Special Section

The Semantic Web

- The Semantic Web: More than a Vision
- An Overview of W3C Semantic Web Activity
- Semantic Web Services
- Metadata: A Fundamental Component of the Semantic Web
- Ontologies and the Semantic Web

Features

Complex Acts of Knowing: Paradox and Descriptive Self-Awareness

Columns

What's New?

BULLETIN

of the American Society for Information Science and Technology *April/May 2003*

Volume 29, No. 4 ISSN: 0095-4403 CODEN:BASICR

APRIL/MAY 2003

BULLETIN

of the American Society for Information Science and Technology

VOLUME 29, NO. 4

SPECIAL SECTION

The Semantic Web

Features

Columns

Departments

The *Bulletin of the American Society for Information Science and Technology*, ISSN 0095-4403, is published bi-monthly, October through September, by the American Society for Information Science and Technology, 1320 Fenwick Lane, Suite 510, Silver Spring, MD 20910; 301/ 495-0900; Fax: 301/495-0810; e-mail: asis@asis.org; http://www.asis.org

POSTMASTER: Send address changes to the *Bulletin of the American Society for Information Science and Technology*, 1320 Fenwick Lane, Suite 510, Silver Spring, MD 20910. Periodicals postage paid at Silver Spring, MD.

The subscription rate for ASIST members is $19, which is included in the annual membership dues. Non-member subscriptions, and additional member subscriptions, are $60 per year U.S., Canada and Mexico; $70 per year other, postpaid in the U.S. Single copies and back issues of the *Bulletin* may be purchased for $10 each. Claims for undelivered copies must be made no later than three months following the month of publication.

The American Society for Information Science and Technology (ASIST) is a non-profit professional association organized for scientific, literary and educational purposes and is dedicated to the creation, organization, dissemination and application of knowledge concerning information and its transfer.

The official ASIST journal is the *Journal of the American Society for Information Science and Technology,* published for the Society by John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158.

The *Bulletin of the American Society for Information Science and Technology* is a news magazine concentrating on issues affecting the information field; management reports; opinion; and news of people and events in ASIST and the information community. Manuscripts are welcomed and reviewed for possible publication. Articles should not exceed 1500 words and may be accompanied by appropriate artwork. All manuscripts are subject to editing. Care is taken in the handling of manuscripts and artwork; however, the *Bulletin* cannot assume responsibility for lost or damaged material or for the return of unsolicited manuscripts. Send manuscripts to the *Bulletin of the American Society for Information Science and Technology*, 1320 Fenwick Lane, Suite 510, Silver Spring, MD 20910. Because the *Bulletin* is a news magazine, authors do not review galleys before publication.

Opinions expressed by contributors to ASIST publications do not necessarily reflect the official position of either their employers or the Society.

Trudi Bellardo Hahn
2003 ASIST President
User Education Services
University of Maryland
Libraries
1103 McKeldin Library
College Park, MD 20742
301-405-9254
th90@umail.umd.edu

What Has Information Science Contributed to the World?

The Council of Scientific Society Presidents recently asked me to respond to a survey question: "What were the most important seminal five to seven discoveries in the field represented by your professional society in the 20th century?" Such a question raises several complex issues, such as what are the most remarkable achievements unique to the field of information science in the past 100 years? Who are the individuals who were responsible for each one? Just what constitutes our field as separate from other fields such as computer science, librarianship, chemistry, engineering, medicine, management, law or education? How do our research methods differ from those of the social sciences, operations research, linguistics and others from which we have obviously borrowed?

Since I could not answer the survey question off the top of my head, I consulted ASIST members who research and write the history of information science. Michael Buckland, Eugene Garfield, Julian Warner and Robert Williams replied. It appeared that "developments" is more apt to describe information science activities than "discoveries."

However, their responses appeared to have discouragingly little consensus or overlap.

By merging their responses into larger categories and consulting some information science textbooks and historical papers, I drafted a list of five major categories of accomplishment that I believe can be attributed directly and solely (well, nearly) to IS researchers and developers.

1. Information science researchers measured the information explosion. They created the field of bibliometrics – the study of published literature and its usage. Bibliometrics has many aspects, including studies of impact, diffusion of innovation, bibliographic coupling, citation and co-citation patterns and other statistical regularities in scientific and scholarly productivity and communication.

2. Information science developers contained the information explosion. Information scientists pioneered innovations in indexing systems that were very different from traditional subject cataloging in libraries – automatic indexing and abstracting, KWIC and KWOC indexing, citation indexing, keyword indexing and post-coordination, text analysis and natural language searching systems. They also developed thesauri or controlled vocabularies for thousands of disciplines and specialties.

3. Information science developers applied computers to manipulating documents and document records in information storage and retrieval systems. This began almost as soon as computers became available in the 1950s, but really took off with third-generation computers in the 1960s. The development of online database systems was accompanied by related telecommunications and networking technologies and specialized search functionalities, as well as large machine-readable databases. The application of formal logic (Boolean operators) to database searching was a major component of these developments.

4. Information science researchers studied users' information seeking, needs and preferences, as well as related areas such as relevance and utility assessment. The sociologists got us started,

(Continued on page 3)

From the Editor's Desktop

Irene L. Travis, Editor
Bulletin of the American Society for Information Science and Technology
Bulletin@asis.org

Feeling insecure about the Semantic Web? This issue will solve your problem. My special thanks go to Jane Greenberg of the University of North Carolina for serving as guest editor and to the six other authors for four excellent and helpful articles on this exciting and expanding area of exploration and development. This issue concludes our two-part treatment of metadata, which we began in the October/November 2002 *Bulletin*.

We have one other major article, a condensation of a piece by Dave Snowden from the *Journal of Knowledge Management*. Dave was a most stimulating and entertaining keynote speaker at 2002 Annual Meeting, and I thank him for letting me chop this very dense publication down to a size we could accommodate. Despite my valiant efforts, I encourage those of you who find it intriguing to read the original and more nuanced version, which is referenced with the article and available on the Web.

Finally Trudi Bellardo Hahn has a true challenge for us. One of her recent tasks as ASIST President was to respond to this query from the Council of Scientific Society Presidents: "What were the most important seminal five to seven discoveries in the field represented by your professional society in the 20th century?" She tackles this daunting question with her usual insight and enthusiasm, shares her thoughts with us and solicits ours.

Andrew Dillon's IA Column will be back next issue with expanded comment on the IA Summit in Portland.

President's Page

(Continued from page 2)

but we quickly developed our own body of research in the second half of the last century.

5. Information science leaders in government and industry contributed to formulating national information policies related to issues of privacy, security, regulating dissemination and access, intellectual property, acceptable use and others. They contributed to developing standards for the processing and communication of information, as well as the monitoring of the national information infrastructure (human, technological, materials and financial) to ensure that information systems and services related to the public interest were maintained.

ASIST members are invited to debate the content of this list, to suggest additions or items that should have high priority, to identify the pioneers and to date seminal discoveries, developments or inventions. We know we are multidisciplinary and cross-disciplinary, but I believe there is a core of knowledge and developments that is uniquely ours – if we can but define it.

I have asked Robert Williams, University of South Carolina, to work with members of the Special Interest Group on History and Foundations of Information Science to refine and expand this list. He has already started the process by compiling a draft of a detailed chronology of information science and technology available at www.libsci.sc.edu/bob/istchron/ISCNET/Ischron.htm. Please help by sending your thoughts and suggestions to Bob (bobwill@sc.edu).

Our goal is to publish an authoritative list of accomplishments on the ASIST website. In addition to the existing "About ASIST" page and the mission and vision statements, it will show ASIST members and potential members what this field is about, what it values and where the greatest potential for future discoveries and contributions lies.

Inside ASIST

ASIST 2003 Annual Meeting

Humanizing Information Technology: From Ideas to Bits and Back

Time to start making your plans for the 2003 ASIST Annual Meeting. This year's gathering will be in Long Beach, California, October 20-23, and will focus on the broad and timely theme of *Humanizing Information Technology: From Ideas to Bits and Back.*

Though presentation submissions for consideration at the meeting were just being evaluated at press time for this issue of the *Bulletin of the American Society for Information Science and Technology*, the program committee has identified the following topics as those on which the meeting will focus:

- **Information management, organization and access:** classification and representation, metadata, taxonomies, indexing, XML, information architecture, digital libraries
- **Information seeking and use:** research on the role of information in daily life and work, use of various types of information technology, social contexts of information seeking
- **Information retrieval:** information system performance, search engines, natural language processing, data mining, intelligent retrieval, multi- and cross-lingual retrieval
- **Interactivity:** user and usability studies, design of human-computer interfaces, visualization
- **Ethical, social, political, legal and economic issues:** privacy, copyright, information policy, the social role of information technologies
- **Information production, transfer and delivery:** electronic publishing and dissemination
- **Technologies for computing and networking:** information communication, collaboration, information security, e-commerce

Conference themes will be explored through refereed papers, posters, panels and technical sessions.

Conference Committee

Marcia J. Bates, UCLA, is chair of the conference committee. The following people are assisting her on the committee: **Eileen Abels, Suresh Bhavnani, Michael Buckland, Donald Case, Chao-chen Chen, Andrew Dillon, Efthimis N. Efthimiadis, Raya Fidel, Jonathan Furner, Andrew Grove, Jenna Hartel, Sandra Hirsh, Joseph W. Janes, Don Kraft, Carol Kuhlthau, Marianne Nelson, Hope Olson, Carole Palmer, Jaime Pontigo, Dragomir Radev, Nancy Roderer, Victor Rosenberg, Linda Rudell-Betts, Bernie Sloan, Ross Todd, Irene Travis, Peiling Wang, Carolyn Watters, Judith Weedman** and **Barbara Wildemuth**.

Local Activities

The Los Angeles Chapter of ASIST is managing all local arrangements for the 2003 Annual Meeting. Beginning in the last issue of the *Bulletin*, with an article by **Bo-Gay Tong Salvador**, and continuing in this issue with the contribution of **Linda Heichman**, LACASIS members will offer regular reports on southern California attractions to lure you to their neck of the woods this fall. In the meantime, you can keep up with all technical and social news about the meeting at the ASIST website.

News from ASIST Chapters

The **Northern Ohio ASIST (NORASIST) Chapter** began the new year with a panel discussion of *The Digital Divide: Research, Thoughts and Action*, co-sponsored by the Cleveland Public Library. Featured speakers were Earnestine Adeyemon, Case Western Reserve University; Joan L. Clark, Cleveland Public Library; and Mary Stansbury, Kent State University.

For the following month's program, **NORASIST** planned *Applications of Full Motion Screen Capture for Library Instruction*, a panel discussion on the use of full motion screen capture and streaming video to record and deliver library instruction, survey current screen capture software and present ideas for a cooperative project. Planned speakers included **Richard Brhel**, director of the Library Resource Center, Myers University, and Ken Burhanna, Instructional Design Librarian, Cleveland State University.

The **Potomac Valley Chapter** presented *The Architecture of Search Engines* as its February meeting at the University of Maryland. **Denise Bedford**, Thesaurus Project Manager, Information Solutions Group, World Bank, Washington DC, looked at some of the most popular search engines and discussed their similarities and differences and looked ahead to see how search engines will improve in the future.

The ASIST Seattle Reading Group of the **Pacific Northwest Chapter** also took a look at search engines for its February meeting with a discussion of Google as a cultural entity. The basis of the group's discussion

LOOKING TO LONG BEACH

Long Beach – Hidden Treasure Trove of Arts & Culture

by Linda Heichman

Imagine a city on the beautiful California coast, a city alive with a bustling art scene – museums, galleries, monthly art walks, symphony, theater and opera. San Francisco? Think again. Los Angeles? Nope. San Diego? Nah. I know, Carmel. Good try, but no. It's Long Beach.

Not only does Long Beach boast an eclectic art scene, the city is home to world-class art museums, internationally renowned theater companies, its own symphony orchestra, opera company and master chorale. Museums include the Long Beach Museum of Art, housed in a Craftsman mansion overlooking the Pacific Ocean; the Museum of Latin American Art, the Western United States' premier museum of Spanish and Caribbean arts and culture; and the University Art Museum at California State University, Long Beach (CSULB).

Performing arts abound in Long Beach. Choose from the Long Beach Symphony Orchestra, check out budding talent at California State University, Long Beach's Carpenter Performing Arts Center, Cal Rep, International City Theatre or Long Beach Playhouse.

Want more? Check these out.

Visual Arts

East Village Arts District – www.eastvillageartsdistrict.com
Long Beach Museum of Art – www.lbma.org
Museum of Latin American Art – www.molaa.com
University Art Museum – www.csulb.edu/org/uam/

Performing Arts

Cal Rep at Edison Theatre – www.calrep.org/
Carpenter Performing Arts Center – www.carpenterarts.org
International City Theatre – www.ictlongbeach.com/
Long Beach Playhouse – www.longbeachplayhouse.com/
Long Beach Symphony – www.lbso.org

was the article "Google vs. Evil" at www.wired.com/wired/archive/11.01/google.html. Optional reading for the discussion were "Google Runs Into Copyright Dispute" at www.nytimes.com/2002/04/22/technology/ebusiness/22NECO.html and "Google May Remove Controversial Links" at www.internetnews.com/bus-news/article.php/1009321

The first **Central Ohio ASIST (CO-ASIST)** meeting of the year featured Patrick Losinski, new executive director of the Columbus Metropolitan Library, sharing his insights about *Public Libraries: Empowering Staff and Managing Technology in Turbulent Times.*

CO-ASIST then followed it up with a March meeting featuring Richard Rubin, interim dean, College of Communication & Information at Kent State University, with a presentation on *Who's Driving, Who's Riding? – The Place of Technology and People in the Workplace.*

The **Southern Ohio Chapter of ASIS (SOASIS)** offered a March presentation on *Information Seeking and Information Avoidance: The Case of Patients and Health Information*, featuring **Donald O. Case**, professor, University of Kentucky, and the author of *Looking for Information*, Academic Press.

News about ASIST Members

Vicki L. Gregory, professor in the School of Library and Information Science at the University of South Florida, has received the President's Award for Excellence, a special merit program at USF. Along with two SLIS colleagues, Vicki was recognized for her role in the SLIS accreditation review in Spring 2002, research and publication, and professional service.

Blaise Cronin, dean of Indiana University's (IU) School of Library and Information Science (SLIS) since 1991, has announced his intention to step down as dean, effective June 30, 2003. After a year's sabbatical leave, he will return to the faculty to pursue his many research and other interests. Cronin, editor of ASIST's *Annual Review of Information Science and Technology*, has served as dean for 12 years.

News from an ASIST SIG

SIG/III has taken a look at the career paths of some of the past winners of the SIG/III International Paper Contest and has been pleased to see the recognition and career advancement of the several of them, including those that follow.

Duncan Wambogo Omole, a 2000 winner from Kenya, is now an information analyst at the World Bank in Washington, DC. He will be working with the World Bank/IMF Library Network Global Outreach Group to provide training and support for the network of World Bank country office libraries in developing countries.

Jagdish Arora, a 2001 winner from India, has been appointed librarian, Central Library, Indian Institute of Technology, Bombay. He was previously head of computer applications at the Central Library, Indian Institute of Technology, New Delhi.

Ismail Fahmi, 2001 winner from Indonesia, received the Most Favorite Program award for the Indonesian Digital Library Network from the Indonesian Business Community in 2002. The award is for companies, organizations or institutes that conduct business or social activities that have a positive impact for ICT development in Indonesia.

P. R. Goswami, 2001 winner from India and also assistant chair (outside the United States) of SIG/III, is now a director of the Indian Council of Social Science Research (ICSSR). His responsibility is to manage its documentation and information dissemination activity.

Introduction
The Semantic Web: More than a Vision

by Jane Greenberg, Guest Editor

Jane Greenberg is an assistant professor in the School of Information and Library Science at the University of North Carolina, Chapel Hill, principal investigator of the Metadata Generation Research project (http://ils.unc.edu/~janeg) and program committee co-chair of the 2003 Dublin Core Conference. \She can be reached by e-mail at janeg@ils.unc.edu

Our dependence on World Wide Web (Web) technology for information, communication and services grows daily. Consider the slightly frantic behaviors people often exhibit when they are unable to access the Web for an extended period of time. Of course there is the other side – a break from interacting with a computer is viewed as a relief for the eyes. Even so, it is clear that our information society is becoming wedded

to Web technology for daily activities. The proliferation of library and information science publications addressing and researching aspects of the Web – still a relatively new phenomenon – provides even further evidence of our dependence on this technology. In fact it's difficult, if not impossible, to find an information science periodical without one article dealing with Web technology.

While all this is exciting, there are many limitations to the current Web. Visionaries and researchers throughout time have talked about exploiting our mass of information to automatically produce new knowledge, build intelligent systems and eliminate human burdens associated with information seeking and problem solving activities. There have been successes, but they are often limited by domain or infrastructure. The Web offers us a new playing field for addressing these goals through the *Semantic Web*, which is an extension aiming to foster communication between computers and people via semantically encoded information.

This special section includes four articles about the Semantic Web. A great deal of the Semantic Web activity is taking place at the World Wide Web Consortium (W3C). In this review of the field, Eric Miller and Ralph Swick provide an overview of W3C Semantic Web activities. They discuss Semantic Web enabling technologies and important Semantic Web Advance Development (SWAD) initiatives. These include SWAD DAML, SWAD-Europe, SWAD Simile and SWAD Oxygen.

Bijan Parsia focuses on Semantic Web services (remote programs). Parsia outlines the shortcomings of the current Web, explaining why current services are severely limited and how they could be improved. Attention is specifically given to the problems of service discovery. Parsia explains current efforts to solve this problem with Universal Description, Discovery and Integration of Web Services (UDDI) and demonstrates the significant role of semantics in problem solving. This article draws from work currently being conducted in MIND's Semantic Web Agents Project at the University of Maryland, College Park.

Jane Greenberg, Stuart Sutton and D. Grant Campbell address the fundamental role that metadata plays in building the Semantic Web. We discuss the vision and architecture underlying the Semantic Web and explain how each layer of the Semantic Web's architecture, as envisioned by Tim Berners-Lee, is connected to or directly involves metadata. Topics include metadata vocabularies, enabling technologies and Semantic Web authoring and annotation. We find ourselves, in some respects, as early pioneers exploring the potential roles and forms of metadata related to the Semantic Web's emerging architecture.

The technology underlying the Web is unprecedented and affords us new opportunities to turn segments of the growing mass of electronic information into new intelligence for both humans and computers.

Elin Jacob's article concludes this special section with an article on ontologies. Jacob offers a philosophical and practical discussion of ontologies and their roles in building the Semantic Web. Specific attention is given to ontology languages, such as RDFS (Resource Description Framework Schemas) and OWL (Web Ontology Language) and their application to the Semantic Web. Jacob urges us to think outside the box and realize that there are indeed new capabilities that we need to explore.

To pick up on Jacob's remarks, I have heard people say the Semantic Web is "old wine in a new bottle." There is likely some truth here, as is always the case with innovations drawing upon developments and ideas from earlier times, but I agree with Jacob's line of thinking. The technology underlying the Web is unprecedented and affords us new opportunities to turn segments of the growing mass of electronic information into new intelligence for both humans and computers. The Semantic Web is an engaging territory to explore and cultivate.

Special Section

An Overview of W3C Semantic Web Activity

by Eric Miller and Ralph Swick

Both authors of this article are with the World Wide Web Consortium (W3C). Eric Miller is Semantic Web Activity Lead and can be reached by e-mail at em@w3.org; Ralph Swick is Technology and Society Domain Technical Lead, e-mail: swick@w3.org.

The Semantic Web is an extension of the current Web in which the meaning of information is clearly and explicitly linked from the information itself, better enabling computers and people to work in cooperation. The World Wide Web Consortium (W3C) Semantic Web Activity, in collaboration with a large number of researchers and industrial partners, is tasked with defining enabling standards and technologies to allow data on the Web to be defined and linked in such a way that it can be used for more effective discovery, automation, integration and reuse across various applications. The Web can reach its full potential if it becomes a place where data can be shared and processed by automated tools as well as by people.

The Semantic Web fosters and encourages greater data reuse by making it available for purposes not planned or conceived by the data provider. Suppose you want, for example, to locate news articles published in the previous month about companies headquartered in cities with populations under 500,000 or to compare the stock price of a company with the weather at its home base or to search online product catalogs for an equivalent replacement part for something. The information may be there in the Web, but currently only in a form that requires intensive human processing.

The Semantic Web will allow two things. First, it will allow this information to surface in the form of data, so that a program doesn't have to strip the formatting, pictures and ads off a Web page and guess at how the remaining page markup denotes the relevant bits of information. Second, it will allow people to write (or generate) files that explain – to a machine – the relationship between different sets of data. For example, one will be able to make a "semantic link" between a database with a "zip-code" column and a form with a "zip" field to tell the machines that they do actually mean the same thing. This will allow machines to follow links and facilitate the integration of data from many different sources. When the relationships among data are fully accessible to our machines, our machines will be able to help us browse those relationships and interpret the data as well as assess the appropriateness of the data for our intended purposes.

This notion of being able to "semantically link" various resources, such as documents, images, people or concepts, is an important one. With semantic links we can move from the current Web of simple relationships like "links-to" to a more expressive, semantically rich Web – a Web where we can incrementally add meaning and express a whole new set of relationships (hasLocation, worksFor, isAuthorOf, hasSubjectOf, dependsOn, etc.). These relationships can make explicit the particular contextual relationships that are either implicit or expressed in the current Web only in prose that is impossible for machines to interpret. This enhancement in turn opens doors for a whole new set of effective information integration, management and automated services.

The Semantic Web is a place where strongly controlled (or centralized) metadata vocabulary registries can flourish alongside special-purpose, small community or even "private" vocabularies. The Semantic Web technology supports free comingling of vocabularies as well as the ad-hoc definition of new relationships to construct data descriptions. In addition, instructions for processing data in specific ways can be expressed in the Semantic Web using the same technologies used to describe the data. So discovery mechanisms that work for data will also work for procedures to operate on the data. Trust mechanisms to permit an application to evaluate whether specific data or procedures are suitable for use in a given context are simply more data and relationships in the Semantic Web architecture; that is, they are an integral part of the Semantic Web vision.

The development of the Semantic Web is well underway in at least two very important areas: (1)

from the infrastructural and architectural position defined by W3C and (2) in a more directed application-specific fashion by those leveraging Semantic Web technologies in various demonstrations, applications and products. This article provides a brief introduction to both of these developmental areas with a specific focus on those in which the W3C is directly involved.

More information on the Semantic Web, including additional projects, products, efforts and future directions, is available on the Semantic Web home page (www.w3.org/2001/sw/).

Enabling Standards

Uniform Resource Identifiers (URIs) (www.w3.org/Addressing/) are a fundamental component of the current Web and are in turn a foundation for the Semantic Web. URIs provide the ability for uniquely identifying resources of all types – not just Web documents – as well as relationships among resources. An additional fundamental contribution toward the Semantic Web has been the development of the Extensible Markup Language (XML) (www.w3.org/XML/). XML provides an interoperable syntactic foundation upon which the languages to represent relationships and meaning are built. The Resource Description Framework (RDF) (www.w3.org/RDF/) family of languages leverages XML, URIs and the Web to provide a powerful means of expressing and representing these relationships and meaning.

The W3C Semantic Web Activity (www.w3.org/2001/sw/) plays a leadership role in both the design of specifications and the open, collaborative development of technologies focused on representing relationships and meaning and the automation, integration and reuse of data. The base level RDF 1.0 standard was defined in 1999. RDF 1.0 and RDF Schema (RDF Vocabularies) are currently being refined based on implementation experience, and more expressive higher layers are being addressed.

The base level standards for supporting the Semantic Web are currently being refined by the RDF Core (www.w3.org/2001/sw/RDFCore/) Working Group. This group is chartered to revise and formalize the original RDF Model and Syntax Recommendation (www.w3.org/TR/1999/REC-rdf-syntax-19990222/), which provides a simple, yet powerful, assertional framework for representing information in the Web. Additionally, this group is tasked to layer upon this general descriptive framework a simple means for defining RDF Vocabularies (www.w3.org/TR/rdf-schema/). RDF Vocabularies are descriptive terms such as *service, book, image, title, description* or *rights* that are useful to communities interested in recoding information in a way that enables effective reuse, integration and aggregation of data. Additional deliverables include a precise semantic theory (www.w3.org/TR/rdf-mt/) associated with these standards useful for supporting future work, as well as a primer (www.w3.org/TR/rdf-primer/) designed to provide the reader the basic fundamentals required to effectively use RDF in their particular applications.

The Web Ontology (www.w3.org/2001/sw/WebOnt/) Working Group standards efforts are designed to build upon the RDF core work a language, OWL (www.w3.org/TR/owl-ref/), for defining structured, Web-based ontologies. Ontologies can be used by automated tools to power advanced services such as more accurate Web search, intelligent software agents and knowledge management. Web portals, corporate website management, intelligent agents and ubiquitous computing are just some of the identified scenarios (www.w3.org/TR/webont-req/) that helped shape the requirements for this work.

Semantic Web Advanced Development (SWAD)

Code modules such as libwww (www.w3.org/Library/) accelerated the early deployment of the Web, and to a similar end the W3C is devoting resources to the creation and distribution of components to assist in the deployment of the Semantic Web.

These W3C Semantic Web Advanced Development initiatives are designed to work in collaboration with a large number of researchers and industrial partners to stimulate various complementary areas of development that will help facilitate further deployment and future standards work associated with the Semantic Web.

SWAD DAML. SWAD DAML is a project within the Defense Advanced Research Project Agency (DARPA) Agent Markup Language (DAML) (www.daml.org/) Program. The SWAD DAML (www.w3.org/2000/01/sw/daml) project combines research and development to define the architectural layering of the languages of the Semantic Web infrastructure. SWAD DAML builds critical components of that infrastructure and demonstrates how those components can be used by practical, user-oriented applications. It both seeks to define a logic language framework on top of RDF and the OWL vocabulary and to build basic tools for working with RDF, OWL and this logic framework.

To demonstrate some practical applications of these tools to manipulate structured information, SWAD DAML is deploying them to maintain the ongoing activities of the W3C, including access control, collaboration, document workflow tracking and meeting management. Another component of SWAD DAML is focused on the informal and often heuris-

tic processes involved in document management in a personalized information environment. Integrated into SWAD DAML will be tools to enable authors to control terms under which personal or sensitive information is used by others, a critical feature to encourage sharing of semantic content.

SWAD-Europe. SWAD-Europe (www.w3.org/2001/sw/Europe/) aims to highlight practical examples of where real value can be added to the Web through Semantic Web technologies. The focus of this Advanced Development initiative is on providing practical demonstrations of how (1) the Semantic Web can address problems in areas such as sitemaps, news channel syndication, thesauri, classification, topic maps, calendaring, scheduling, collaboration, annotations, quality ratings, shared bookmarks, Dublin Core (http://dublincore.org/) for simple resource discovery, Web service description and discovery, trust and rights management and (2) effectively and efficiently integrate them.

The focus of the SWAD-Europe deliverables are to exploit the enabling standards that have already been developed and not to depend upon future technologies identified with the Semantic Web architecture. Thus, the SWAD-Europe work is demonstrating the potential of what can be built on existing Semantic Web standards.

SWAD-Europe will additionally engage in exploratory implementation and pre-consensus design in such areas as querying and the integration of multiple Semantic Web technologies. This effort will provide input and experiences to future standards work.

SWAD Simile. Under the SWAD initiatives, W3C is also working with Hewlett-Packard (www.hp.com/), Massachusetts Institute of Technology (MIT) Libraries (http://libraries.mit.edu/), and MIT's Laboratory for Computer Science (MIT LCS) (www.lcs.mit.edu/) on Project Simile (http://web.mit.edu/simile/www/). Simile seeks to enhance interoperability among digital assets, schemas, metadata and services across distributed individual, community and institutional stores and across value chains to provide useful end-user services by drawing upon the assets, schemas and metadata held in such stores. Simile will leverage and extend DSpace (http://dspace.org/), also developed by MIT and HP, enhancing DSpace's support for arbitrary schemas and metadata, primarily through the application of RDF and Semantic Web techniques. The project also aims to implement a digital asset dissemination architecture based upon Web standards, enabling services to operate upon relevant assets, schemas and metadata within distributed stores.

The Simile effort will be grounded by focusing on well-defined, real-world cases in the libraries domain. Since parallel work is underway to deploy DSpace at a number of leading research libraries, we hope that such an approach will lead to a powerful deployment channel through which the utility and readiness of Semantic Web tools and techniques can be demonstrated compellingly in a visible and global community.

SWAD Oxygen. The Oxygen Project (http://oxygen.lcs.mit.edu/), a joint effort of the MIT LCS and the MIT Artificial Intelligence Laboratory (MIT AI), is designed to make pervasive, human-centered computing a reality through a combination of specific user and system technologies. Oxygen's user technologies directly address human interaction needs: automation (http://oxygen.lcs.mit.edu/Automation.html), individualized knowledge access (http://oxygen.lcs.mit.edu/KnowledgeAccess.html) and collaboration (http://oxygen.lcs.mit.edu/Collaboration.html) technologies help us perform what we want to do in the ways we like to do them. In Oxygen, these technologies enable the formation of spontaneous collaborative regions that provide support for recording, archiving and linking fragments of meeting records to issues, summaries, keywords and annotations.

A goal of the Semantic Web is to foster similar collaborative environments – human-to-human and human-to-machine – and the W3C is working with project Oxygen to help realize this goal. The ability for "anyone to say anything about anything" is an important characteristic of the current Web and is a fundamental principal of the Semantic Web. Knowing who is making these assertions is increasingly important in trusting these descriptions and enabling a "Web of Trust." The Annotea (www.w3.org/2001/Annotea/) advanced development project provides the basis for associating descriptive information, comments, notes, reviews, explanations or other types of external remarks with any resource. Together with XML digital signatures, the Annotea project will provide a test-bed for "Web-of-Trust" Semantic Web applications.

Education and Outreach

To fulfill its leadership role and facilitate the effectiveness and efficiency of the W3C Semantic Web Activity, a strong focus on education and outreach is important. The RDF Interest Group (www.w3.org/RDF/Interest/) continues to be an extremely effective forum in which developers and users coordinate public implementation, share deployment experiences of RDF and help each other promote the Semantic Web.

Arising out of RDF Interest Group discussions are several

The Semantic Web provides an infrastructure that enables not just Web pages, but databases, services, programs, sensors, personal devices and even household appliances to both consume and produce data on the Web.

public issue-specific mailing lists, including RDF-based calendar and group scheduling systems (http://lists.w3.org/Archives/Public/www-rdf-calendar/), logic-based languages (http://lists.w3.org/Archives/Public/www-rdf-logic/), queries and rules for RDF data (http://lists.w3.org/Archives/Public/www-rdf-rules/) and distributed annotation and collaboration (http://lists.w3.org/Archives/Public/www-annotation/) systems. Each of these discussion groups focuses on complementary areas of interest associated with the Semantic Web activity.

Future education and outreach plans include the formation of a Semantic Web education and outreach group designed to develop strategies and materials to increase awareness among the Web community of the need for and benefits of the Semantic Web and to educate the Web community regarding best practice solutions and enabling technologies associated with the Semantic Web.

Conclusion

The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. It is based on the idea of having data on the Web defined and linked such that it can be used for more effective discovery, automation, integration and reuse across various applications.

The Semantic Web provides an infrastructure that enables not just Web pages, but databases, services, programs, sensors, personal devices and even household appliances to both consume and produce data on the Web. Software agents can use this information to search, filter and prepare information in new and exciting ways to assist Web users. New languages make significantly more of the information on the Web machine-readable to power this vision and will enable the development of a new generation of technologies and toolkits.

The seeds of the Semantic Web have been present within the Web from the time of Tim Berners-Lee's original Web proposal. For the Web to reach its full potential, it must grow and incorporate this Semantic Web vision, providing a universally accessible platform that allows data to be shared and processed by automated tools as well as by people. The W3C Semantic Web Activity is a multi-faceted program of basic research, collaborative technology development and open consensus-based standards setting to bring the Semantic Web to a reality and open the door to a whole new set of effective information integration, management and automation services.

For Further Reading

Resource Description Framework (RDF)
W3C Working Draft (work in progress) 11 November 2002, RDF Primer; www.w3.org/TR/rdf-primer/

Semantic Web Home Page
W3C, Semantic Web; www.w3.org/2001/sw/

URIs
W3C, Naming and Addressing: URIs, URLs, ...; www.w3.org/Addressing/

Web Ontology Language (OWL)
W3C Working Draft (work in progress) 12 November 2002, Web Ontology Language (OWL) Reference Version 1.0; www.w3.org/TR/owl-ref/

Semantic Web Services

by Bijan Parsia

Bijan Parsia, a Semantic Web researcher in the MIND Laboratory at the University of Maryland, College Park, can be reached by e-mail at bparsia@isr.umd.edu

The World Wide Web allows people to follow bewildering paths through hundreds of documents on a variety of topics without losing equilibrium. People shop, plan and book travel, check accounts and pay bills, publish articles and artwork, share photos with family and friends and total strangers, build complex information systems and manage a variety of business interactions. Web programmers have done very well in building programs that help people write Web pages, build and maintain Web sites and develop sophisticated Web stores. It is much more challenging to develop programs that can *use* the Web with more human-like flexibility without a human being constantly in the loop. The Semantic Web and Web Services are two visions of how to make the Web more amenable to automated use.

The Semantic Web

"The Semantic Web is the web of connections between different forms of data that allow a machine to do something it wasn't able to do directly." (*Weaving the Web*, p. 185)

Programs can do a lot with the current Web, much of it critical to successful and pleasant human interaction with it. Web crawlers and spiders make link checking and site archiving easy and are absolutely essential for search engines. What makes Google *the* search engine for the Web is its scope (achieved by crawling the Web), the "goodness" of its results (achieved, in part, by automated reasoning about the links between pages) and the fact that it doesn't require massive, disciplined human intervention by the search engine operators or by Web authors. The last point needs a bit of explication. Google does require a staff of brilliant people who constantly enhance and tune the crawling, indexing, ranking, storage and retrieval programs. This is a significant amount of human effort, but it is dwarfed by the alternative: a team of catalogers to categorize three billion Web pages by hand. Google also doesn't require that page authors supply correct metadata for their pages *above* what they do naturally in writing link-rich hypertext pages. There's no need for an explicit cataloging step, either by page authors or by Google operators. Moreover, it's not clear that adding that step would produce better results (or even results as good).

A striking feature of this sort of automation is that it depends on the interconnectedness of the Web – its *link structure*. Web links are what make the Web work for human browsers, authors and, it turns out, for some content-sensitive programs. This suggests that a richer link structure could support not just better search, but other activities. Hyperlinking Web pages together for the convenience of human browsers turns out to be useful for building search engines that cater to the needs of those human browsers. What can we build if we have different kinds of links supporting activities besides human browsing? The best way to find out what we can do is to do it. The Semantic Web is driven by a specific vision: to explore what sorts of links between which kinds of representations supply the greatest achievable potential for the Web.

The Semantic Web is rooted in Tim Berners-Lee's original conception of the Web. Web links were seen not just as providing a navigatory connection for the reader, but also as (partially) constituting the meaning of the linked representations. On this view, the Web is a kind of *semantic net* with Web pages as nodes and hyperlinks as arcs. The machine processable meanings of the nodes are constituted by the patterns of arcs between them. The original Web was a hypertext system instead of a pure semantic net. The nodes were bits of text with quite a bit of meaning all on their own – meaning that is largely inaccessible to current programs, though quite handy for literate people. So, the extra meaning of the Web based on knowledge representation (KR) tends to take a back seat to that based on natural language. Web links only give us

a little semantics, but it turns out that a little semantics goes a long way.

Still, it would be nice to have a little more semantics. If the original Web is a hypermedia system with aspirations toward KR, the Semantic Web seeks to be a KR system deeply integrated with global, distributed hypermedia. More precisely, the Semantic Web is a Web-like – global, distributed, universal, loosely coupled – KR extension of the human/hypermedia Web. There are no content-directed constraints on putting up a Web page or with making a Web link to anything else on the Web. These design desiderata are commonly expressed with the slogan, "Anyone can say anything about anything." It's not enough, therefore, to have a KR system with the same scope of subject matter as the Web (that is, any topic you care to write about), but that system must also accept input from practically anyone. This vision is as much a departure from traditional KR as the Web is from traditional hypertext. The Semantic Web requires a blending of KR and Web architecture. The difference between the current Web and the Semantic Web is that the Semantic Web asks people to say their anythings about anything in a way that's more amenable to significant, content-sensitive machine processing.

If the Semantic Web requires more effort from the vast Web-authoring masses only to make it easier for some programmers to do useful and intelligent things with Web content, it's not going to fly. Arguably, there's a fairly firm limit to how much effort can sanely and fruitfully be required of Web authors. If you have to be a knowledge engineer to slap up a Semantic Web page, then most people, even smart, interested people, aren't going to do it. On the other hand, a lot of websites already require a bit of reasonably heavy modeling, typically in the form of relational databases. Homebrew content management systems are everywhere, and many go beyond modeling the generic aspects of a website (authors, articles, pages, ads, bookmarks, etc.) and add support for various sorts of classification and content modeling.

The Semantic Web offers common languages and techniques to share and use such models in a naturally Web-like way. Just as HTML expanded from a drop-dead-simple, static hypertext format to a very rich, complex language capable of expressing very dynamic, media rich pages and websites, the Resource Description Framework (RDF) is being enriched and extended into more generally powerful modeling languages such as the forthcoming Web Ontology Language (OWL). OWL is an expressive, declarative language for describing sophisticated terminologies, classification schemes, taxonomies and ontologies in a way that allows machine-based reasoners to derive significant, and often surprising, results. OWL reasoners can check the consistency of a model or classify resources with greater specificity as you add information or derive new classifications from your data (among other things). OWL also has constructs to express mappings between ontologies, and a hot area of research – and tentative deployment – is automating these mappings.

Web Services and The Problem of Discovery

As described, the Semantic Web emphasizes making Web-based information more friendly for programs (especially reasoning programs). Web Services, by contrast, are rooted in making it more straightforward for programs to do things on the Web. The early conception of Web Services was primarily of remote procedure calls over the Web with both the invocations (the request) and the returned value (the response) encoded in XML (Extensible Markup Language). This makes it easier to write programs that use other programs across the Web by making the remote programs (i.e., the services) feel very similar to code libraries locally provided by the operating system and programming environment.

There are two strong motivations for using Web technologies for services. The first is to reuse existing deployment of and expertise with those technologies. Web servers are everywhere, Web client libraries are even more prevalent and one can scarcely turn around without tripping over an XML parser. This motivation holds even if the services in question are to be used solely on private intranets or, for that matter, for inter-process communication on a single machine.

The second motivation is to deploy the service on the Web. A Web Service is much like a Web page or site: You've put it up – now you want people to use it. Often, you want a lot of people to use it. If your service is primarily intended for human beings who have already found your associated website, then the problem of how to get people using your Web Service reduces to getting them to your website and, once they are there, giving them enough information so that they can figure out how to use the associated service. Getting people to come to your website has known solutions, and a programmer can simply read your service's documentation to learn how to invoke it. Notice that both steps rely heavily on human intervention: A human programmer has to find the service and then figure out how and when to use it. If there are only 10 Web Services, finding the right one isn't hard. If there are a few hundred, it's tedious, but feasible. Even these simple cases, however, can defeat automated discovery.

Furthermore, while useful Web Services may be scarce,

The big Web Services hopefuls recognize the importance of the problem of discovery and have moved to address it with Universal Description, Discovery and Integration of Web Services (UDDI).

they aren't that scarce. The hope of a lot of players, big and small, is that we'll have problems of Web Service superabundance rather than scarcity. On the Web perhaps most searches aren't critical, or, at least, aren't time critical. After all, if you're simply looking for a recipe for honey-almond nougat, spending a bit of time using search engines or browsing various recipe sites is quite reasonable. In fact, like wandering the stacks in a library, this sort of surfing has pleasures and rewards. Casual consumer shopping also works well with a leisurely air. However, these examples have a critical feature: low cost of failure. If you pay a dollar more for the book or never find an appealing recipe, the loss of money or opportunity is not very significant. However, if you are trying to secure fairly large supplies of a part (and only if you can get enough of another part), or make reservations for an emergency trip as quickly and cheaply as possible, failure can be a serious and costly problem.

Solving the Problem with Semantics

The big Web Services hopefuls recognize the importance of the problem of discovery and have moved to address it with Universal Description, Discovery and Integration of Web Services (UDDI). The *UDDI Executive White Paper* argues fervently that the current crop of search engines isn't up to the task – in fact, that a service-oriented Web requires, according to the UDDI 3.0 specification, "a 'meta service' for locating Web services by enabling robust queries against rich metadata."

"Enabling robust queries against rich metadata" sounds like a pitch for the Semantic Web. But the UDDI has two additional components: UDDI is a meta-service, and it specifically focuses on locating Web Services. These constrain the task of UDDI considerably, especially in contrast with the Semantic Web's goal of helping machines deal with anyone saying anything about anything. Indeed, making UDDI a Web Service seems to have introduced a very peculiar, centralized and anti-Web attitude into the effort. For example, the *UDDI Executive White Paper* claims "[u]ntil now, there has been no central way to easily get information about what standards different services support and no single point of access to all markets of opportunity, allowing them to easily connect with all potential service consumers." (p. 1)

To Web people, the lack of a central way and a single point of access are good things, indeed critical features of the Web. The White Paper goes on to claim

> Publishing your URL's [sic] to the UDDI registry is much more of an exact science that [sic] it is trying to get your Web site's URL into an Internet search engine. With UDDI, you have complete control over which of your business and service and service address information is published and when. This is because you are the one who publishes it. (p. 4)

Of course, "being the one who publishes" is what you are for your own website, which is a perfectly sane place to put your business and service and service address information. Being the publisher is also exactly what you are not when you put that information into "public" UDDI repositories, at least in any normal sense of being the publisher. The companies running the repositories, the "UDDI Operators," are the publishers. You give your data to them, and they publish it. Is it surprising that the examples of UDDI Operators are huge corporations like Hewlett-Packard, IBM and Microsoft?

The restriction of the subject matter of searches to Web Services seems to have encouraged the development of less sophisticated metadata techniques – perhaps the constrained domain suggests that you can get away with (much) less. The metadata representation "system" at the core of UDDI looks especially impoverished next to OWL. Version 3.0 of the UDDI specification has begun to take taxonomies seriously, but its way of associating properties with services, combining

them, deriving sets of them from each other is horribly kludgy. It is a warning sign when a large number of your search techniques involve prefix, suffix, substring, approximate or "fuzzy" matching on strings – you are going to need a person driving that search. The lack of a logic for UDDI "tModels" (the key metadata construct) makes it practically impossible to write a system that does anything more than heuristic reasoning with them, and even that is fairly difficult and often highly application or domain specific.

Furthermore, the ontologies are coming. OWL's predecessor language, DAML+OIL, already enjoys wide interest and use (see the ontology library at www.daml.org for a sampling). As the Web Ontology Working Group heads toward completion, it's clear that OWL the language and associated documents defining it are very solid. The two years experience with DAML+OIL has produced a mature user base that is not only comfortable with its understanding of the language and how to use it, but of its ability to motivate and explain it to the wider world. It would be silly if the UDDI community didn't exploit this expertise.

But if they don't, others will. The DAML Web Services (DAML-S) joint committee is an initiative to develop Semantic Web Services to deal with the challenges of Web Services in the context of the Semantic Web. DAML-S goes far beyond discovery to provide ontologies for the composition, execution and monitoring of Web Services. One key goal is to support rich enough descriptions of how a service works that we can derive facts about what it does, how much it would cost and other features of the service. Thus, the service provider (or broker) can concentrate on describing the service from a fairly narrow, functional perspective and let AI planners figure out if their service can be fruitfully combined with others at a reasonable price and in a timely manner to produce the desired result. In other words, the DAML-S committee is developing a language for describing Web-based services to support programs that can write other programs with little or no human intervention.

Staying on the Web

When caught up with enthusiasm for a technique or technology that seems to promise a significant new bit of coolness for the Web, it's easy to confuse coolness inherent in the new toy with coolness derived from enhancing (or being enhanced by) the Web. Expert systems are neat, but wrapping one in an HTML user interface doesn't really change it in any interesting way. Distributed computing is sexy, but using Web servers just because they're there doesn't just miss one opportunity, but many.

The Semantic Web and Web Services turn out not to be quite the rivals that they sometimes seem to be. DAML-S makes good use of many Web Service technologies – SOAP (Simple Object Access Protocol) and WSDL (Web Service Definition Language), for example – and UDDI already reflects the need for rich metadata descriptions. So it seems that more Web Semantics are in the cards, assuming that either camp is basically right about the future of the Web. One thing that the history of the Web should teach us is that the unexpected is the norm.

Acknowledgements

This research comes out of work of members of MIND's Semantic Web Agents Project, especially the work of the director, James Hendler, in collaboration with Fujitsu Laboratories of America, College Park, and also from the DAML-S coalition.

References and Further Reading

Theory of the Web

Weaving The Web: The original design and the ultimate destiny of the World Wide Web by its inventor, Tim Berners-Lee (with Mark Fischetti), HarperBusiness, 2000.

"Information management: A Proposal", Tim Berners-Lee, www.funet.fi/index/FUNET/history/internet/w3c/proposal.html, (also in *Weaving the Web*).

Architectural Styles and the Design of Network-based Software Architectures, Roy T. Fielding, www.ics.uci.edu/~fielding/pubs/dissertation/top.htm.

Architecture of the World Wide Web, W3C Working Draft (work in progress) of the Technical Architecture Group, www.w3.org/TR/Webarch/.

Semantic Networks, John F. Sowa, www.jfsowa.com/pubs/semnet.htm

OWL/DAML-S

Web Ontology Working Group website, www.w3.org/2001/sw/WebOnt/

DAML website, www.daml.org

DAML Web Services (DAML-S), www.daml.org/services/

TheTrue Meaning of Service, Kendall Grant Clark, www.xml.com/pub/a/2002/07/17/daml-s.html?page=1

UDDI

UDDI 3.0 specification, http://uddi.org/pubs/uddi_v3.htm

UDDI Technical White Paper, http://uddi.org/pubs/UDDI_Executive_White_Paper.pdf

Metadata: A Fundamental Component of the Semantic Web

by Jane Greenberg, Stuart Sutton and D. Grant Campbell

Jane Greenberg, assistant professor, School of Information and Library Science, University of North Carolina at Chapel Hill, can be reached by e-mail at janeg@ils.unc.edu. Stuart Sutton, associate professor, The Information School, University of Washington, can be reached at sasutton@u.washington.edu. D. Grant Campbell, assistant professor, Faculty of Information and Media Studies, University of Western Ontario, is at gcampbel@uwo.ca.

In their widely discussed May 2001 article on the Semantic Web in *Scientific American* (www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21) Tim Berners-Lee, James Hendler and Ora Lassila present a scenario in which a person named Pete is listening to the Beatles through his home entertainment center. Lucy, Pete's sister, phones from the doctor's office to explain that their mother needs a series of bi-weekly physical therapy sessions. The first few paragraphs of this article tell how both Pete's and Lucy's Semantic Web agents (hereafter referred to as *agents*) communicate with each other and transverse the Semantic Web to schedule their mother's physical therapy session, how Pete is not pleased with the initial plan and how later that evening Pete sends his agent back onto the Semantic Web to find an alternative plan. Pete's Web agent completes this second task and reschedules several of his personal and less important work appointments.

Realizing this scenario is dependent not only on the ability of Pete's and Lucy's agents to communicate with each other, but their ability to transverse a system of structured semantic knowledge that is forming the Semantic Web. This system of semantics is *metadata*. With efforts to build the Semantic Web, we are beginning to see the metadata infrastructure that agents need to carry out tasks that the initial Web has not been able to support. To this end, implementing and harvesting metadata is fundamental to the success of the Semantic Web. This article provides an overview of metadata as a key component of the Semantic Web – its vision and architecture; metadata vocabularies; enabling technologies; and authoring and annotation.

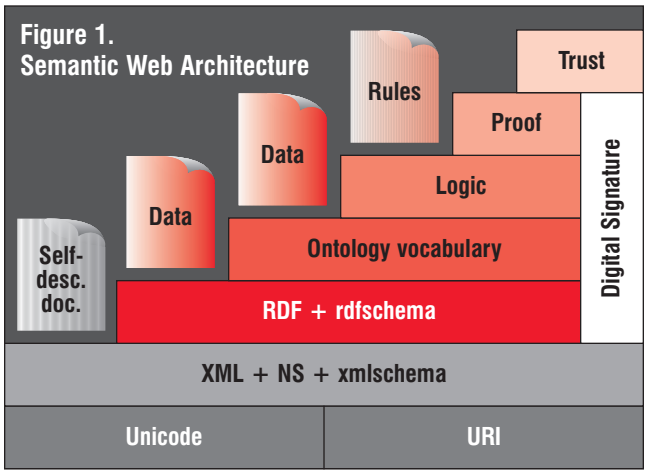
Semantic Web: Vision & Architecture

Clearly articulated visions and architectural plans, drawn by great thinkers and experts of the time, form the underpinnings of many of the world's most significant structures. Consider the Panama Canal. Crossing the mass of land called the Americas at its narrowest point by means of a waterway was a vision shared by many throughout history, including indigenous people of the Americas, Christopher Columbus and merchants worldwide. Execution of architectural plans, first by leading French engineers in 1879 and then by a U.S. Commission, led to completion of the Panama Canal in 1914. In like fashion, an evolving and shared vision, supported by an architectural plan, underlies the development of the Semantic Web.

The Vision. The Semantic Web was envisioned by Tim Berners-Lee, inventor of the World Wide Web (Web), and is now being further defined by researchers and visionaries, but it was inspired by a host of creative thinkers who have, throughout history, looked to technological innovation as a way not only to control, but also to transform the world's mass of information into intelligence. Vannevar Bush was one early pioneer in this area with his vision of the Memex (www.theatlantic.com/unbound/flashbks/computer/bushf.htm) – a mechanism using associative indexing to link the world's vast body of scientific recorded knowledge to discover new knowledge. Another significant idea is Alan Turing's conceptualization of the Turing Machine and its use of logic to transform numbers into intelligence. (See www.turing.org.uk/turing/scrapbook/machine.html.) The vision supporting the Semantic Web draws upon these ideas and new ideas inspired by technological developments to create intelligence.

The Architecture. The Semantic Web's architecture, captured by Berners-Lee, is represented in Figure 1. Each layer supports or has a connection to metadata.

- **URIs and Unicode.** URIs (uniform resource identifiers) are unique identifiers for *resources* of all types—from schemas to people. A major



(Berners-Lee, T., 2000) www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html

component of the base layer, URIs are metadata and function like ISBNs (international standard book numbers) or Social Security numbers in the context of the Web.

- **XML + NS + XMLschema.** Extensible Markup Language (XML) and more recently XML schemas facilitate the creation, use, and syntactic interoperability of metadata vocabularies. NS (namespaces), which are identified via URIs, secure semantic interoperability among metadata vocabularies.
- **RDF and RDFschema.** The RDF family further supports interoperability at the semantic level. RDF developments comprise the base Web language, so that agents, like Pete's and Lucy's discussed above, can make logical inferences, based on metadata, to perform tasks.
- **Ontology vocabulary.** Ontologies are metadata systems (referred to as metadata vocabularies in this article). The ontology layer represents the Semantic Web's central metadata artery, where simple descriptive to complex classificatory schemas are to be created and registered so that agents can intelligently interpret data, make inferences, and perform tasks. Jacob's article in this issue discusses ontologies in detail.
- **Logic.** We make logical inferences in our performance of daily tasks. For example: If N denotes new unread email in an email inbox, then if an *N* appears by a particular message, the message is new unread email. This inference is based on evidence provided by the letter *N*. The logic layer of the Semantic Web works on this basic principle through First Order Predicate Logic. An agent can derive a logical conclusion (or reason) in the process of completing a task based on what are essentially "facts" rendered from semantically encoded metadata. Other types of logic may also be applicable in the Semantic Web.
- **Proof and Trust.** The last two horizontal layers build off of the logic layer, *proof* being a validation of the "evidence" stemming from the inferential logic activity and *trust* relating to the integrity of the proof, which may be traced back down through the other layers in Berners-Lee's diagram. The functionality of these two layers is highly dependent on creation of accurate and trustworthy metadata.
- **Digital signature.** Digital signatures run horizontal to the RDF family up through the proof layer and support the notion of trust. Developments in the area of digital signatures are progressing, and could eventually help validate the integrity of metadata that an agent will use for reasoning and task completion.

The synopsis provided here shows that metadata permeates each layer of the Semantic Web's architecture, although it is not the only piece, as agents, enabling technologies, authoring and annotation programs, and metadata vocabularies need to be further developed to realize the full potential of the Semantic Web.

Metadata Vocabularies

Metadata vocabularies are synonymous with onotologies, as discussed above. A vocabulary, in a general sense, is a shared system of semantics for concepts or objects. Think about the vocabulary that comprises Portuguese. It is a system of agreed upon meanings permitting intelligible communication among Portuguese people and other persons who speak this language.

The metadata world hosts a range of systems to which the label of "metadata vocabulary" is applied. These vocabularies range from basic descriptive metadata systems, with limited or single-focused functionalities, to more complex "member/class" semantic vocabularies. An example of a simple ontology is the Dublin Core Metadata Initiative (DCMI) Elements and Element Refinements (www.dublincore.org/usage/terms/dc/current-elements/), a metadata schema developed mainly to facilitate resource discovery. A more complex metadata vocabulary to which the word *ontology* is applied is the Ariadne Genomics ontology (www.ariadnegenomics.com/technology/ontology.html). This ontology is used to formalize data on cell proteins for computer analysis: the ontology defines the various proteins, classifies them into taxonomic trees and defines the semantic relationships among them. While many metadata vocabularies in operation were not necessarily created with the Semantic Web in mind, they may be able to play a significant role in its development. For existing and developing ontologies to be used and function fully in the Semantic Web environment, they need to adhere to standards supported by enabling technologies.

Enabling Technologies for Metadata

Although metadata is integral to the Semantic Web, metadata on its own is far from sufficient. Needed are standards to syntactically encode and represent semantic knowledge so that agents can perform tasks in an efficient and comprehensible manner. A variety of enabling technologies have developed over the last few years that are proving significant to the

construction of the Semantic Web. Some have developed prior to the conceptualization of the Semantic Web, while others are being developed and refined specifically with the Semantic Web in mind. Among four key developments that are critical for metadata encoding and manipulation by agents are:

- XML and XML schema
- RDF and RDFschema
- DAML+OIL (DARPA Agent Metadata Language/ Ontology Inference Layer) (www.w3.org/TR/daml+oil-reference)
- OWL (Web Ontology Language) (www.w3.org/2001/sw/WebOnt).

The other articles in this special section of the *Bulletin of the American Society for Information Science and Technology* give attention to each of these and other important technologies.

Semantic Web Authoring and Annotation

As we move on with our discussion of the Semantic Web, we need to account for the notion that all of these impressive developments assume that metadata will exist – that authors and third parties will take the time and trouble to create metadata for Web content which can be read, understood and harvested by intelligent agents. This is a daunting proposition since accurate, consistent metadata is notoriously difficult to create, as librarians and information scientists over the world will verify. The challenge, then, is to decentralize a task that has traditionally been centralized in libraries and other information centers, carried out by professionally trained catalogers. The glories of the Semantic Web will ultimately depend on tools that will enable authors to create with very little effort RDF annotations and other useful semantic metadata on their Web pages.

Annotation has been one of the slower developments on the World Wide Web: Berners-Lee's vision of a Web that permits collaborative authoring, in addition to hyperlinked pages, has not yet materialized. At last, however, an encouraging number of annotation tools are appearing, ranging from simple captioning systems to ambitious and sophisticated systems that provide multiple views of annotation in multiple formats. At the most basic level, programs such as Annotea enable multiple users to comment on and provide metadata for a single pool of documents for purposes of collaborative writing and research. Other programs more directly geared to the Semantic Web provide ways in which RDF and ontological data may be easily created and stored in the headers of the document.

Painless creation of RDF metadata depends on two things: a predefined ontology that spares the author the task of creating terms and relationships and a user-friendly interface that permits the author to create metadata instances intuitively. One such annotation program, OntoMat, permits the author to download a predefined ontology that appears in one window, while the HTML document being annotated appears in the other. The author highlights elements of the document to be annotated and places them into a third window, and then uses the ontology to define the data elements and their relationships to each other. The program then generates highly detailed RDF metadata without the author having seen a single angle bracket. With tools such as these, conference organizers, for instance, can require contributing authors to annotate their abstracts with RDF metadata or members of a particular community can annotate their Web pages using a common ontology. Equally important, the tasks can be integrated fairly painlessly into the normal workflow of Web authoring.

A variety of annotation tools are listed at the Semantic Web Authoring and Annotation site: http://annotation.semanticweb.org/tools. They include

- OntoMat – http://annotation.semanticweb.org/tools/ontomat
- Annotea – www.3.org/2001/Annotea/
- Annozilla – http://annozilla.mozdev.org/
- COHSE – http://cohse.semanticweb.org/software.html
- SMORE – www.mindswap.org/~aditkal/editor.shtml

These tools, and the tools which improve on them, will hopefully provide the simplicity that is essential for the Semantic Web to grow. The Semantic Web, after all, was not envisioned as a tool for information professionals and computer scientists, but as a tool for everyone. And if Pete and Lucy are ever going to prefer this new Web to their date books and palm pilots, it has to stay simple.

Conclusion

Our perceptions of metadata's role in both the vision and architecture of the Semantic Web are not yet fully focused. While the vision embraces metadata as a first-order prerequisite to that architecture, its roles and mechanisms currently resonate with the evolution of earlier technological achievements. At the turn of the last century, as ships began to pass through the Panama Canal, the fledgling horseless carriage traversed other byways in forms we would hardly recognize today. They asked then, Shall we steer with a wheel or with a stick? Will the brake be on the left or the right of the steering column? Competing variations on the vision's theme sought to dominate the architecture of the evolving automobile. In many ways, we stand in a place quite like that occupied by those earlier pioneers when we view the potential roles and forms of metadata in the emerging architecture of the Semantic Web.

Further Reading

Berners-Lee, T. (1997). Axioms of Web architecture: Metadata: Metadata architecture. In *Design issues: Architectural and philosophical points (Semantic Web roadmap)*. Available at www.w3.org/DesignIssues/Metadata.html

World Wide Web Consortium. *Metadata activity statement.* Available at www.w3.org/Metadata/Activity.html

Semantic Web authoring and annotation. Available at http://annotation.semanticweb.org/

Special Section

Ontologies and the Semantic Web

by Elin K. Jacob

Elin K. Jacob is associate professor, School of Library and Information Science, Indiana University-Bloomington and can be reached by e-mail at ejacob@indiana.edu

For those interested in the continuing evolution of the Web – and particularly for those actively engaged in development of the Semantic Web – ontologies are "sexy." But even though ontologies are currently a very popular topic, there appears to be some confusion as to just what they are and the role that they will play on the Semantic Web. Ontologies have been variously construed as classification schemes, taxonomies, hierarchies, thesauri, controlled vocabularies, terminologies and even dictionaries. While they may display characteristics reminiscent of each of these systems, to equate ontologies with any one type of representational structure is to diminish both their function and their potential in the evolution of the Semantic Web.

Ontology (with an upper-case "O") is the branch of philosophy that studies the nature of existence and the structure of reality. However, the definition provided by John Sowa (http://users.bestweb.net/~sowa/ontology/index.htm) is more appropriate for understanding the function of ontologies on the Semantic Web. Ontology, Sowa explains, investigates "the categories of things that exist or may exist" in a particular domain and produces a catalog that details the types of things – and the relations between those types – that are relevant for that domain. This catalog of types is an *ontology* (with a lower-case "o").

The term *ontology* is frequently used to refer to the semantic understanding – the conceptual framework of knowledge – shared by individuals who participate in a given domain. A semantic ontology may exist as an informal conceptual structure with concept types and their relations named and defined, if at all, in natural language. Or it may be constructed as a formal semantic account of the domain with concept types and their relations systematically defined in a logical language and generally ordered by genus-species – or type-subtype – relationships. Within the environment of the Web, however, an ontology is not simply a conceptual framework but a concrete, syntactic structure that models the semantics of a domain – the conceptual framework – in a machine-understandable language.

The most frequently quoted definition of an ontology is from Tom Gruber. In "Ontologies as a specification mechanism" (www-ksl.stanford.edu/kst/what-is-an-ontology.html), Gruber described an ontology as "an explicit specification of a conceptualization." This definition is short and sweet but patently incomplete because it has been taken out of context. Gruber was careful to constrain his use of *conceptualization* by defining it as "an abstract, simplified view of the world that we wish to represent for some purpose" – a partial view of the world consisting only of those "objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them." Following Gruber's lead, an ontology can be defined as a partial, simplified conceptualization of the world as it is assumed to exist by a community of users – a conceptualization created for an explicit purpose and defined in a formal, machine-processable language.

Why Do We Need Ontologies?

Because the Web is currently structured to support humans, domain terms and HTML metadata tags that are patently transparent to human users are meaningless to computer systems, to applications and to agents. XML is gaining increasing acceptance and is rapidly replacing HTML as the language of the Web. But XML schemas deal primarily with the physical structure of Web documents; and XML tag names lack the explicit semantic modeling that would support computer interpretation. If the Semantic Web is to realize the goal of enabling systems and agents to "understand" the content of a Web resource and to integrate that understanding with the content of other resources, the system or agent must be able to interpret the semantics of each resource, not only to

Ontologies are not new to the Web. Any metadata schema is, in effect, an ontology specifying the set of physical and/or conceptual characteristics of resources that have been deemed relevant for a particular community of users.

accurately represent the content of those resources but also to draw inferences and even discover new knowledge. In the environment of the Semantic Web, then, an ontology is a partial conceptualization of a given knowledge domain, shared by a community of users, that has been defined in a formal, machine-processable language for the explicit purpose of sharing semantic information across automated systems.

An ontology offers a concise and systematic means for defining the semantics of Web resources. The ontology specifies relevant domain concepts, properties of those concepts – including, where appropriate, value ranges or explicit sets of values – and possible relationships among concepts and properties. Because an ontology defines relevant concepts – the types of things and their properties – and the semantic relationships that obtain between those concepts, it provides support for processing of resources based on meaningful interpretation of the content rather than the physical structure of a resource or syntactic features such as sequential ordering or the nesting of elements.

An Example of an Ontology

Ontologies are not new to the Web. Any metadata schema is, in effect, an ontology specifying the set of physical and/or conceptual characteristics of resources that have been deemed relevant for a particular community of users. Thus, for example, the set of elements and element refinements defined in the Dublin Core [DC] is itself an ontology. The most current version of the DC element set (http://dublincore.org/usage/terms/dc/current elements/) consists of 16 attributes (element types) and 30 qualifiers (element refinements or subtypes) that are defined and maintained by the Dublin Core Metadata Initiative Usage Board. DC is intended to support consistency in the description and semantic interpretation of networked resources. To this end, declaration of the vocabulary of DC (the set of elements and element refinements) in the machine-processable language of RDF/RDFS (see below) is projected to be available in early 2003.

DC is a relatively simple representational structure applicable to a wide range of non-domain-specific resources. Nonetheless, it is an ontology, albeit a very general one, because it imposes a formally defined conceptual model that facilitates the automated processing necessary to support the sharing of knowledge across systems and thus the emergence of the Semantic Web. While an ontology typically defines a vocabulary of domain concepts in an is-a hierarchy that supports inheritance of defining features, properties and constraints, DC illustrates that hierarchical structure is not a defining feature of ontologies. The 16 elements currently defined by DC are independent of each other: none of the elements is required by the conceptual model and any one may be repeated as frequently as warranted for any given resource.

The Role of RDF/RDFS

Although hierarchy is not a defining characteristic of ontologies, it is an important component in the representational model prescribed by the Resource Description Framework (RDF) Model and Syntax Specification (www.w3.org/TR/REC-rdf-syntax/) and the RDF Vocabulary Description Language schema (RDFS) (www.w3.org/TR/rdf-schema). RDF and RDFS have been developed by the W3C and together comprise a general-purpose knowledge representation tool that provides a neutral method for describing a resource or defining an ontology or metadata schema. RDF/RDFS doesn't make assumptions about content; it doesn't incorporate semantics from any particular domain; and it doesn't depend on a set of predetermined values. However, it does support reuse of elements from any ontology or metadata schema that can be identified by a Uniform Resource Identifier (URI).

RDF defines a model and a set of elements for describing resources in terms of named properties and values. More importantly, however, it provides a syntax that allows any resource description community to create a domain-specific representational schema with its associated semantics. It also supports incorporation of elements from multiple metadata schemas. This model and syntax can be used for encoding information

in a machine-understandable format, for exchanging data between applications and for processing semantic information. RDFS complements and extends RDF by defining a declarative, machine-processable language – a "metaontology" or core vocabulary of elements – that can be used to formally describe an ontology or metadata schema as a set of classes (resource types) and their properties; to specify the semantics of these classes and properties; to establish relationships between classes, between properties and between classes and properties; and to specify constraints on properties. Together, RDF and RDFS provide a syntactic model and semantic structure for defining machine-processable ontologies and metadata schemas and for supporting interoperability of representational structures across heterogeneous resource communities.

RDFS

In order to understand more clearly both the nature and the function of ontologies, it is helpful to look more closely at the schema structure of RDFS. While an XML schema places specific constraints on the structure of an XML document, an RDFS schema provides the semantic information necessary for a computer system or agent to understand the statements expressed in the language of classes, properties and values established by the schema. One of the more important mechanisms that RDFS relies on to support semantic inference and build a web of knowledge is the relationship structure that typifies the hierarchy and is so characteristic of traditional classification schemes. The creation of generic relationships through the nesting structure of genus-species (or type-subtype) capitalizes on the power of hierarchical inheritance whereby a subclass or subproperty inherits the definition, properties and constraints of its parent.

An RDFS ontology differs from taxonomies and traditional classification structures, however, in that the top two levels of the hierarchy – the superordinate class *resource* and its subordinate classes *class* and *property* – are not determined by the knowledge domain of the ontology but are prescribed by the RDFS schema. Every element in the ontology is either a type of *class* or a type of *property*. Furthermore, the relationships between classes or properties are potentially polyhierarchical: thus, for example, a particular class may be a subclass of one, two, three or more superordinate classes.

A taxonomy or traditional classification scheme systematically organizes the knowledge of a domain by identifying the essential or defining characteristics of domain entities and creating a hierarchical ordering of mutually exclusive classes

The *BULLETIN OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY* is a BIMONTHLY PUBLICATION that serves as the newsletter of the Society. It publishes short articles on a BROAD RANGE OF TOPICS of current concern to ASIST MEMBERS, focusing particularly on material of interest to practitioners. Readers are ENCOURAGED TO SUGGEST topics of interest or alert the Editor of suitable material that may have been presented at ASIST-sponsored events or elsewhere. In addition, authors are ENCOURAGED TO SUBMIT articles on topis such as CURRENT PRACTICE, PUBLIC POLICY, LEGISLATION, STANDARDS, PILOT PROJECTS, STATE-OF-THE ART REVIEWS or OVERVIEWS OF EVOLVING TECHNOLOGY AND ITS IMPACT. Articles informing the membership about various developments within ASIST are very welcome, as are articles reporting on ACTIVITIES OUTSIDE THE UNITED STATES. The *Bulletin* encourages original articles, but will consider TIMELY MATERIAL that has been presented or published elsewhere. Articles are posted in full on the ASIS Web Site at http://www.asis.org/Bulletin/index.html

Authors interested in developing material for a focused issue are urged to contact the Editor directly.

Authors are encouraged to discuss article ideas with the Editor if there are questions about suitability or relevance.

Irene L. Travis, Editor
Bulletin of the American Society for Information Science and Technology
1320 Fenwick Lane,
Silver Spring, MD 20910
(301) 495-0900
Bulletin@asis.org

BULLETIN
of the American Society for Information Science and Technology

to which the entities themselves are then assigned. In contrast, an RDFS ontology does not create classes into which domain resources are slotted. Rather, the ontology defines a set of elements (or slots) to which values may be assigned, as appropriate, in order to represent the physical and conceptual features of a resource. And, unlike a classification scheme, the ontology may also incorporate a set of inference rules that allows the system or agent to make inferences about the represented knowledge, to identify connections across resources or to discover new knowledge.

In an RDFS ontology, relationships between classes and properties are created by specifying the *domain* of a property, thereby constraining the class or set of classes to which a given property may be applied. In this respect, the structure of an RDFS schema is reminiscent of a faceted representational language or thesaurus. However, unlike a thesaurus, which authorizes a controlled vocabulary of terms (values) that can be assigned to represent the content of a resource, the structure of an RDFS ontology consists of a system of elements or slots whose possible range of values may or may not be established by the ontology. RDFS does provide for establishment of a controlled vocabulary (or vocabularies) within the structure of the ontology: specifying the *range* of a property stipulates that any value of that property must be an instance of a particular class of resources (e.g., the class *Literal*). An RDFS ontology is further distinguished from a traditional thesaurus in that it does not incorporate a lead-in vocabulary. And, while it is possible to map natural language synonyms to the appropriate classes or properties in the ontology, this must be accomplished through a domain lexicon that is external to the ontology itself.

The argument that an ontology constitutes a controlled vocabulary is only valid if the standard concept of a controlled vocabulary is redefined. A controlled vocabulary is generally understood to consist of a set of terms (values) that have been authorized to represent the content of a resource. In contrast, an ontology consists of a catalog of types and properties – a catalog of controlled and well-defined element slots – that are meaningless when applied to a resource unless they are paired with an appropriate value. And, although an ontology defines a catalog of types, it is not a dictionary. A dictionary is a list of terms and associated definitions arranged in a meaningful order; but, because that order is generally alphabetical, it does not establish the meaningful relationships among terms (elements) that are characteristic of an ontology.

An ontology is not a taxonomy, a classification scheme or a dictionary. It is, in fact, a unique representational system that integrates within a single structure the characteristics of more traditional approaches such as nested hierarchies, faceted thesauri and controlled vocabularies. An ontology provides the semantic basis for metadata schemes and facilitates communication among systems and agents by enforcing a standardized conceptual model for a community of users. In so doing, ontologies provide the meaningful conceptual foundation without which the goal of the Semantic Web would be impossible.

Future Directions

Much still must be done to extend the capabilities and effectiveness of current ontological models. While there is ongoing work to refine the RDF/RDFS model and schema, other efforts such as the DAML+OIL Web Ontology Language (www.w3.org/TR/daml+oil-reference) and the Web Ontology Language [OWL] (www.w3.org/TR/2002/WD-owl-ref-20021112/) seek to build on the foundation established by RDF and RDFS.

Conclusion

It is simply not true that there is nothing new under the sun. This is aptly underscored not only by the history of the Web itself but also by ongoing efforts to realize the potential of the Semantic Web. Limiting responses to these new challenges by adhering to traditional representational structures will ultimately undermine efforts to address the unique needs of these new environments. As recent developments with ontologies illustrate, the knowledge accrued across generations of practical experience must not be discarded; but there must be the conscious effort to step outside the box – to rethink traditional approaches to representation in light of the changing requirements occasioned by the constantly evolving environment of the Web.

Recommended Reading

Guarino, N. (1998). Formal ontology and information systems. In N. Guarino (Ed.), *Formal ontology in information systems: Proceedings of FOIS '98* (pp. 3-15). Amsterdam: IOS Press. Available at www.ladseb.pd.cnr.it/infor/ontology/PUBL15.html

Guarino, N., & Giaretta, P. (1995). Ontologies and knowledge bases: Towards a terminological clarification. In N. Mars (Ed.), *Towards very large knowledge bases: Knowledge building and knowledge sharing* (pp. 25-32). Amsterdam: IOS Press. Available at www.ladseb.pd.cnr.it/infor/Ontology/Papers/KBKS95.pdf

Holsapple, C.W., & Joshi, K.D. (2002). A collaborative approach to ontology design. *Communications of the ACM*, *45*(2), 42-47.

Kim, H. (2002). Predicting how ontologies for the Semantic Web will evolve. *Communications of the ACM*, *45*(2), 48-54.

Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: a guide to creating your first ontology*. Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880. Stanford Knowledge Systems Laboratory. Available at www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html

Uschold, M., & Grüninger, M. (1996). Ontologies: principles, methods and applications. *Knowledge Engineering Review*, 11(2), 93-155. Available at http://citeseer.nj.nec.com/uschold96ontologie.html

Feature

Complex Acts of Knowing: Paradox and Descriptive Self-Awareness

by Dave Snowden

Dave Snowden is director of IBM's newly created Centre for Action Research in Organisational Complexity (CAROC) and was formerly a Director of IBM's Institute for Knowledge. He is a fellow of the Information Systems Research Unit at Warwick University. He can be contacted via e-mail at snowded@uk.ibm.com

Editor's Note: This article has been extracted and condensed from one that first appeared in the *Journal of Knowledge Management*, v. 6, no2 (May 2002), p. 100-111. A copy of the original article also appears on the Cynefin website (http://www-1.ibm.com/services/cynefin/). The *Bulletin* wishes to thank Emerald (publisher of the *Journal*) and the author for permission to publish this version.

The contention of this paper is that we are entering a third age in the management of knowledge. Further, that the conceptual changes required for both academics and management are substantial, effectively bounding or restricting over a hundred years of management science in a way similar to the bounding of Newtonian science by the discoveries and conceptual insights of quantum mechanics. These changes are not incremental, but require a phase shift in thinking that appears problematic, but once made reveals a new simplicity without the simplistic and formulaic solutions of too much practice in this domain.

The First Age: Information for Decision Support

The first age, prior to 1995, sees knowledge being managed, but the focus is on the appropriate structuring and flow of information to decision makers and the computerization of major business applications leading to a revolution dominated by the perceived efficiencies of process reengineering. For many, reengineering was carried out with missionary enthusiasm as managers and consultants rode roughshod across pre-existing "primitive" cultures with positive intent that too frequently degenerated into rape and pillage. By the mid- to late-90s disillusionment was creeping in. Organizations were starting to recognize that they might have achieved efficiencies at the cost of effectiveness and laid off people with experience or natural talents vital to the operation of which they had been unaware. The failure to recognize the value of knowledge gained through experience, through traditional forms of knowledge transfer such as apprentice schemes and the collective nature of much knowledge, was such that even the word *knowledge* became problematic.

1995: The Transition to the Second Age

To all intents and purposes knowledge management started circa 1995 with the popularization of the SECI model (Nonaka & Takeuchi, 1995) with its focus on the movement of knowledge between tacit and explicit states through the four processes of socialization, externalization, combination and internalization. An irony is that Nonaka and Takeuchi were only seeking to contrast a claimed Japanese tradition of "Oneness" with a rational, analytical and Cartesian western tradition. Their work derived in the main from the study of innovation in manufacturing processes where tacit knowledge is rendered explicit *to the degree necessary to enable that process to take place*; it did not follow that all of the knowledge in the designers heads and conversations had, should or could have been made explicit. In partial contrast, early knowledge programs attempted to disembody all knowledge from its possessors to make it an organizational asset. Nonaka attempted to restate his more holistic and dialectical view of tacit and explicit knowledge (Nonaka & Konno 1998), but

by this time the simple two by two of the SECI model was too well established to be restored to its original intent.

The Paradoxical Nature of Knowledge

Some of the basic concepts underpinning knowledge management are now being challenged: "Knowledge is not a 'thing,' or a system, but an ephemeral, active process of relating. If one takes this view then no one, let alone a corporation, can own knowledge. Knowledge itself cannot be stored, nor can intellectual capital be measured, and certainly neither of them can be managed." (Stacy 2001).

Stacy summarizes many of the deficiencies of mainstream thinking and is one of a growing group of authors who base their ideas in the science of complex adaptive systems. That new understanding does not require abandonment of much of which has been valuable, but it does involve a recognition that most knowledge management has been content management. In the third generation we grow beyond managing knowledge as a *thing* to managing knowledge as a *flow* and *thing*, which requires focusing more on context and narrative than on content.

The question of the manageability of knowledge is not just an academic one. Organizations have increasingly discovered that the tacit and explicit distinction tends to focus on the *container*, rather than *the thing contained* (Snowden, 2000). Three heuristics illustrate the change in thinking required to manage knowledge:

- Knowledge can only be volunteered; it cannot be conscripted.
- We can always know more than we can tell, and we will always tell more than we can write down.
- We only know what we know when we need to know it; that is human knowledge is deeply contextual – it is triggered by circumstance.

The three heuristics partially support Stacy's view of knowledge as an active process of relating. However, it does not follow that we have to abandon second-generation practice, but we must recognize its limitations. We can encompass both Stacy and Nonaka if we embrace knowledge as both a *thing* and a *flow*. In the second age we looked for things and in consequence found things; in the third age we look for both in different ways and must therefore embrace the consequent paradox.

Context: The Dimension of Abstraction

The issue of content and context, which runs through all three heuristics, is key to understanding the nature of knowledge transfer. At the highest level of abstraction, in a context where I share knowledge with myself, there is a minor cost; I may keep notes but no one else has to read them. At the other extreme if I want to share with everyone the cost becomes infinite, as the audience not only needs to share the same language, but also the same education, experience, values, etc.

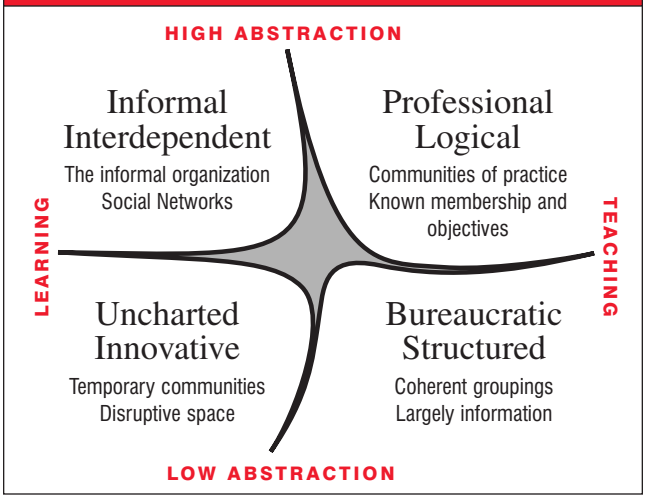
Context: The Dimension of Culture

Abstraction is one dimension of context; the other is culture. The term *culture* is used both to describe socio-cultural systems, which are artifactual and knowable, and ideational systems, which are systems of shared ideas, rules and meanings that underlie and are expressed in the way that humans live (Keesing & Strathern, 1998). Both cultures are key to the flow of knowledge within an organization. We need to transfer to new members, in both society and the organization, knowledge that has been painfully created at cost over previous generations.

Cynefin: Diversity over Time and Space

The dimensions of abstraction and culture create the sense-making model, shown below in Figure 1.

Figure 1. Cynefin: Common *Sense* Making



Cynefin (pronounced kun-ev'in) is a Welsh word with no direct equivalent in English. As a noun it is translated as *habitat*, as an adjective *acquainted* or *familiar*, but dictionary definitions fail to do it justice. It links a community into its shared history – or histories – in a way that paradoxically both limits the perception of that community while enabling an instinctive and intuitive ability to adapt to conditions of profound uncertainty. In general, if a community is not physically, temporally and spiritually rooted, then it is alienated from its environment and will focus on survival rather than creativity and collaboration. In such conditions, knowledge hoarding will predominate and the community will close itself to the external world. If the alienation becomes extreme, the community may even turn in on itself, atomizing into an incoherent babble of competing self interests. Critically it emphasizes that we never start from a zero base when we design a knowledge system, all players in that system come with the baggage, positive and negative derived from multiple histories.

Cynefin creates four open spaces or domains of knowledge, all of which have validity within different contexts. They are domains, not quadrants, as they create boundaries within a center of focus, but they do not pretend to fully encompass all possibilities.

Bureaucratic/Structured: Teaching, Low Abstraction. This is the formal organization, the realm of company policy, procedures and controls. It is a training environment. Its language is known, explicit and open. It is the legitimate domain of the corporate intranet and its shared context is the lowest common denominator of its target audience's shared context.

Professional/Logical: Teaching, High Abstraction. Commonly professional individuals, who through defined training programs, acquire a specialist terminology; codified in textbooks. The high level of abstraction is teachable given the necessary time, intelligence and opportunity. This is one of the most important domains as knowledge communication is at its most efficient due to the high level of abstraction; in second generation thinking this is the domain of communities of practice.

Informal/Interdependent: Learning, High Abstraction. In this domain we have the abstraction of shared experiences, values and beliefs. This is the domain of the shadow or informal organization, that complex network of obligations, experiences and mutual commitments without which an organization could not survive. Trust in this domain is a naturally occurring phenomenon as all collaboration is voluntary in nature. In some primitive societies the symbols are stories, often unique to a particular family who train their children to act as human repositories of complex stories that contain the wisdom of the tribe. The ability to convey high levels of complexity through story lies in the highly abstract nature of the symbol associations in the observer's mind when s/he hears the story. It triggers ideas, concepts, values and beliefs at an emotional and intellectual level simultaneously. A critical mass of such anecdotal material from a cohesive community can be used to identify and codify simple rules and values that underlie the reality of that organization's culture (Snowden, 1999). At its simplest manifestation this can be a coded reference to past experience. "You're doing a Margi" may be praise or blame – without context the phrase is meaningless, with context a dense set of experiences is communicated in a simple form.

Uncharted/Innovative: Learning, Low Abstraction. We now reach a domain in which we have neither the experience nor the expertise because the situation is new, the ultimate learning environment. The organization will tend to look at such problems through the filters of past experience. But here we can act to create context to enable action through individuals or communities who have either developed specific understanding or who are comfortable in conditions of extreme uncertainty. Such individuals or communities impose patterns on chaos to make it both comprehensible and manageable.

The Third Age: Complicated, Complex and Chaotic

The above description of the Cynefin common-sense making model relates to its use in the context of communities. It is based on an understanding of the distinctiveness of three different types of system – complicated, complex and chaotic, best understood through two distinctions.

Complex vs. Complicated. An aircraft is a *complicated* system; all of its thousands of components are knowable, definable and capable of being catalogued as are all of the relationships between and among those components, while human systems are *complex*. A complex system comprises many interacting agents, an agent being anything that has identity. We all exist in many identities in our personal and work lives. As we move among identities, we observe different rules, rituals and procedures unconsciously. In a complex system, the components and their interactions are changing and can never be quite pinned down. The system is irreducible. Cause and effect cannot be separated because they are intimately intertwined (Juarrero 1999).

Two examples make this clearer:

- When a rumor of reorganization surfaces: the complex human system starts to mutate and change in unknowable ways; new patterns form in anticipation of the event. If you walk up to an aircraft with a box of tools in your hand, nothing changes.
- Another feature of a complex system is *retrospective coherence* in which the current state of affairs always makes logical sense, but only when we look backwards. The current pattern is logical, but is only one of many patterns that could have formed, any one of which would be equally logical.

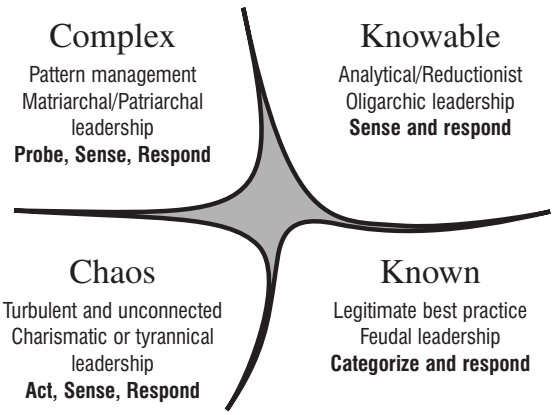
Scientific management served well in the revolutions of total quality management and business process re-engineering and continues to be applicable in the domain of the complicated; however, just as Newtonian physics was bounded by the understandings of quantum mechanics, so scientific management has been bounded by the need to manage knowledge and learning.

Complex vs. Chaotic. A complex system comprises many interacting identities in which, while I cannot distinguish cause and effect relationships, I can identify and influence patterns of interactivity. With a chaotic system all connections have broken down and we are in a state of turbulence. In a complex domain we manage to recognize, disrupt, reinforce and seed the emergence of patterns; we allow the interaction of identities to create coherence and meaning. In a chaotic domain no such patterns are possible unless we intervene to impose them; they will not emerge through the interaction of agents.

System States and the Cynefin Model

The three types of system map on to the Cynefin model, with a separation of complicated systems into those in which we know all of the cause and effect relationships and those

Figure 2. Cynefin: Decision making



that are knowable if we had the resource, capability and time (Figure 2). Each of the domains contains a different model of community behavior; each requires a different form of management and a different leadership style.

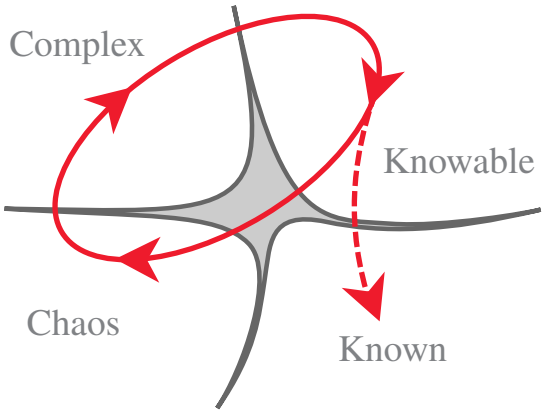
Known space is the only legitimate domain of best practice. Within known limits we can both predict and prescribe behavior. Humans, acting collectively, can make systems that might otherwise be complex or chaotic into known systems; we impose order through laws and practices that have sufficient universal acceptance to create predictable environments. On the negative side, the imposed structure can continue beyond its useful life. In this domain we categorize incoming stimuli, and once categorized we respond in accordance with predefined procedures. Leadership tends to a feudal model, with budget having replaced land as the controlling mechanism.

Knowable space is the domain of good practice. We do not yet know all the linkages, but they can be discovered. This is the domain of experts, whose expertise enables us to manage by delegation without the need for categorization. Again there is a human imposition of order but it is more fluid than in the space of the known. A major issue in the space of the knowable is entrainment of thinking. The very thing that enables expertise to develop, namely the codification of expert language, leads inevitably to entrainment of thinking. Exhortations to remain open to new ideas are unlikely to succeed. Management of this space requires the cyclical disruption of perceived wisdom. The common context of expertise is both an enabler and blocker to knowledge creation, and from time to time context must be removed to allow the emergence of new meaning. In this space we sense and respond based on our expert understanding of the situation while the leadership models are oligarchic – requiring consent of the elders of the community and interestingly oligarchies are often less innovative than the idiosyncrasies of feudalism.

The nature of the complex domain is the management of patterns. We need to identify the early signs of a pattern forming and disrupt those we find undesirable while stabilizing those we want. If we are really clever then we seed the space to encourage the formation of patterns that we can control. These patterns are emergent properties of the interactions of the various agents. By increasing information flow, variety and connectiveness either singly or in combination we can break down existing patterns and create the conditions under which new patterns will emerge, although the nature of emergence is not predictable. Entrepreneurs manage in this space instinctively while large organizations find it more uncomfortable. In this domain leadership cannot be imposed, it is emergent based on natural authority and respect but it is not democratic, it is matriarchal or patriarchal.

Chaos represents the consequence of excessive structure or massive change, both of which can cause linkages to sunder. As such it is a space that requires crisis management and is not comfortable or entered with any enthusiasm by other than the insane. However it is one of the most useful spaces, and one that needs to be actively managed. It provides a means by which entrainment of thinking can be disrupted by breaking down the assumptions on which expertise is based. It is also a space into which most management teams and all knowledge programs will be precipitated; however, regular immersion in a controlled way can immunize the organization and create patterns of behavior that will pay dividends when markets create those conditions. We also need to remember that what to one organization is chaotic, to another is complex or knowable. In the chaotic domain the most important thing is to act, then we can sense and respond. Leadership in this domain is about power – either tyranny or charisma. Both models impose order, and if order is imposed without loss of control, then the new space is capable of being used to advantage.

Figure 3. Cynefin: Knowledge Flows



The Knowledge Spiral and Cynefin

The Cynefin model allows us to see knowledge as both *thing* and *flow*, and this allows us to continue to use the insights and practices of scientific management, while embracing the new learnings and insights from the new sciences of complexity and chaos. Cynefin focuses on creating the conditions for the emergence of meaning. In its two *complicated* domains – known and knowable – these conditions are rationalist and reductionist, and the SECI model works. In the *complex* and *chaotic* domains new science and new approaches are required. The range of possible flows within the Cynefin model across its various boundary transformations is large, but here we will look at an idealized model of knowledge flow involving three key boundary transitions: the just-in-time transfer of knowledge from informal to formal, the disruption of entrained thinking and the creation and stimulation of informal communities. These transitions are shown in Figure 3.

Just-in-Time Knowledge Management: From *Complex* to *Knowable*

Like manufacturing before just-in-time (JIT) inventory management was introduced, second-generation knowledge management tries to anticipate demand. In the third generation we create ecologies in which the informal communities of the *complex* domain can self-organize and self-manage their knowledge in such a way as to permit that knowledge to transfer to the formal, *knowable* domain on a JIT basis.

The sheer number of informal and semi-formal communities within an organization is too great to permit formal management. The informal, complex space also contains much knowledge that never needs to be an organizational asset; the issue is that even if we knew what we know, we cannot distinguish in advance what we need to know as an organization, and critically when we need to know it. Techniques for the informal-formal JIT transfer include:

- Flagging by subject matter. To take an example from the author's own experience, during the early stage of pioneering work on narrative techniques for knowledge disclosure a private collaboration space was created within IBM's network, but not as a part of a formal community of practice. This contained a record of significant mistakes and associated learning that would only be shared in a small trusted community. The subject matter was flagged in the formal community under the more colloquial label of "organizational story telling." When story telling became fashionable, e-mail volume increased to a painful level. At this point a document answering the most frequently answered questions was written in self-defense. The socialization pressure of the ecology forced the voluntary codification of knowledge and provided the context that allowed the production of material at an appropriate level of abstraction.
- Expertise location systems replace the second-generation technique of yellow pages making connections between people and communities. One example, "Tacit" will trawl e-mail records to identify where expertise lies, but allow the individual knowledge holder to determine if his or her expertise is to be public, which has many advantages in building context and trust.
- We can use the complex domain as a means of creating communities in the formal space. *Clustering* is the identification of like-minded or like interested individuals within the organization, who already form the nucleus of a community. Such clusters will have already worked out the upper and lower levels of acceptable abstraction and will have sufficient shared context to create a sustainable, low cost formal community. *Swarming* is used where no naturally occurring cluster can be found, either to create a cluster or make one visible. Swarming involves creating the equivalent of a bright light and seeing what comes to it – a Web discussion group, evening lecture series, an open competition. Only if we cannot either find a cluster or a swarm do we build a formal community with all the associated costs of creating something from scratch.

Organizations need to realize the degree of their dependence on informal networks. The danger is of chronic self-deception in the formal organization, partly reinforced by the camouflage behavior of individuals in conforming to the pseudo-rational models. A mature organization will recognize that such informal networks are a major competitive advantage and while ensuring scalability through automated process and formal constructions will leave room for the informal communities to operate.

Disruption: From *Knowable* to *Chaotic*

The second key transition is to provide cyclical disruption of the entrained thinking in expert communities. Perspective shift, when necessary, is not easy to achieve and needs to be handled with care if operational efficiency is to be maintained. However there are various techniques that do work, such as taking deep experts in one field and linking them with experts in a radically different field, which will challenge their assumptions. Often it is sufficient to take only the leadership of a community into a chaotic environment. The ritual is important – humans manage boundary transitions through rituals that create awareness of the transition, but equally awareness of the new roles, responsibility and social mores associated with the new space. If the disruption is cyclical and expected, then we are closer to a learning ecology, and we have also to some degree immunized the group in respect of involuntary moves into the chaotic space.

Creating New Identities and Interactions: From *Chaotic* to *Complex*

We use the domain of chaos to disrupt in advance of need, in order to break down inappropriate or overly restrictive mod-

els, combined with constrained starvation, pressure and access to new concepts and ideas. As a result we create radically new capability within the ecology, which will both transform the *knowable* domain of experts and stimulate the creation of new networks, communities and trust/experience relationships, while new alliances and relationships form from the creative stimulus of chaos.

The chaotic space is not of itself the only source of natural communities; new people join the organization, existing projects create new informal communities and trusted links; the normal day to day interaction of human agents is a constant source of new communities. Chaos is particularly productive, but is not the only source.

The Natural Flow of Knowledge

We can now see the sensible pattern of flow of knowledge within an organization. Communities form naturally in the complex domain and as a result of activity both voluntary and involuntary within the domain of chaos. JIT techniques allow us to use the complex domain to create through a process of formalization, more natural and sustainable communities in the *knowable* domain. We can also commence operations here, but the cost will be high. A limited amount of codified knowledge can be fully separated from its owners and transferred to the best practice domain, that of the *known*. On a cyclical basis we disrupt the assumptions and models of the *knowable* domain of experts allowing new meaning to emerge. From this perspective we see knowledge as flowing between different states, with different rules, expectations and methods of management. We do not have to choose between views and approaches, but we bound those approaches to their appropriate domains. The Cynefin model allows the creation of multiple contexts.

Conclusion

We are reaching the end of the second generation of knowledge management, with its focus on tacit-explicit knowledge conversion. Triggered by the SECI model of Nonaka, it replaced a first generation focus on timely information provision for decision support and in support of business process re-engineering. Like re-engineering it has substantially failed to deliver on its promised benefits.

The third generation requires the clear separation of context, narrative and content management and challenges the orthodoxy of scientific management. Complex adaptive systems theory has been used to create a sense-making model that utilizes self-organizing capabilities of the informal communities and identifies a natural flow model of knowledge creation, disruption and utilization. Knowledge is seen paradoxically, as both a thing and a flow requiring diverse management approaches.

In the new, "complexity informed" but not "complexity constrained" third generation, content, narrative and context management provide a radical synthesis of the concepts and practices of both first and second generation. By enabling descriptive self-awareness within an organization, rather than imposing a pseudo-analytic model of best practice, it provides a new simplicity, without being simplistic, enabling the emergence of new meaning through the interaction of the formal and the formal in a complex ecology of knowledge.

Additional Acknowledgements

Some parts of this paper were originally published in the conference proceedings of KMAC at the University of Aston, July 2000. The idea of "knowledge" becoming a problematic concept comes from J C. Spender.

The views expressed in this paper are those of the author and are not intended to represent the views of either IBM or IBM's Institute for Knowledge Management.

The Cynefin Centre

Membership of the Cynefin Centre, which focuses on action research in organizational complexity, is open to individuals and to organizations. It focuses on high-participation action research projects seeking new insights into the nature of organizations and markets using models derived from sciences that recognize the inherent uncertainties of systems comprising interacting agents. The basis of all center programs is to look at any issue from multiple new perspectives and to facilitate problem solving through multiple interactions among program participants. Programs run on a national, international and regional basis and range from investigation of seemingly impossible or intractable problems to pragmatic early entry into new methods and tools such as narrative databases, social network stimulation and asymmetric threat response.

References

Juarrero, A (1999). *Dynamics in action: Intentional behaviour as a complex system.* Cambridge, MA: MIT Press.

Keesing, R. & Strathern, A. (1998). *Cultural anthropology: A contemporary perspective*. Orlando, FL: Harcourt Brace.

Nonaka, I. & Konno, N. (1998). The concept of "Ba": Building a foundation for knowledge creation. *California Management Review*, *40*(3), 40-54.

Nonaka, I. & Takeuchi, H. (1995). *The Knowledge-creating company.* London: Oxford University Press.

Snowden, D. (1999). The paradox of story. *Scenario and Strategy Planning, 1* (5).

Snowden, D. (2000). Organic knowledge management: Part I The ASHEN model: An enabler of action. *Knowledge Management, 3* (7), 14-17.

Stacey, R. D. (2001). *Complex responsive processes in organizations: Learning and knowledge creation.* New York: Routledge.

What's New?

In the last issue we began an experiment in publishing structured, "bottom-line" abstracts of selected *JASIST* articles to improve dissemination of research findings that might be of general interest. The fact that an article does not appear here certainly does not mean that it is of no interest to practitioners. First, there was a start date when *JASIST* began notifying authors whose articles had been accepted of the opportunity to submit abstracts to the *Bulletin*, and no attempt was made to solicit retrospectively. Some articles may, therefore, have been accepted before we initiated this project. Second, submission is optional, and third, the Editor can only choose a few due to space restrictions. We would appreciate your comments and input to Bulletin@asis.org.

FROM *JASIST*, V. 53 (13)

Gu, Yinian (2002). An exploratory study of Malaysian publication productivity in computer science and Information technology, pp. 974-986.

Study and Results: A total of 547 unique Malaysian authors, affiliated to 52 organizations in Malaysia, contributed 197 (42.7%) journal articles, 263 (57.1%) conference papers and 1 (0.2%) monograph chapters between 1990 and 1999 as indicated by data collected from three Web-based databases. The results indicate that the scholars published in a few core proceedings but contributed to a wide variety of journals. Thirty-nine fields of research undertaken by the scholars are also revealed.

What's New? The paper presents Malaysia's contribution to world publication productivity in the fields of computer science and information technology for the period 1990-1999, and identifies the main interests of academic activity of Malaysian professional scholars. The findings and conclusion would definitely be informative for interested colleagues and researchers and can subsequently be used by funding agencies to ascertain the ratio of published output to fund allocations for the years under study to determine the benefits obtained. Moreover, the study of a country's scientific output does help to provide a general view of its scientific community's activity and contributions to world scientific literature.

Limitations: Due to constrained facilities, human and financial resource as well as obstacles of language, the investigation was restricted to the three international Web-based databases with selective coverage of academic publications. Therefore, some kinds of data, e.g., technical reports, dissertations and monographs, may have been missed.

Thelwall, M. (2002). Conceptualizing documentation on the Web: An evaluation of different heuristic-based models for counting links between university web sites, pp. 995-1005.

Study and Results: The individual pages of books or journals are rarely studied as entities in their own right, yet on the Web the page is the standard unit of content for the majority of research, as well as for online tools such as search engines. But should Web pages be aggregated, perhaps binding together all pages in the same site into a single document, in the same way that pages are bound into a book or journal? This is an issue particularly for those counting objects on the Web, such as how many links point to a given website. A common technique that can artificially inflate link counts is to place an identical navigation bar or on each page of a site: if there are a thousand pages with this device then one decision has created a thousand links. So is there a more useful Web document definition than the Web page? In the study, three alternative levels of document are defined for university websites based upon the directory, the domain and the whole institutional site. These are then compared on a set of 108 UK university websites under the assumption that a more effective document heuristic will tend to produce link counts that correlate more highly with institutional research productivity. The domain and directory models produced more statistically significant results, showing that the alternative definitions are both practical and useful.

What's New?: The document models introduced have the potential to add a new perspective to the Web for those that seek to measure it, assess its use or design information retrieval and storage tools for it.

Limitations: No simple document model can on its own eliminate all anomalies in Web publishing behavior. The data set for the study only covers the UK academic Web.