Arno Scharl
Klaus Tochtermann (Eds.)

The Geospatial Web

How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society

Springer

Advanced Information and Knowledge Processing

Series Editors
Professor Lakhmi Jain
Lakhmi.jain@unisa.edu.au
Professor Xindong Wu
xwu@cs.uvm.edu

Also in this series

Gregoris Mentzas, Dimitris Apostolou, Andreas Abecker and Ron Young
Knowledge Asset Management
1-85233-583-1

Michalis Vazirgiannis, Maria Halkidi and Dimitrios Gunopulos
Uncertainty Handling and Quality Assessment in Data Mining
1-85233-655-2

Asunción Gómez-Pérez, Mariano Fernández-López and Oscar Corcho
Ontological Engineering
1-85233-551-3

Arno Scharl (Ed.)
Environmental Online Communication
1-85233-783-4

Shichao Zhang, Chengqi Zhang and Xindong Wu
Knowledge Discovery in Multiple Databases
1-85233-703-6

Jason T.L. Wang, Mohammed J. Zaki, Hannu T.T. Toivonen and Dennis Shasha (Eds.)
Data Mining in Bioinformatics
1-85233-671-4

C.C. Ko, Ben M. Chen and Jianping Chen
Creating Web-based Laboratories
1-85233-837-7

Manuel Graña, Richard Duro, Alicia d'Anjou and Paul P. Wang (Eds.)
Information Processing with Evolutionary Algorithms
1-85233-886-0

Colin Fyfe
Hebbian Learning and Negative Feedback Networks
1-85233-883-0

Yun-Heh Chen-Burger and Dave Robertson
Automating Business Modelling
1-85233-835-0

Dirk Husmeier, Richard Dybowski and Stephen Roberts (Eds.)
Probabilistic Modeling in Bioinformatics and Medical Informatics
1-85233-778-8

Ajith Abraham, Lakhmi Jain and Robert Goldberg (Eds.)
Evolutionary Multiobjective Optimization
1-85233-787-7

K.C. Tan, E.F. Khor and T.H. Lee
Multiobjective Evolutionary Algorithms and Applications
1-85233-836-9

Nikhil R. Pal and Lakhmi Jain (Eds.)
Advanced Techniques in Knowledge Discovery and Data Mining
1-85233-867-9

Amit Konar and Lakhmi Jain
Cognitive Engineering
1-85233-975-6

Miroslav Kárný (Ed.)
Optimized Bayesian Dynamic Advising
1-85233-928-4

Yannis Manolopoulos, Alexandros Nanopoulos, Apostolos N. Papadopoulos
and Yannis Theodoridis
R-Trees: Theory and Applications
1-85233-977-2

Sanghamitra Bandyopadhyay, Ujjwal Maulik, Lawrence B. Holder and Diane J. Cook (Eds.)
Advanced Methods for Knowledge Discovery from Complex Data
1-85233-989-6

Marcus A. Maloof (Ed.)
Machine Learning and Data Mining for Computer Security
1-84628-029-X

Sifeng Liu and Yi Lin
Grey Information
1-85233-995-0

Vasile Palade, Cosmin Danut Bocaniala and Lakhmi Jain (Eds.)
Computational Intelligence in Fault Diagnosis
1-84628-343-4

Mitra Basu and Tin Kam Ho (Eds.)
Data Complexity in Pattern Recognition
1-84628-171-7

Samuel Pierre (Ed.)
E-Learning Networked Environments and Architectures
1-84628-351-5

Arno Scharl and Klaus Tochtermann (Eds.)

The Geospatial Web

How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society

Springer

Prof. Arno Scharl
Prof. Klaus Tochtermann
Know-Center & Graz University of Technology, Graz, Austria

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2007926113

AI&KP ISSN 1610-3947
ISBN-13: 978-1-84628-826-5
e-ISBN-13: 978-1-84628-827-2

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

Springer Science+Business Media
springer.com

Foreword

The most important attribute of geospatial platforms is their unique potential to aggregate a multitude of public and private geographic data sets, providing access to data from government agencies, industry and the general public. NASA and other organizations have a wealth of planetary science data – representing the output from thousands of satellites in earth-orbit, and from dozens of costly missions to other planets. Benefits derived from both the data and visual interfaces to access the data represent a significant return on investment for the public. Integrating geospatial data with semantic and collaborative Web technology multiplies the public benefits and represents the main focus of this book.

The user interfaces of geobrowsers are designed for the layperson, giving convenient access to all kinds of geographically referenced information. Geobrowsers hide the technical details related to finding, accessing and retrieving such information. The daunting challenge of the Geospatial Web is to seamlessly integrate and display vastly different information modes. Nowadays, it is not enough to simply display a map of some region; additional dynamic information modes need to be displayed and put into context – from weather sensor readings and live aerial video feeds to daily news updates, photo collections and video archives.

The open-source community plays a crucial role in driving the development of the Geospatial Web. Collaborative efforts have provided a large number of add-ons for popular platforms. In the case of NASA World Wind, several of these external modules have been integrated into the core system. Participants in open-source projects identify, track and resolve technical problems, suggest new features and source code modifications, and often provide high-resolution data sets and other types of user-generated content.

This book presents the state-of-the-art in geospatial Web technology. It gradually exposes the reader to the technical foundations of the Geospatial Web, and to new interface technologies and their implications for human-computer interaction. Several chapters deal with the semantic enrichment of electronic resources, a process that yields extensive archives of Web documents, multimedia data, individual user profiles and social network data. The following chapters then demonstrate the use of geospatial technologies for managing virtual communities, and for monitoring, analyzing and mapping environmental indicators. Finally, the last four chapters address service-oriented architectures, and describe how distributed Web services facilitate the integration of knowledge repositories with geospatial platforms and third-party applications.

I congratulate the authors for their excellent and timely work. The book is not only a comprehensive, interdisciplinary collection of current research; it also introduces visionary concepts and outlines promising avenues for future research.

Patrick J. Hogan
Program Manager, NASA World Wind
worldwind.arc.nasa.gov

Preface

Contrary to early predictions that the Internet will obsolete geography, the discipline is increasingly gaining importance. In a 1998 speech at the California Science Center, former U.S. Vice President Al Gore called for replacing the prevalent desktop metaphor with a "multi-resolution, three-dimensional representation of the planet, into which we can embed vast quantities of geo-referenced data" (Gore 1998). After the successful introduction of three-dimensional geospatial platforms such as NASA World Wind,[1] Google Earth[2] and Microsoft Live Local 3D,[3] achieving the vision of a *Geospatial Web* seems more realistic than ever.

Dubbed the "holy grail of mapping" (Levy 2004), these geobrowsers aggregate and project layers of metadata onto scale-independent spherical globes. They are an ideal platform to integrate (i) cartographic data such as topographic maps and street directories, (ii) geotagged knowledge repositories aggregated from public online sources or corporate intranets, and (iii) environmental indicators such as emission levels, ozone concentrations and biodiversity density. By integrating cartographic data with geotagged knowledge repositories, the Geospatial Web will revolutionize the production, distribution and consumption of media products.

The appearance of geobrowsers in mainstream media coverage (see Chapter 1) increases public acceptance of geospatial technology and improves geospatial literacy, which today exists only among a small portion of highly educated people (Erle et al. 2005). Geospatial literacy includes the ability to understand, create and use geospatial representations for Web navigation, narrative descriptions, problem-solving and artistic expression (Liebhold 2004). In light of the explosive growth and diminished lifespan of information, geospatial literacy is becoming increasingly important, as the thought that needs to be followed in information discovery tasks is often spatial in nature (McCurley 2001). Geobrowsing platforms support such information discovery tasks by allowing users to switch between or integrate a large number of heterogeneous information services.

The 25 chapters contained in this edited volume summarize the latest research on the Geospatial Web's technical foundations, describe information services and collaborative tools built on top of geobrowsers and investigate the environmental, social and economic impacts of knowledge-intensive applications. Supplemental material including author biographies and bibliographic resources is available from the book's official Web site at

www.geospatialweb.com

The book emphasizes the role of contextual knowledge in shaping the emerging network society. Several chapters focus on the integration of geospatial and semantic technology to extract geospatial context from unstructured textual resources; e.g., to automatically identify and map the most relevant content for customized news services. Hybrid models combine such automated services with the advantages of individual and collaborative content production environments – for example by integrating "edited" material from newspapers and traditional encyclopedias with "evolving" content from collaborative Wiki applications.

Automatically annotating content acquired from these different sources creates knowledge repositories spanning multiple dimensions (space, time, semantics, etc.). Geospatial exploration systems will improve the accessibility and transparency of such complex repositories.

Keen competition between software and media companies surrounds the provision of geospatial exploration systems. The platforms are evolving quickly, gaining new functionality, data sources and interface options in rapid succession. But the currently available applications only hint at the true potential of geospatial technology. The Geospatial Web will have a profound impact on managing individual and organizational knowledge. It will not only reveal the context and geographic distribution of a broad range of information services and location-based resources but also help create and maintain virtual communities by matching people of similar interests, browsing behavior or geographic location.

Acknowledgements

This book would not have been possible without the help and contributions of many colleagues. Our first word of appreciation goes to the authors for their excellent work and active participation in the peer-review process. Each chapter was evaluated by three or four referees and revised at least once on the basis of their comments and criticism.

We would like to thank the following colleagues who generously provided additional reviews and feedback: Albert Weichselbraun, Andrea Polli, Andreas Juffinger, Benno Stein, Deana Pennington, Eva Micietova, Fridolin Wild, Herwig Rollett, Joachim P. Hasebrook, Jörg Westbomke, Klaus Leopold, Kostas Karatzas, Marc Van Liedekerke, Markus Strohmaier, Mathias Lux, Michael Granitzer, Nguyen Xuan Thinh, Panos Panagos, S.K. Ghosh, Soenke Dohrn, Stefan Kollarits, Tomas Pitner, Wei Liu, Wolf-Fritz Riekert, Wolfgang Kienreich and Yiwei Cao.

Gabriele Zorn-Pauli is to be commended for her valuable assistance in the editorial process. We would also like to thank the series editors of *Advanced Information and Knowledge Processing*, Xindong Wu and Lakhmi Jain, as well as the staff at Springer for their support and help in the materialization of this book.

Arno Scharl
Klaus Tochtermann
Graz, April 2007

Know-Center Graz – Austria's Competence Center for Knowledge Management
Graz University of Technology, Knowledge Management Institute
Inffeldgasse 21a, A-8010 Graz, Austria

www.know-center.at ▪ kmi.tugraz.at ▪ www.idiom.at ▪ www.ecoresearch.net

The Know-Center is funded by the Austrian Competence Center program Kplus under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT), and by the State of Styria. The IDIOM (Information Diffusion across Interactive Online Media) research project is funded by BMVIT and the Austrian Research Promotion Agency (FFG) within the strategic objective FIT-IT Semantic Systems (www.fit-it.at).

Table of Contents

FOUNDATIONS OF THE GEOSPATIAL WEB

NAVIGATING THE GEOSPATIAL WEB

BUILDING THE GEOSPATIAL WEB

GEOSPATIAL COMMUNITIES

ENVIRONMENTAL APPLICATIONS

GEOSPATIAL WEB SERVICES

List of Authors

Dipl.-Umweltwiss.
Christian Aden
PhD Candidate
University of Vechta, Chair of Landscape Ecology; Vechta, Germany

Dr
Pragya Agarwal
Lecturer
Department of Geomatic Engineering
University College London
London, UK

Dipl. Inf.
Dirk Ahlers
Research Assistant
OFFIS Institute for Information Technology; Oldenburg, Germany

MAMS
Boanerges Aleman-Meza
Research Assistant
University of Georgia, Computer Science Department, LSDIS Lab; Athens, GA, USA

Dr
Steve Battle
Research Engineer
Hewlett-Packard Laboratories
Bristol, UK

Mr
Torsten Becker
Publicist and Author
Founder of ExploreOurPla.net
Cologne, Germany

Dr
Roderic Bera
Lecturer
Department of Geomatic Engineering
University College London
London, UK

MA
Susan J. Bergeron
PhD Candidate
West Virginia University
Department of Geology and Geography
Morgantown, WV, USA

MD PhD
Daniel C. Berrios
Scientist
University of California, Santa Cruz
NASA Ames Research Center
Moffett Field, USA

Dr
Susanne Boll
Professor of Media Informatics and Multimedia Systems
University of Oldenburg
Oldenburg, Germany

MS
Gabriella Castelli
PhD Candidate
University of Modena and Reggio Emilia, Department of Science and Methodologies of Engineering; Reggio Emilia, Italy

Dr
Steve Cayzer
Research Engineer
Hewlett-Packard Laboratories
Bristol, UK

MSc
Jérôme Chemitte
PhD Candidate
Mission Risques Naturels,
Ecole des Mines de Paris, Pôle Cindyniques
Sophia Antipolis, France

Dr
Christophe Claramunt
Professor and Director
Naval Academy Research Institute
Brest, France

BA MSc
Rob Davies
Partner
MDR Partners
London, UK

BS
Mike Dean
Principal Engineer
BBN Technologies
Arlington, Virginia, USA

Dr
Bruce Denby
Senior Researcher
Norwegian Institute for Air Research
Kjeller, Norway

Dr
Catherine Dolbear
Senior Research Scientist
Ordnance Survey Research Labs
Southampton, United Kingdom

Dr
John Domingue
Deputy Director
Knowledge Media Institute (KMI)
The Open University
Milton Keynes, UK

Mr
Jim Farley
Vice President
Galdos Systems
Vancouver, Canada

Mr
Matthew Fleagle
Senior Technical Writer
LizardTech
Seattle, WA, USA

Dr
Mauro Gaio
Professor of Computer Science
LIUPPA Laboratory
Laboratoire d'Informatique de l'Université
de Pau et des Pays de l'Adour; Pau, France

Dr
Linlin Ge
Senior Lecturer
University of New South Wales
School of Surveying and Spatial
Information Systems
Sydney, Australia

MS
Michael P. Gerlek
Engineering Manager/Software Architect
LizardTech
Seattle, WA, USA

Dr
Alessio Gugliotta
Research Fellow
Knowledge Media Institute (KMI)
The Open University
Milton Keynes, UK

MSc
Leticia Gutierrez
Ontology Engineer
Essex County Council
Chelmsford, Essex, UK

PhD
Farshad Hakimpour
Research Associate
University of Georgia, Computer Science
Department, LSDIS Lab
Athens, GA, USA

Dr
Trevor M. Harris
Eberly Distinguished Professor of Geography
West Virginia University, Department of
Geology and Geography
Morgantown, WV, USA

BSc (Hons), BA (Hons)
Glen Hart
Principal Research Scientist
and Research Manager
Ordnance Survey Research Labs
Southampton, UK

Mr
Marcel Holy
Student Academic Staff
University of Vechta
Chair of Landscape Ecology
Vechta, Germany

Mgr
Jan Horalek
Air Quality Researcher
Czech Hydrometeorological Institute
Prague, Czech Republic

Dr
Jirí Hrebícek
Professor of Information Systems
Masaryk University
Institute of Biostatistics and Analyzes
Brno, Czech Republic

Mr
You-Heng Hu
PhD Candidate
University of New South Wales
School of Surveying and Spatial
Information Systems
Sydney, Australia

MSc
Julien Iris
PhD Candidate
Ecole des Mines de Paris
Pôle Cindyniques
Sophia Antipolis, France

Dipl.-Lök
Krzysztof Janowicz
Research Associate/PhD Student
Münster Semantic Interoperability Lab
Institute for Geoinformatics
University of Münster, Germany

BS
William Kammersell
Software Engineer
BBN Technologies
Arlington, Virginia, USA

PhD
Richard M. Keller
Computer Scientist
NASA Ames Research Center
Intelligent Systems Division
Moffett Field, Mountain View, CA, USA

Dipl.-Umweltwiss.
Lukas Kleppin,
PhD Candidate
University of Vechta
Chair of Landscape Ecology
Vechta, Germany

Dipl.-Landsch.-Ökol.
Eva Klien
Research Associate
Institute for Geoinformatics
University of Münster
Münster, Germany

Dr
Milan Konecný
Associate Professor and President of the International Cartography Association
Masaryk University, Laboratory of Cartography & Geography
Brno, Czech Republic

BA MA CGS (GIS/LIS)
Athanasios Tom Kralidis
Senior Systems Scientist
Environment Canada
Toronto, Ontario
Canada

Mr
Ron Lake
CEO
Galdos Systems
Vancouver, British Columbia
Canada

PhD
Julien Lesbegueries
PIV Project Member
LIUPPA Laboratory
Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour; Pau, France

PhD
Pierre Loustau
PIV Project Member
LIUPPA Laboratory
Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour; Pau, France

PhD
Marco Mamei
Researcher
University of Modena and Reggio Emilia
Department of Science & Methodologies of Engineering; Reggio Emilia, Italy

Dr
Ernesto Marcheggiani
Post Doc Researcher
Università Politecnica delle Marche,
Department of Applied Science of Complex Systems, DiSASC; Università Politecnica delle Marche; Ancona, Italy

MA
Graeme McFerren
Senior Researcher
Information & Communications
Technology for Earth Observation Group
Meraka Institute, CSIR
Pretoria, South Africa

Dr
Christian Morbidoni
Post Doc Researcher
Semantic Web and Multimedia Group (SEMEDIA); D.E.I.T, Università Politecnica delle Marche; Ancona, Italy

Dr
Aldo Napoli
Researcher
Ecole des Mines de Paris, Pôle Cindyniques
Sophia Antipolis, France

Ing.
Michele Nucci
PhD Candidate
Semantic Web and Multimedia Group (SEMEDIA); D.E.I.T, Università Politecnica delle Marche; Ancona, Italy

MSc
Matthew Perry
Research Assistant
University of Georgia, Computer Science Department, LSDIS Lab
Athens, GA, USA

Dr
Roland Pesch
Research Assistant
University of Vechta, Chair of Landscape Ecology; Vechta, Germany

BSc
Marc Richardson
Semantic Web Researcher
Next Generation Web Research
BT Group Chief Technology Office
Ipswich, UK

Dipl.-Eng.
Dumitru Roman
Researcher
DERI Innsbruck
University of Innsbruck
Innsbruck, Austria

MSc
Stacey Roos
Researcher
Information & Communications Technology for Earth Observation Group
Meraka Institute, CSIR
Pretoria, South Africa

MS
Alberto Rosi
PhD Candidate
University of Modena and Reggio Emilia
Department of Science and Methodologies of Engineering; Reggio Emilia, Italy

Mr
L. Jesse Rouse
PhD Candidate
West Virginia University, Department of Geology and Geography
Morgantown, WV, USA

BA MSc
Mary Rowlatt
Customer Relations Manager
Essex County Council
Chelmsford, Essex, UK

Dr
Christian Sallaberry
Assistant Professor of Computer Science
LIUPPA Laboratory
Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour; Pau, France

DDr
Arno Scharl
Professor of New Media and Knowledge Management
Know-Center and Graz University of Technology, Knowledge Management Institute; Graz, Austria

Dr
Gunther Schmidt
Research Assistant
University of Vechta
Chair of Landscape Ecology
Vechta, Germany

Dr
Winfried Schröder
Professor
University of Vechta
Chair of Landscape Ecology
Vechta, Germany

PhD
Amit Sheth
Professor of Computer Science
University of Georgia, Computer Science Department, LSDIS Lab
Athens, GA, USA

Dr Ing
Maarten Sierhuis
Senior Scientist
Research Institute for Advanced Computer Science, NASA Ames Research Center, Moffett Field
Mountain View, CA, USA

Msc
Peter A. M. de Smet
Senior Policy Researcher
Netherlands Environmental Assessment Agency; Bilthoven, The Netherlands

BSc
Sandra Stinčić
Semantic Web Researcher
Next Generation Web Research
BT Group Chief Technology Office
Ipswich, UK

BA MSc
Vlad Tanasescu
PhD Student
The Open University
Knowledge Media Institute (KMi)
Milton Keynes, UK

MSc
Andrew Terhorst
Research Group Leader
Information and Communications Technology for Earth Observation Group
Meraka Institute, CSIR
Pretoria, South Africa

Dr
Carlo Torniai
Researcher
Multimedia Integration and Communication Center, Universita' di Firenze
Firenze, Italy

Dr
Giovanni Tummarello
Post Doc Researcher
Semantic Web and Multimedia Group (SEMEDIA); D.E.I.T, Università Politecnica delle Marche; Ancona, Italy

Mr
Marc Wick
Software Engineer
Project Lead Geonames.org
St. Gallen, Switzerland

PhD
Franco Zambonelli
Professor
University of Modena and Reggio Emilia
Department of Science and Methodologies of Engineering; Reggio Emilia, Italy

Chapter 1

Towards the Geospatial Web: Media Platforms for Managing Geotagged Knowledge Repositories

Arno Scharl

Abstract. *International media have recognized the visual appeal of geobrowsers such as NASA World Wind and Google Earth, for example, when Web and television coverage on Hurricane Katrina used interactive geospatial projections to illustrate its path and the scale of destruction in August 2005. Yet these early applications only hint at the true potential of geospatial technology to build and maintain virtual communities and to revolutionize the production, distribution and consumption of media products. This chapter investigates this potential by reviewing the literature and discussing the integration of geospatial and semantic reference systems, with an emphasis on extracting geospatial context from unstructured text. A content analysis of news coverage based on a suite of text mining tools (webLyzard) sheds light on the popularity and adoption of geospatial platforms.*

1.1 Introduction

Historically, media technology enters the market via new types of content that drive adoption and validate emerging business models. For true media innovation to have human impact, however, it must affect the imagination – creating an associated magic “behind the eyeballs” that changes people’s behavior in their commercial, academic and personal environments (Stapleton and Hughes 2006). The following hypothetical scenario outlines how geospatial technology may radically change working environments, impact workflows within and across organizations, and enrich the interaction between content providers and their target audience.

Kathryn O'Reilly is a freelance editor who sells her ability to gather, filter and prioritize electronic content. In a virtual world built on contextualized information spaces, Kathryn seamlessly switches between geographic and semantic topologies. She begins her typical working day floating in the virtual space above Earth, ready to navigate the globe and semantic structures via subtle movements of her eyes. An extensive portfolio of add-on functionality is accessible through haptic devices. From her elevated position, Kathryn not only observes the rise and decay of topics, but also the unfolding of social structures based on the unique social networks of her friends and business contacts. Across these networks she builds and shares her knowledge repository and composes media products that are continuously being validated and enriched by the latest news feeds and third-party sources.

The underlying content management system tailors the format of her articles to the changing preferences of her regular readers. Kathryn adds, selects, categorizes, aggregates, filters and extrapolates information along multiple dimensions, with minimal cognitive requirements. She can structure her daily workflows, access archives of historic *textual and multimedia data and customize her virtual environment. Adaptive communication services allow her to interact with predefined or dynamically assembled groups of like-minded individuals. At any point in time, Kathryn may use portions of the information space to initialize simple what-if scenarios or advanced socioeconomic simulations, investigating the complex interplay among computer-generated avatars, automated information services and real-world participants.*

In the words of McLuhan, media as an extension of ourselves provide new transforming vision and awareness (McLuhan 1964). In the early 1940s, the first images of Earth from space eroded limitations to human perception, triggered profound self-reflexive experiences (DeVarco 2004) and revitalized public desire to preserve a beautiful but vulnerable planet (Biever 2005). Thanks to human space exploration, therefore, most users will instantly recognize our planet and find it an intuitive and effective metaphor to access and manage geotagged information: "There it is, that good old pale blue dot in all its earthly glory, right there on your computer screen. It's a familiar sight, even from a sky-high perspective experienced only by astronauts and angels" (Levy 2004, 56).

As the concepts of "desktop," "village" and "landscape" have shown, well-known interface metaphors are powerful instruments to gain market acceptance (Fidler 1997). Geobrowsers promote the "planet" metaphor by providing users with an accurate visual representation and allowing them to browse geospatial data from a satellite perspective. Using standardized services such as the bitmap-based Web Map Service (WMS)[4] and the vector-based Web Feature Service (WFS),[5] image tiles and vector data including geo-positioning information are retrieved from a central server, arranged into real-time mosaics and mapped onto three-dimensional representations of the globe. Altering the field-of-view angle allows users to switch between detailed views and highly aggregated representations. Users can effortlessly zoom from Blue Marble Data[6] at a 1-kilometer-per-pixel rate, for example, to the detailed mosaic of LandSat 7 Data[7] at 15 meters per pixel (Hogan and Kim 2004). Adding the option to tilt the display relative to the spectator's point of view adds altitude as a third dimension.

Given the potential of the "planet" metaphor, academia and industry alike call for a new generation of geospatial interfaces with simple yet powerful navigational aids that facilitate the real-time access and manipulation of geospatially and semantically referenced information.

1.2 Geospatial Reference System

Observing, aggregating and visualizing human behavior are common activities, from tracking customers in retailing outlets to monitoring traffic in congested urban areas, or analyzing the clickstreams of online shoppers based on Web server logfiles (Scharl 2001). Prior to the advent of the Global Positioning System (GPS) and Radio Frequency Identification (RFID), the lack of appropriate technology to pinpoint a user's precise location restricted the functionality of many applications. Nowadays, aggregated visualizations of individual actions are a familiar sight, as geobrowsers redefine the look and feel of user interfaces and leverage the knowledge about a user's precise location to unlock organized indices to the physical world (Kendall 2005).

Information retrieval research has also discovered geobrowsers as an effective platform to identify and access relevant information more effectively. An increasing number of applications use geospatial extensions for specifying queries and structuring the presentation of results.

Most providers of geobrowsing platforms offer *Application Programming Interfaces (APIs)* or *XML scripting* to facilitate building third-party online services on top of their platforms (Roush 2005). Multiple layers of icons, paths and images can be projected via these services. Such visual elements are scaled, positioned on the globe, and linked to (Web) documents,[8] photo collections[9] and other external resources (Neches et al. 2001). Latitude and longitude variables determine the symbols' position, while distance above surface values specify whether symbols hover above ground. A good example is data from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS),[10] providing daily updated planetary imagery at resolutions up to 250 meters per pixel that documents natural events such as fires, floods, storms and volcanic activity (Hogan and Kim 2004). The left screenshot of Figure 1.1 shows an MODIS overlay of Hurricane Katrina as of August 29, 2005.

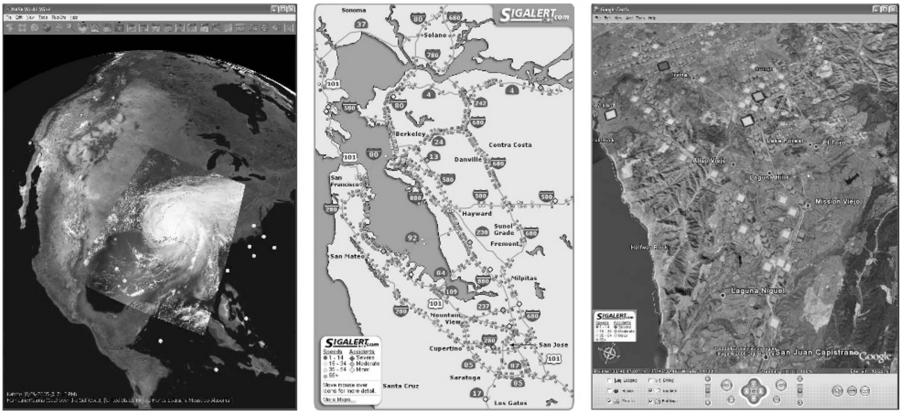


Figure 1.1: Data integration with NASA World Wind and Google Earth

The availability of APIs is largely responsible for the growing popularity of mash-ups. Often released by individuals or the open-source community, mash-ups combine publicly available data and interface services from multiple providers into an integrated user experience (Hof 2005). The map in the center of Figure 1.1 displays the original Sigalert.com service that aggregates real-time traffic data from the San Francisco Bay Area. The screenshot on the right exemplifies the idea of a mash-up, using the Sigalert.com traffic data to project symbols for accidents and current traffic speeds onto the Google Earth representation of Southern Los Angeles.[11]

1.2.1 Extraction and Disambiguation of Geospatial Context

Concentrated efforts are under way to geotag as much existing information as possible. Geotagging refers to the process of assigning geospatial context information, ranging from specific point locations to arbitrarily shaped regions. Different sources of geospatial context information for annotating Web resources often complement each other in real-world applications (McCurley 2001):

- Annotation by the author, manually (Daviel and Kaegi 2003) or through location-aware devices such as car navigation systems, RFID-tagged products and GPS-enabled cellular handsets. These devices geotag information automatically when it is being created.

- Determining the location of the server – e.g., by querying the Whois[12] database for domain registrations, monitoring how Internet traffic is routed on a macro level, or by analyzing the domain of a Web site for additional cues.
- Automated annotation of existing documents. The processes of recognizing geographic context and assigning spatial coordinates are commonly referred to as *geoparsing* and *geocoding*, respectively.

Once geospatial context information becomes widely available, any point in space will be linked to a universe of commentary on its environmental, historical and cultural context, to related community events and activities and to personal stories and preferences. Even locative spam will become a common phenomenon (Erle et al. 2005) with the widespread introduction of location-based services, geospatial gaming environments and other commercial applications.

At present, however, many metadata initiatives still suffer from the chicken and egg problem, wishing that existing content was retrofitted with metadata (McCurley 2001). This "capture bottleneck" results from the beneficiaries' lack of motivation to devote the necessary resources for providing a critical mass of metadata (Motta et al. 2000). Geotagging projects are no exception. Acknowledging calls to automate the semantic annotation of documents (Benjamins et al. 2004; Domingue and Motta 2000), the following sections focus on the third category, the automated geoparsing and geocoding of existing Web resources – online news, for example, or other types of unstructured textual data found on the Web.

1.2.1.1 Geoparsing

All human artifacts have a location history, which commonly includes a creation location and a current location (Spohrer 1999). Given the availability of metadata, geospatial applications can map the whole life cycle of such artifacts. Electronic resources contain metadata as explicit or implicit geographic references. This includes references to physical features of the Earth's surface such as forests, lakes, rivers and mountains, and references to objects of the human-made environment such as cities, countries, roads and buildings (Jones et al. 2001). Addresses, postal codes, telephone numbers and descriptions of landmarks also allow us to pinpoint exact locations (Ding et al. 2000; McCurley 2001).

At least 20 percent of Web pages contain easily recognizable and unambiguous geographic identifiers (Delboni et al. 2005). News articles are particularly rich in such identifiers, since they usually discuss the location where an event took place, or where it was reported from (Morimoto et al. 2003). The BBC article "Vienna Marking Mozart Milestone" (Bell 2006), for example, has a target geography of EUROPE/AUSTRIA/VIENNA and a source geography of EUROPE/UNITED KINGDOM/LONDON. In addition to target and source geography (Amitay et al. 2004), natural language processing can also be used to extract the geographic scope (i.e., intended reach) of Web resources (Wang et al. 2005).

Identifying and ranking spatial references by semantically analyzing textual data is a subset of the more general problem of *named entity recognition*, which locates and interprets phrasal units such as the names of people, organizations and places (Cowie and Lehnert 1996; Weiss et al. 2005). As with most named entity recognition tasks, false positives are inevitable – e.g., documents that quote addresses unrelated to the their actual content (Morimoto et al. 2003).

Ambiguity, synonymy and changes in terminology over time further complicate the geoparsing of Web documents (Amitay et al. 2004; Kienreich et al. 2006; Larson 1996). Identical lexical forms refer to distinct places with the same name (VIENNA refers to the capital of Austria as well as a town in Northern Virginia, USA) or have geographic and non-geographic meanings: TURKEY (large gallinaceous bird; bicontinental country between Asia and Europe), MOBILE (capable of moving; city in Alabama, USA), or READING (processing written linguistic messages; town in Massachusetts, USA). Geoparsers also need to correctly process references to identical or similar places that may be known under different names, or may belong to different levels of administrative or topographical hierarchies (Jones et al. 2001).

1.2.1.2 Geocoding

Once a location has been identified, precise spatial coordinates – latitude, longitude and altitude – can be assigned to the documents by querying structured geographic indices (gazetteers) for matching entries (Hill et al. 1999; Tochtermann et al. 1997). This process of associating documents with formal models is also referred to as document enrichment (Domingue and Motta 2000; Motta et al. 2000). Examples of formal geographic models are the Geographic Names Information System (GNIS),[13] the World Gazetteer,[14] the classifications of the United Nations Group of Experts on Geographical Names,[15] the Getty Thesaurus of Geographic Names[16] and ISO 3166-1 Country Codes.[17]

While simple gazetteer lookup has the advantage of being language-independent, advanced algorithms consider lexical and structural linguistic clues as well as contextual knowledge contained in the documents; e.g., dealing with ambiguity by removing stop-words, identifying references to people and organizations (Clough 2005) and applying contextual rules like "single sense per document" and "co-occurring place names indicate nearby locations". Each identified reference is assigned a probability *P(name, place)* that it refers to a particular place (Amitay et al. 2004). The location that receives the highest probability is then assigned a canonical taxonomy node such as EUROPE/AUSTRIA/VIENNA; 48°14' N, 16°20' E.

1.2.2 Managing Geospatial Context

Standardized metadata frameworks often include geospatial attributes like the Dublin Core Metadata Initiative's "Coverage" tag (McCurley 2001).[18] The need for controlled vocabularies and shared meaning suggests that ontologies are going to play a key role in managing geospatial context information. While conflicting definitions of "ontology" abound (Guarino 1997), most researchers agree that the term refers to a designed artifact formally representing shared conceptualizations within a specific domain (Gahleitner et al. 2005; Jarrar and Meersman 2002).

Geo-ontologies encode geographical terms and semantic relationships such as containment, overlap and adjacency (Tochtermann et al. 1997). Spatially aware search engines use ontological knowledge for query term expansion and disambiguation, relevance ranking and Web resource annotation (Abdelmoty et al. 2005). Geo-ontologies can either be represented through generic markup languages like the Web Ontology Language (OWL)[88] endorsed by the World Wide Web Consortium (Horrocks et al. 2003; Smith et al. 2004) or more specific approaches like the Geography Markup Language (GML)[28] developed by the Open Geospatial Consortium (Lake et al. 2004; see Chapter 2 "Infrastructure for the Geospatial Web" for a more detailed discussion).

1.3 Semantic Reference System

Geospatially referenced information enables geobrowsers to map annotated content units from various sources, track human activities and visualize the structure and dynamics of virtual communities. But geobrowsers can also serve as a generic image rendering engine to project other types of imagery. Diverting them from their traditional purpose and connecting them to *semantically* referenced information, they can be used to visualize *knowledge planets* based on layered thematic maps. Such maps are visual representations of semantic information spaces based on a landscape metaphor (Chalmers 1993).

Generally, two sets of information need to be integrated and mapped to latitude and longitude – image tiles and terrain information. Knowledge planets are generated by orthographically projecting and tiling thematic maps. The planet metaphor allows visualizing massive amounts of textual data. At the time of map generation, the knowledge planet's topology is determined by the content of the knowledge base. The peaks of the virtual landscape indicate abundant coverage on a particular topic, whereas valleys represent sparsely populated parts of the information space.

Extending the planet metaphor, search results can be visualized as cities, landmarks or other objects of the manmade environment. Zooming provides an intuitive way of selecting the desired level of aggregation. Unique resource identifiers link concepts embedded in the thematic maps to related news articles, encyclopedia entries or papers in scientific journals. With such a query interface that hides the underlying complexity, exploring complex data along multiple dimensions is as intuitive as using a geobrowser to get a glimpse of the next holiday destination.

VisIslands, a thematic mapping algorithm similar to SPIRE's Themescape (Wise 1999) and its commercial successor Cartia/Aureka (see Figure 1.2),[19] supports dynamic document clustering (Andrews et al. 2001; Sabol et al. 2002). Initially, the document set is pre-clustered using hierarchical agglomerative clustering (Jain et al. 1999), randomly distributing the cluster centroids in the viewing rectangle. The documents belonging to each cluster, as determined by the pre-clustering, are then placed in circles around each centroid. The arrangement is fine-tuned using a linear iteration force-directed placement algorithm adapted from Chalmers (1996). The result resembles a contour map of islands. Fortunately, algorithms based on force models easily generalize to the knowledge planets' spherical geometries.

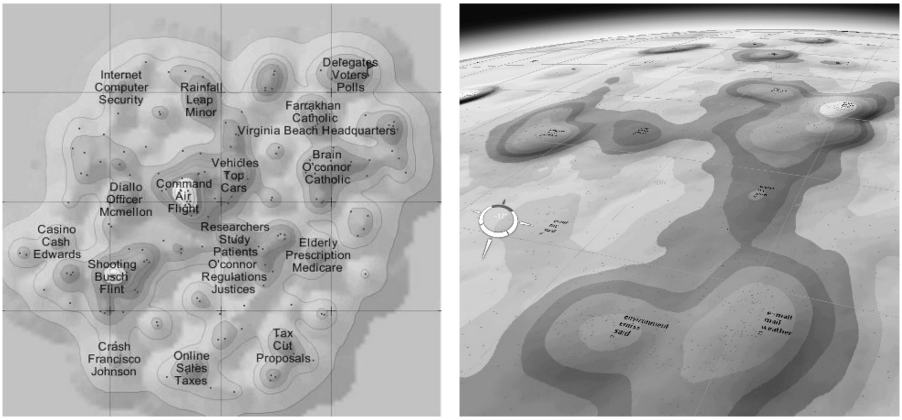


Figure 1.2: Thematic mapping exemplified by Cartia (left) and a knowledge planet prototype (right)

The IDIOM (Information Diffusion across Interactive Online Media)[20] research project extends and refines the VisIslands thematic mapping component to improve throughput and scalability, generate layered thematic maps, and provide a Web Map Service (WMS) that serves these maps as image tiles for various geobrowsing platforms. The NASA World Wind screenshot of Figure 1.2 shows an early prototype of this service. The transition from two-dimensional thematic maps to three-dimensional knowledge planets poses conceptual and technical challenges – the initial arrangement of major concepts, for example, which should be guided by domain ontologies. Users will expect a consistent experience when rotating the planet. This requires a seamless flow of concepts when crossing the planet's 0° meridian line. The same principle applies to zooming operations. Analogous to Landsat-7 data, multiple layers of thematic maps in different resolutions and with appropriate sets of captions have to be synchronized with each other.

1.4 Geospatial Publishing

Technological convergence and the move towards digital media continue to drive today's newsrooms (Pavlik 1998). While many innovations that gain ground in the media industry are largely invisible to the end user, geobrowsers directly impact the consumption of news media, change mainstream storytelling conventions and provide new ways of selecting and filtering news stories. By facilitating the access of annotated knowledge repositories, geobrowsers set the stage for the Geosptial Web as a new platform for content production and distribution.

1.4.1 Content Production and Distribution

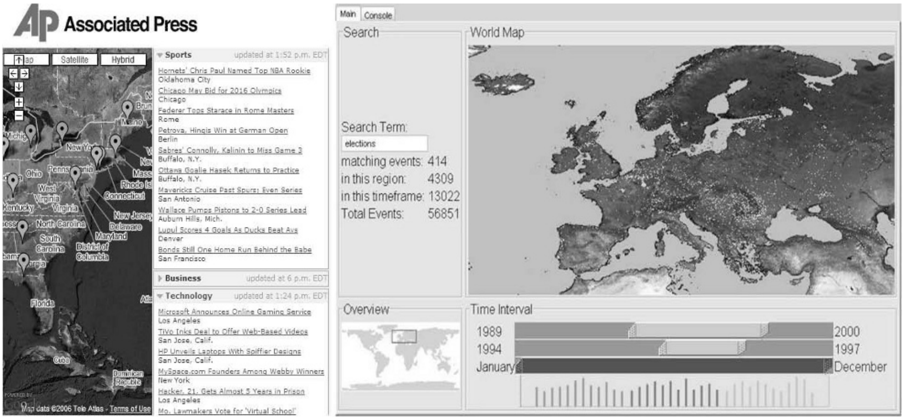
Hybrid models of individual and collaborative content production are particularly suited for geobrowsers, which can integrate and map *individual sources* (monographs, commentaries, blogs), *edited sources* (encyclopedias, conference proceedings, traditional news services), *evolutionary sources* (Wiki applications, open-source project documentations) and *automated sources* (document summarizers, news aggregators). Geobrowsing technology not only impacts the production of content, but also its distribution, packaging and consumption. When specifying preferences for personalized news services, for example, geobrowsers are effective tools to pinpoint locations and specify geographic areas to be covered by the news service.

Personalized news services require content that is correctly annotated along at least three dimensions: (i) *spatial* – e.g., distinguishing between source and target geography; see Section 1.2.1; (ii) *semantic* – e.g., assigning the most relevant concepts from a controlled vocabulary; and (iii) *temporal* – e.g., adding timestamps for the reported event, the initial publication and subsequent revisions. Online news can be organized, indexed, searched and navigated along these dimensions:

- The *geographical scope* of an article allows filtering and prioritizing electronic content in line with the user's area of interest, which is often different from his or her actual location.
- *Topical similarity* is another common dimension to tag and filter news content, often matched against user-specific degree of interest functions.
- Finally, by adding a *temporal dimension* through time distribution graphs or visual animation, change over time along any other dimension can be captured; e.g., the unfolding of events, news distribution patterns or the inter-individual propagation of personal messages.

Most geographic information systems, however, treat time as an attribute rather than a separate dimension (Johnson 2004). This is about to change as geobrowsing platforms prepare the transition towards a fully functional Geospatial Web. Dynamic queries, interactive time displays and playback controls will enable users to identify the rise and decay of topics – the diffusion of news coverage on natural disasters, for example, or the impact of political events.

The simplest way of developing a news browser is to combine existing data sources and interface services (see Section 1.2). The news summary[21] on the left side of Figure 1.3 receives the News Feeds of Associated Press,[22] processes this stream of data with the Yahoo! Geocoding API[23] and displays the results via the Google Maps interface. More specific requirements or research interests often result in standalone applications. The second screenshot of Figure 1.3 (Rüger 2005) shows a geotemporal news browser that allows users to search a news database via query terms and time-interval sliders and presents matching articles mapped onto a region of interest. It follows Shneiderman and Plaisant's (2004) information seeking mantra: generate an overview, provide zooming and filtering, and present details on demand. These guidelines avoid clutter in the display, which results from projecting too many content items from a large knowledge repository simultaneously (Larson 1996). Instead of showing the complete set of available news items, for example, a user may wish to restrict the display to articles on climate change that were published in the online editions of Italian newspapers within the last 48 hours.



Sports updated at 1:52 p.m. EDT

Hornets' Chris Paul Named Top NBA Rookie
Oklahoma City
Chicago May Bid for 2016 Olympics
Chicago
Federer Tops Starace in Rome Masters
Rome
Petrova, Hingis Win at German Open
Berlin
Sabres' Connolly, Kalinin to Miss Game 3
Buffalo, N.Y.
Ottawa Goalie Hasek Returns to Practice
Buffalo, N.Y.
Mavericks Cruise Past Spurs; Even Series
San Antonio
Wallace Pumps Pistons to 2-0 Series Lead
Auburn Hills, Mich.
Lupul Scores 4 Goals As Ducks Beat Avs
Denver
Bonds Still One Home Run Behind the Babe
San Francisco

Business updated at 6 p.m. EDT

Technology updated at 1:24 p.m. EDT

Microsoft Announces Online Gaming Service
Los Angeles
TiVo Inks Deal to Offer Web-Based Videos
San Jose, Calif.
HP Unveils Laptops With Spiffier Designs
San Jose, Calif.
MySpace.com Founders Among Webby Winners
New York
Hacker, 21, Gets Almost 5 Years in Prison
Los Angeles
Mo. Lawmakers Vote for 'Virtual School'

Search Term: elections
matching events: 414
in this region: 4309
in this timeframe: 13022
Total Events: 56851

Time Interval
1989 2000
1994 1997
January December

Figure 1.3: Interfaces for accessing geo-referenced news archives

1.4.2 Economic Implications

In light of the observable technical trends, irrespective of their reach and target audience, newsrooms will have to come to terms with metadata (Schutzberg 2005). The widespread availability of metadata will drive the transition towards the Geospatial Web. Emerging geospatial technology supports restructuring processes within the media sector, enhances the workflows of virtual newsrooms and promotes locally dispersed content production. It also facilitates the distribution of (customized) electronic content, which is usually characterized by network effects. Metcalfe's law describes such effects by stipulating that the aggregate value of networks increases with approximately the square number of adopters (Swann 2002), which suggests first-mover advantages and lock-in effects due to high switching costs once a network technology dominates the market. Consequently, successful business strategies for providers of geobrowsing platforms and distributors of media products built on top of these platforms use innovation to attract and retain users, quickly grow a community of like-minded individuals around a new technology and successively enlarge this community through synergy effects with other products and services (Wilk 2005). The rise of Google Earth as outlined in Section 1.4.3 and the ever-increasing number of mash-ups leveraging this platform exemplify a successful implementation of this strategy.

But the race to provide the dominant geospatial platform is far from over and might trigger a new standard war (Google's purchase of Keyhole[24] and Microsoft's purchase of GeoTango[25] and Vexcel[26] demonstrate the perceived strategic potential of three-dimensional platforms for aerial imagery). Strong network effects in markets with powerful positive feedback loops tend to increase the likelihood and intensity of standard wars. In addition to the first-mover advantage of controlling a large base of loyal or locked-in customers, success factors in a standard war include brand name and reputation, intellectual property rights, the ability to innovate, manufacturing capabilities and strength in complements (Shapiro and Varian 1999).

For the Geospatial Web, such complements range from repositories of geotagged documents and user-generated content (e.g., tags and other types of annotation) to location-based services and third-party applications (e.g., simulation games within a geospatial context). With its Flight Simulator,[27] for example, Microsoft looks back on more than 25 years of developing a successful geospatial game engine. Considering its unique position in the operating systems market and large base of locked-in customers, it does not come as a surprise that the company has joined the race to provide the underlying infrastructure for a three-dimensional Geospatial Web. This strategy has worked before. From the first browser war fought against Netscape in the 1990s, Microsoft is known for its "embrace and extend" strategy – imitating technological advances and successfully incorporating them into its flagship products (Shapiro and Varian 1999). It remains to be seen whether the three-dimensional capabilities and high-resolution city textures of Microsoft Virtual Earth 3D will suffice in light of Google's dominance in the search engine market and obvious opportunities to geo-enable the popular and rapidly growing portfolio of Google services.

While platform providers hope to become the substratum upon which all types of electronic content are layered (Levy 2004), first-mover advantages gained through network effects might allow innovative media companies to dominate the information spaces built on top of these platforms. The content management systems of media companies often contain rich geospatial annotations, reflecting both the source and target geography of articles. For articles without geospatial references or only partial annotations, geotagging as outlined in Section 1.2.1 can add the missing information.

Previous geotagging research has developed methods not only to identify a location referenced in a Web resource but also to capture the geographical distribution of its target audience. The *geographical scope* describes the geographical area that its creator intends to reach (Ding et al. 2000). Distinguishing globally relevant material from publications targeting the national, state or city level is particularly relevant, as virtually all media planning models consider gross impressions, reach and frequency of media products (Cannon 2001).

1.4.3 Media Coverage on Geospatial Platforms

Geo-informatics represents an established discipline that has created an industry with remarkable revenues (Wilk 2005), largely hidden from the public eye. The launch of powerful yet intuitive-to-use geobrowsers has increased public awareness of geospatial technology considerably. Spurred by space photography, global satellite positioning, mobile phones, adaptive search engines and new ways of annotating Web content, the “ancient art of cartography is now on the cutting edge” (Levy 2004, 56). Many current articles shine a spotlight on geospatial technologies, describe trends in mobile geospatial applications, investigate the emerging industry of local search or report unusual objects found on satellite images.

In the past, the process of collecting and analyzing such articles was time-consuming and expensive and often yielded incomplete information. Nowadays, information is readily available online, allowing for inexpensive, fast and topical research. As traditional media extend their dominant position to the online world, analyzing their Web sites reflects an important portion of Web content that the average Internet user accesses. On a macro level, analysts gain insights into publicity through incidental news coverage by monitoring information flows across media sites (Scharl et al. 2005). On a micro level, documents retrieved from Web sites contain valuable information about trends and organizational strategies (Scharl 2000).

To investigate the media coverage on geospatial platforms, 129 Web sites were sampled in quarterly intervals between May 2005 and January 2006, drawing upon the *Newslink.org*, *Kidon.com* and *ABYZNewsLinks.com* directories to compile a list of international media sites from seven English-speaking countries: United States, United Kingdom, Canada, Australia, South Africa, New Zealand and Ireland. The *webLyzard.com* crawler followed the Web sites’ hierarchical structure until reaching 50 megabytes of textual data, a limit that helped reduce the dilution of top-level information by content in lower hierarchical levels (Scharl 2004). Updated news articles often result in multiple versions of the same content (Kutz and Herring 2005). The system thus identified and removed redundant segments like headlines and news summaries, whose appearance on multiple pages would otherwise distort frequency counts.

Media attention was calculated as the relative number of references to a platform (in occurrences per million tokens). A pattern matching algorithm processed a list of regular expressions, considering common term inflections while excluding potentially ambiguous terms. Table 1.1 categorizes these regular expressions into references to either three-dimensional (3D) or two-dimensional (2D) platforms.

Table 1.1: Regular expression query for geospatial platforms

<table>
<tr><th>Geospatial Platforms (3D)</th><th>Geospatial Platforms (2D)</th></tr>
<tr><td>(earth|globe|planet)(-|)?(browser|tool|viewer)s?
(microsoft|msn?)(-|)(visual|virtual)(-|)?earth
(virtual|digital)(-|)?(earth|globe|planet)s?
geo(fusion|matrix)
google(-|)?earth
keyhole(-|)?(2|earthviewer|pro|inc|markup)
nasa(-|)?world(-|)?wind
terrafly
terra(-|)?(suite|explorer|builder|gate)
skyline software
world(-|)?wind central
geo(-|)?tango</td><td>map(-|)?(browser|tool|viewer)s?
(microsoft|msn?) map(-|)?point
google(-|)?(local|maps?)
map(-|)?quest
map(-|)?machine
windows live(-|)?local
yahoo!?(-|)?maps?
parc map viewer
terrain(-|)?(browser|tool|viewer)s?</td></tr>
</table>

Figure 1.4 summarizes the number of occurrences identified by the pattern-matching algorithm. Between Q2/2005 and Q1/2006, coverage on 2D and 3D platforms increased significantly by more than 300 and 1,100 percent, respectively (Wilcoxon Signed Ranks; $p < 0.05$). In the second quarter of 2005, coverage on 2D platforms exceeded coverage on their 3D counterparts (Mann-Whitney; $p < 0.05$). Results from the first quarter of 2006 showed a different picture. There was no significant difference between the categories, although 3D platforms took a slight lead with an average relative frequency of exactly one occurrence per million tokens. Receiving 83 percent of the coverage, Google Earth has been the primary driver behind the observable increase in popularity. This represents a remarkable feat with a product only launched in June 2005, not receiving any mentions in the second quarter of 2005. As of January 2006, MapQuest still dominated the 2D category with 46 percent of total coverage, but Google Maps was catching up rapidly with a share of 44 percent.

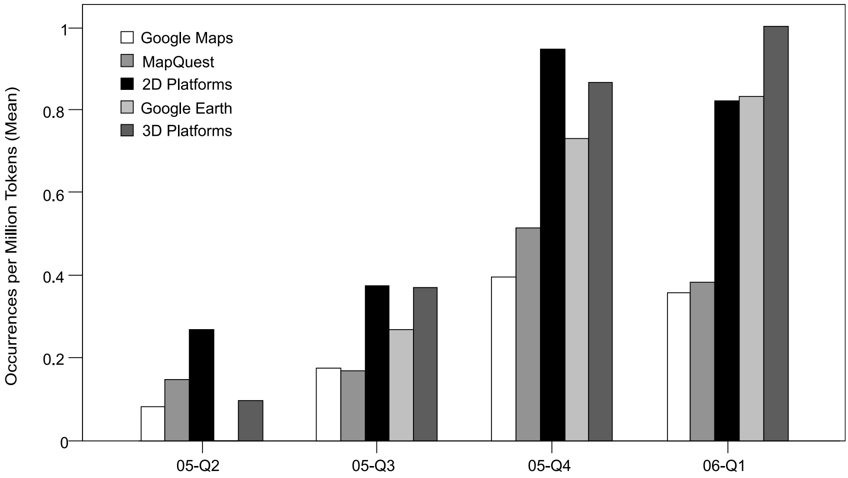


Figure 1.4: Media coverage of geospatial platforms

1.5 Conclusions and Outlook

By integrating cartographic geodata with geotagged hypermedia, the Geospatial Web "may ultimately be the big disruptive innovation of the coming decade" (Erle et al. 2005, xxv). As such, it will serve as a catalyst of social change and enabler of a broad range of as yet unforeseen applications.

The introduction of geobrowsing platforms has popularized the process of "annotating the Planet" (Udell 2005). This chapter outlined the underlying technology, discussed methods to "geo-enable" existing knowledge repositories through parsing geospatial references, and presented several geospatial applications in a media context. A quarterly snapshot of international media coverage revealed the increasing popularity of geospatial technology, particularly as far as three-dimensional platforms are concerned.

Science and technology's accelerated advancement demands constant media innovation, from idea to utility (Stapleton and Hughes 2006). In this competitive environment, geography is emerging as a fundamental principle for structuring the Web (Roush 2005) – a principle that yields the world's knowledge through the lens of location (Levy 2004, 58). The strategy of adding location metadata to existing databases and accessing the vast amounts of information stored in these databases via geospatial services weds physical and virtual spaces, deepens our experiences of these spaces and incorporates them into our everyday lives (Roush 2005). Coupling tagged knowledge repositories with satellite surveillance and other real-time data sources is a further step towards the *Earth as Universal Desktop*, an idea widely popularized in Neal Stephenson's 1992 novel "Snow Crash":

> *"A globe about the size of a grapefruit, a perfectly detailed rendition of Planet Earth, hanging in space at arm's length in front of his eyes. … It is a piece of CIC [Central Intelligence Corporation] software called, simply, Earth. It is the user interface that CIC uses to keep track of every bit of spatial information that it owns … It's not just continents and oceans. It looks exactly like the Earth would look from a point in geosynchronous orbit directly above L.A., complete with weather systems – vast spinning galaxies of clouds, hovering just above the surface of the globe, casting gray shadows on the oceans and polar ice caps, fading and fragmenting into the sea. … The computer, bouncing low-powered lasers off his cornea, senses this change in emphasis, and then Hiro gasps as he seems to plunge downward toward the globe, like a space-walking astronaut who has just fallen out of his orbital groove." (Stephenson 1992, 100ff.)*

Besides changing individual working environments, geobrowsers are ideally suited for creating and maintaining location-aware communities, bringing people together who share common needs or desires – e.g., communities of friends and social contacts, gaming enthusiasts, political activists or professional acquaintances. Within these communities, geospatial technology helps analyze topics of interest, from the state of the environment to political campaigns, demographic disparity, the progress of civil and urban planning efforts or the structure and efficiency of telecommunications or transportation networks (Erle et al. 2005).

The popularity of contextual advertising and location-based services indicates the technology's remarkable commercial potential. For marketers exploring new media for emerging business opportunities, for instance, the Geospatial Web is "the equivalent of a virgin continent waiting to be planted with billboards" (Roush 2005, 58f.). But established media companies often base strategic decisions on repeating financial successes, a practice that discourages radical innovation (Stapleton and Hughes 2006) and favors nondisruptive technologies. The fact that geospatial technology is compatible with current Internet communication models might help explain its unprecedented rate of adoption, from both organizational and individual perspectives. It integrates well with current protocols and therefore does not replace but complements established modes of navigating Internet resources. This process goes hand-in-hand with the transition towards the Web 2.0, a term that describes advances in Web technology governed by strong network effects and the harnessing of collective intelligence through customer-self service and algorithmic data management (O'Reilly 2005).

Acknowledgements. The author wishes to thank colleagues at the Know-Center and at Vienna University of Economics & Business Administration for their feedback on earlier versions of this chapter. Particular thanks go to Albert Weichselbraun for his help in acquiring the media coverage data (Section 1.4.3). The Know-Center is funded by the Austrian Competence Center program K+ under the auspices of the Austrian Ministry of Transport, Innovation & Technology (BMVIT), and the State of Styria. The IDIOM Project (www.idiom.at) is funded by BMVIT and the Austrian Research Promotion Agency within the strategic objective FIT-IT (www.fit-it.at).

Chapter 2

Infrastructure for the Geospatial Web

Ron Lake • Jim Farley

Abstract. *Geospatial data and geoprocessing techniques are now directly linked to business processes in many areas. Commerce, transportation and logistics, planning, defense, emergency response, health care, asset management and many other domains leverage geospatial information and the ability to model these data to achieve increased efficiencies and to develop better, more comprehensive decisions. However, the ability to deliver geospatial data and the capacity to process geospatial information effectively in these domains are dependent on infrastructure technology that facilitates basic operations such as locating data, publishing data, keeping data current and notifying subscribers and others whose applications and decisions are dependent on this information when changes are made. This chapter introduces the notion of infrastructure technology for the Geospatial Web. Specifically, the Geography Markup Language (GML) and registry technology developed using the ebRIM specification delivered from the OASIS consortium are presented as atomic infrastructure components in a working Geospatial Web.*

2.1 What Is the Geospatial Web?

This article considers the technical foundations for the development, evolution and deployment of a Geospatial Web. For the purposes of this discussion, the Geospatial Web is an integrated, discoverable collection of geographically related Web services and data that spans multiple jurisdictions and geographic regions. In a broad sense, the Geospatial Web refers to the global collection of general services and data that support the use of geographic data in a range of domain applications. Regional and/or domain-specific expressions of the global Geospatial Web exist as well. The global, national, state/provincial or local Spatial Data Infrastructure, or SDI (NRC 1993), are each instances of the Geospatial Web. Like the Internet, which is composed of many local extranets and intranets, the Geospatial Web rests on a common framework of open standards and standards-based technologies. The importance of such open platforms is firmly established (Cargill 1997). This discussion provides a clear description of the Geospatial Web and its role. Key standards and capabilities are highlighted. The notion of these standards-based technologies as an *infrastructure for the Geospatial Web* is introduced. Specific examples of real-world applications being deployed on the Geospatial Web using this infrastructure are discussed.

At the outset there needs to be a clear differentiation between geographic data and map images. While maps may be the most commonly recognized product associated with geospatial data and applications, they are just that: *one product at the end of a long supply chain*. This supply chain acquires and fabricates geospatial data to develop new information, to make decisions, to assist in a broad-range of modeling and simulation and for many other purposes, one of which is to create a map. Existing standards such as WMS support the reliable delivery of map images in an interoperable framework. As such they are a component in the fabric of the Geospatial

Web. However, the realization of the Geospatial Web requires a much richer set of features distributed over a broader, interjurisdictional audience. It is this richer set of features, the standards that support the requirements implied by these features and this broader audience that occupy the remainder of this discussion.

2.2 Organizations, Integration and Data

Critical to the idea of data integration across jurisdictional or administrative boundaries is the recognition that business processes within each jurisdiction or administrative unit are fundamentally autonomous. While there might be changes in corporate or government departmental organization, our assumption is fundamentally that each organization acquires, analyzes and deploys geographic and geographically related information in order to deal with a business issue confronting that organization. Notions of information sharing that depend on new cross-organizational business processes or that demand the integration of such processes are, in our view, doomed to failure, since they conflict with the basic needs of the organization itself. Cross-organizational integration and information sharing can only succeed if they are accommodated within and transparent to the core business processes of the organization itself. Only when information sharing is achieved on the basis of such organizational autonomy should we consider integration that depends on integrated business processes. This leads to two basic premises regarding the use of geographic information in and between organizations: (i) the acquisition and use of geographic information is driven by real-world business problems; (ii) new business processes will not succeed if they are created solely to support artificial cross-organizational integration (integration for the sake of integration).

2.2.1 Primary Resources and Objects of Interest

If we begin with the principle of organizational autonomy expressed above, we must consider the persistent information stores that support the organization's business process(es) as the primary resource. Given this, the applications that update, process and display that information (underwriting organizational decision making) constitute an essential secondary class of resource. The persistent information stores contain the objects of primary interest to the organization. These objects are modeled and represented in ways that address the organization's immediate and long-term concerns and objectives. This basic set of relationships and operational dynamics is independent of organizational size and exists in application domains that include resource exploitation, transportation security, environmental protection, or the planning or operation of new urban infrastructure.

These primary resources or objects of interest can be very dynamic. They might be generated or modified by an organization as a result of its core business processes; e.g., tax parcels, property ownership and the location and condition of mobile assets. Primary resources can also be effectively static, seldom updated but providing critical input for data processing and decision making. For instance, a police department will clearly want to update objects that relate to crime incidents, traffic accidents or the whereabouts of serious criminals. However, it is unlikely that they would be interested in recapturing the layout of buildings or roadways. (Note, however, that the police may close or block a road or deny access to a building). The same police department will want to have access to additional information that they do not "own" or update such as the location of fires, major civic events and other items that could become of concern to the police. In a similar manner, the fire department has a primary interest in the location of fires and hazardous materials and would be responsible for the update of these objects and would have an interest – but not an update interest – in objects such as the location of roads, buildings and large civic events. In the former case the fire department acts as the primary custodian and/or author or authority for specific data (e.g., locations of fires, profiles and locations of hazardous materials), while in the latter instance the fire department's role is primarily that of a consumer, incorporating relevant data produced by other organizations (e.g., road networks, address ranges, building footprints and construction materials) into models and workflows to improve decisions.

One should also note that objects updated by one organization may significantly impact the business processes of other organizations. For example, the occurrence of a large fire (primary interest of the fire department) will quickly become of interest to the police or other traffic management organizations.

2.2.2 Objects and Roles

While these issues have been expressed in terms of organizations, however, they may be more accurately considered in terms of roles. Roles are usually mapped to a particular organization, but this need not be the case in all circumstances. In most application domains we will recognize the notion of an authorized observer, meaning someone who can report the existence of or change in some object, which they are not primarily responsible for, the reporting typically taking place to the responsible organization. Thus, we have the common citizen able to report a fire to the fire department or an emergency of some sort to a centralized 911 service or contacting the police to report a traffic accident.

2.2.3 Standards for the Geospatial Web

The Geospatial Web is ultimately enabled by widely adopted, open standards. These standards emerge primarily from the mainstream Web communities (XML, W3C, etc.) and from industry consortia that focus on specific functional areas or topical areas. For instance, the OASIS Consortium concentrates on Registries and Registry Services in its specification (Fuger et al. 2005). The Open Geospatial Consortium (OGC) has had a committed global membership that has worked for more than 10 years to specify a framework supporting interoperability in geographic applications. To underscore the global nature of these initiatives, many of the specifications are rationalized under the umbrella of the International Standards Organization (ISO).

With this background in hand, we can now move on to the consideration of the key standards and technologies that enable information sharing and deliver infrastructure for the Geospatial Web.

2.3 GML: A Lingua Franca for the Geospatial Web

A common language capable of expressing geographic information is required to enable information to be shared in the various ways discussed above. The Geography Markup Language (GML) provides such a language (Lake et al. 2004).[28] As such, GML provides a fundamental infrastructure that enables the Geospatial Web.

2.3.1 GML Basics

GML is an XML language for the encoding of geographic and geographically related information. While it includes geographic information, actually any information can be encoded in GML, and there is no requirement that the information be related to location or time.

GML is written in XML Schema, which delivers inherent extensibility to GML. Users of GML create their own object vocabulary (i.e., object types) by writing GML Application Schemas. These XML Schemas make use of GML schema components (e.g., time, geometry, etc.) and follow simple, structural rules for GML. The ability of users to create more or less any object in GML is essential. It is this capacity that enables GML to function both as an information transport and as a means of exposing the information model of a chosen persistent store. These capabilities are required to support information requests and transactions.

2.3.2 Type Definitions and Encoding

GML can be viewed either in terms of type definitions (schema components) or in terms of the instance data itself. Which one is more important depends on the specific application (Galdos Systems 2003). For example, if a Web service is being specified, GML can be used to define the types of the arguments in the inputs and outputs of the Web service, even if GML is not used in the actual data transport.

GML can also be viewed as an XML encoding of an extended E-R diagram, representing entities and their attributes, attribute inheritance and relationships between entities. Related entities (or classes) can be related (associated) regardless of their relative location. For example, a road object can say that it crosses a bridge object even when the respective objects are located in different spatial databases and belong to different organizations. In GML this is stated in terms of the Object-Property-Value rule, as illustrated in Figure 2.1.

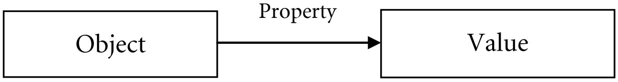


Figure 2.1: GML object-property-value relationship

The GML Object-Property-Value rule determines the structure of a GML (XML) instance document. GML Objects are those whose content model derives from a content model in the GML core schemas (GML namespace). Their children are always properties of the GML Object, and such properties can be either attributes or express relationships between objects (e.g., if the value is a GML object, the property expresses an object relationship).

2.4 Working with GML

GML was devised in order to support a fine-grained, feature-based relationship between geographic (and other) databases. GML is specified to support both transactions and requests. GML is not simply another file encoding that supports ad hoc file exchange (Galdos Systems 2003; Lake et al. 2004).

GML provides a rich collection of primitives for the encoding of time, geometry, topology, coordinate reference systems, units of measure, map styling, observations, coverages and general geographic features. This means that a GML encoding can