) System for Health Information Retrieval</td>
<td>Sukjin You, Joel Desarmo, Xiangming Mu, Sukwon Lee, Jessica C. Neal</td>
<td>429</td>
</tr>
<tr>
<td>Posters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modeling Abstractions for Dance Digital Libraries</td>
<td>Katerina El Raheb, Yannis Ioannidis</td>
<td>431</td>
</tr>
<tr>
<td>Text touching effects: Why it is difficult to do active reading with a touch-based tablet device</td>
<td>Hirohito Shibata, Kentaro Takano</td>
<td>433</td>
</tr>
<tr>
<td>Vector-Borne Disease Network Digital Library</td>
<td>Michelle Barker, Donald Brower, Natalie Meyers</td>
<td>435</td>
</tr>
<tr>
<td>CKGHV: a Comprehensive Knowledge Graph For History Visualization</td>
<td>Yingzhen Zhu, Xinyi Cao, Yali Biao, Jiangqin Wu</td>
<td>437</td>
</tr>
<tr>
<td>The Value of Risk Management for Data Management in Science and Engineering</td>
<td>Filipe Ferreira, Ricardo Vieira, José Borbinha</td>
<td>439</td>
</tr>
<tr>
<td>Quantifying the State Trials</td>
<td>Tracy Bergstrom, Donald Brower, Natalie Meyers, Sonia Howell, Eric Morgan, Elliott Visconti</td>
<td>441</td>
</tr>
<tr>
<td>On Cloud Deployment of Digital Preservation Environments</td>
<td>Daniel Pop, Marian Neagul, Dana Petcu</td>
<td>443</td>
</tr>
<tr>
<td>Microscopic analysis of document handling while reading paper documents to improve digital reading device</td>
<td>Kentaro Takano, Hirohito Shibata, Junko Ichino, Tomonori Hashiyama, Shun ‘ichi Tano</td>
<td>445</td>
</tr>
<tr>
<td>Amplifying Scientific Paper’s Abstract By Leveraging Data-weighted Reconstruction</td>
<td>Shansong Yang, Weiming Lu, Baogang Wei, Wenjia An</td>
<td>447</td>
</tr>
<tr>
<td>Unified Digital Library and Data Access Framework For Heterogeneous and Federated Computational Epidemiology Datasets</td>
<td>S.M. Shamimul Hasan, Sandeep Gupta, Edward A. Fox, Keith Bisset, Madhav V. Marathe</td>
<td>449</td>
</tr>
<tr>
<td>Large-Scale Extraction and Analysis of Web Link References in Scholarly Articles</td>
<td>Ke Zhou, Richard Tobin, Claire Grover</td>
<td>451</td>
</tr>
<tr>
<td>What Is This Song about Anyway?: Automatic Classification of Aboutness Using User Interpretations, Social Tags, and Lyrics</td>
<td>Kahyun Choi, Jin Ha Lee, J. Stephen Downie</td>
<td>453</td>
</tr>
<tr>
<td>REEL: A Relation Extraction Learning Framework</td>
<td>Pablo Barrio, Gonçalo Simões, Helena Galhardas, Luis Gravano</td>
<td>455</td>
</tr>
<tr>
<td>Linking Thesaurus for the Social Sciences to the Web of Linked Data</td>
<td>Andias Wira Alam, Andreas Oskar Kempf, Benjamin Zapilko</td>
<td>457</td>
</tr>
<tr>
<td>A Context Model for Digital Preservation of Processes and its Application to a Digital Library System</td>
<td>Rudolf Mayer, Andreas Rauber, Gonçalo Antunes</td>
<td>459</td>
</tr>
<tr>
<td>The Organization information integration in the management of a Digital Library System</td>
<td>Angela Di Iorio, Marco Schaerf</td>
<td>461</td>
</tr>
<tr>
<td>Life span of web pages: A survey of 10 million pages collected in 2001</td>
<td>Teru Agata, Yasuake Miyata, Emi Ishida, Atsushi Ikeuchi, Shuichi Ueda</td>
<td>463</td>
</tr>
<tr>
<td>PageRank-based Word Sense Induction within Web Search Results Clustering</td>
<td>Jose Moreno, Gaël Dias</td>
<td>465</td>
</tr>
</tbody>
</table>
Posters (Continued)

Enabling Multilingual Information Access to Digital Collections:  
An Investigation of Metadata Records Translation  
Jiangping Chen, Olajumoke Azogu, Ryan Knudson  

Mink: Integrating the Live and Archived Web Viewing Experience Using  
Web Browsers and Memento  
Mat Kelly, Michael L. Nelson, Michele C. Weigle  

Cross-Cultural Mood Regression for Music Digital Libraries  
Xiao Hu, Yi-Hsuan Yang  

Establishing an Online Access Panel for Interactive Information Retrieval Research  
Dagmar Kern, Peter Mutschke, Philipp Mayr  

Mood Metadata for Video Games and Interactive Media  
Stephanie Rossi, Jin Ha Lee, Rachel Ivy Clarke  

Balancing Factors Affecting Virtual Reference Services: Identified from  
Academic Librarians’ Perspective  
Sukjin You, Joel Desarmo, Xiangming Mu, Alexandra Dimitroff  

Articles, Papers, Chapters, Theses – who wins the Visibility Wars?  
Melius Weideman.  

Exploring Relationships Among Video Games  
Rachel Ivy Clarke, Jin Ha Lee, Jacob Jett, Simone Sacchi  

Where should I publish?--Detecting journal similarity based on what have been  
published there  
Rob Koopman, Shenghui Wang  

Organization Structures of Personal Information: A Comparison on Two Groups of  
Information Workers  
Hong Zhang, Xiao Hu  

A computational approach to understanding and predicting the behavior of educators  
using an online curriculum planning tool  
Ogheneovo Dibie, Keith Maull, Tamara Sumner  

An Approach to Named Entity Extraction from Historical Documents in  
Traditional Mongolian Script  
Biligsaikhan Batjargal, Garmaabazar Khaltarkhuu, Fuminori Kimura, Akira Maeda  

Europeana Cloud: Towards a Shared Cloud Computing Infrastructure for  
European Aggregators  
Pavel Kats, Petr Knoth, Georgios Mamakis, Marcin Mielnicki, Markus Muhr, Marcin  
Werla  

A Shared Language for Building a Dataset of Sensitive Information  
Clare Llewellyn, Laine Ruus, Ros Burnett, Steve Kirkwood, Mark Smith,  
Rocio Von-Jungenfeld  

Page

467
469
471
473
475
477
479
481
483
485
487
489
491
493