Leonardo Candela Yannis Ioannidis
Paolo Manghi
(Editors)

Proceedings of the Fourth Workshop on Very Large Digital Libraries VLDL2011

...on the marriage between Very Large Digital Libraries and Very Large Data Archives...

A Workshop in conjunction with the First Conference of Theory and Practice of Digital Libraries

Organized by the Institute of Information Science and Technology of the National Research Council (ISTI-CNR) Pisa, Italy

Held in Berlin, Germany,
29th of September, 2011

Editors

Leonardo Candela
Istituto di Scienza e Tecnologie dell'Informazione (ISTI),
Consiglio Nazionale delle Ricerche,
Pisa, Italy

Yannis Ioannidis
Department of Informatics,
National and Kapodistrian University of Athens,
Athens, Greece

Paolo Manghi
Istituto di Scienza e Tecnologie dell'Informazione (ISTI),
Consiglio Nazionale delle Ricerche,
Pisa, Italy

Preface

Very Large Digital Library workshop series Todays information society is confronted with the data deluge issue and is requesting for Digital Libraries that are capable to effectively serve broader and challenging scenarios. Information seekers are willing to have immediate and effective access to human knowledge that differently from the past very often is produced and disseminated through heterogeneous sources. Such scenarios call for the development of Very Large Digital Libraries (VLDLs), which are "very large" with respect to various aspects including the number and type of information objects and collections to be made available, the number of users to be served, the number of "systems" to be federated, the degree of heterogeneity to be reconciled. The goal of the Very Large Digital Library workshop series is to provide researchers, practitioners and application developers with a forum fostering a constructive exchange among all key actors in the field of Very Large Digital Libraries.

The Fourth VLDL edition Scholarly communication itself has changed a lot during the last years and is today driven by "open access" and "enhanced publications" having datasets (e.g., images, videos, XML/RDF files, Linked Data, web sites, databases, queries, benchmark files) as first class citizens among the research outcomes, on par of papers. The fourth edition of the Workshop focused on the problems emerging from the marriage of "very-large" Digital Libraries and "very-large" Data Archives, in an attempt to manage, combine and interlink datasets with publications.

Acknowledgements Our sincere gratitude goes to all the people who have directly or indirectly made this event possible. Among these our colleagues at ISTI-CNR, Donatella Castelli, Pasquale Pagano, and Costantino Thanos for their research inspiration, the members of the program committee, who devoted part of their precious time to ensure the success of this workshop, and of course the authors, whose passion and ideas constitute the real fuel of VLDL.

Leonardo Candela, Yannis Ioannidis, Paolo Manghi
Organizers and Editors of the fourth VLDL workshop

Program Committee

Daan Broeder
Max Planck Institute for Psycholinguistics, The Netherlands

Norbert Fuhr
Department of Computer Science
University of Duisburg-Essen, Germany

Kat Hagedorn
University of Michigan Digital Library Production Service, USA

Leonid Andreevich Kalinichenko
Institute of Informatics Problems
Russian Academy of Science, Moscow

Fabrizio Silvestri
Institute of Information Science and Technologies
National Research Council, Italy

Hussein Suleman
Department of Computer Science
University of Cape Town, South Africa

Table of Content