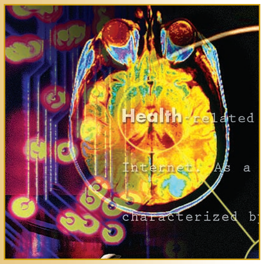

Integrating Knowledge on the Web

The Web provides a ubiquitous medium for seamlessly integrating distributed applications, formats, and content, making it well suited for enterprise knowledge management.

Amrit Tiwana
Emory University

Balasubramaniam Ramesh
Georgia State University

Organizations increasingly create massive internal digital repositories of customer, product development, and collaborative knowledge. While it is easy to convert raw input such as spreadsheets and design documents into storage-friendly formats, few organizations have been able to standardize use of platforms, networks, and operating systems, internally or interorganizationally. Knowledge management systems and knowledge networks attempt to mitigate these problems by seamlessly linking explicit, structured information across organizational rungs.[1]

While most of the enabling technologies for knowledge management have been around for many years, the ability to seamlessly integrate the components into a cohesive infrastructure evaded organizations until the advent of the Internet. Open Internet standards present unprecedented opportunities for implementing knowledge management solutions, and the Web provides a ubiquitous binding medium for integrating distributed applications, formats, and content.

In this article, we discuss a framework for characterizing knowledge management technology. We identify the desirable attributes of a knowledge management system and describe how Web-centric approaches can support these requirements. A sidebar reviews related research in enterprise knowledge management.

Framework for Knowledge Management

Knowledge management requires a suitable infrastructure for creating and managing tacit and explicit process and artifactual knowledge.[2] Although conventional enterprise information systems typically support *explicit* knowledge (that is, knowledge that can be explicated or codified), few support *tacit* knowledge (that is, knowledge that cannot be codified or expressed in written form). Providing pathways, channels, and mechanisms for sharing, distributing, and locating tacit knowledge sources is therefore a challenge that businesses look to technology developers to address.

Managing Knowledge in the Enterprise

Knowledge management systems integrate existing components at both the infrastructural and content levels, bringing

 1089-7801/01/$10.00©2001 IEEE

together the people and information systems associated with collaborative, knowledge-intensive tasks.[3] Tools that handle explicit, codifiable content (project management systems, data warehouses, and so on), as well as components that enable sharing and distributing tacit, or contextualized, content (for example, digital whiteboards, case-based reasoning tools, and concept-mapping systems), must be temporally and spatially integrated. Knowledge held at one location must be transparently accessible from other locations irrespective of protocols and formats. Recent research suggests that a process-centric approach to knowledge management might be more effective than an artifact-centric approach.[4,5] The Minimally Invasive Long-term Organizational Support (Milos) system, for example, uses a process-modeling language for knowledge management in software engineering.[6]

Table 1 shows the components of knowledge management systems and the organizational processes they enable. Together, these components provide integrative, complementary mechanisms for creating and managing knowledge. Table 2 lists several common objectives we found in our analysis of 12 knowledge management systems.[1]

Table 1. Organizational processes supported by the components of a knowledge management technology framework.

Component	Organizational process supported
Document management tools	Publishing, distribution
Workflow tools	Organizational procedures and routines
Transparent capture tools	Nonintrusive capture
Web conferencing tools/Expertise pointers	Communication, dialogue
VoIP telephones	Informal conversations
Electronic "watercoolers"	Conversations
Visual thinking tools	Knowledge maps
Decision support systems	Problem solving, decision support
Digital whiteboards/Shared workspaces	Brainstorming, tacit knowledge capture
Data warehouses/Webhouses	Operational data, data mining, knowledge discovery, validation, cleansing
Groupware, extranets	Collaboration, coordination
Intranets	Distribution, connectivity, publishing

Table 2. Knowledge management system objectives and corresponding technology enablers.

System objective	Technology enablers
Find knowledge	Knowledge bases; search-and-retrieval tools that scan both formal and informal sources of knowledge; employee skills directories.
Create new knowledge	Collaboration support tools; groupware; rationale capture tools; relational databases; decision repositories; externalization tools.
Package and assemble knowledge	Customized publishing tools; information-refinery tools; push technology; customized discussion groups.
Apply knowledge	Search, retrieval, and storage tools to help organize and classify both formal and informal knowledge.
Reuse and revalidate knowledge	Customer-support knowledge bases; services discussion databases; project databases and communities of practice.

Components of a Knowledge Management Network

Knowledge management involves collecting and assimilating information within informal and formal networks of people and artifacts spread throughout an organization. Knowledge management systems help establish, maintain, and support these networks.

Figure 1 (next page) shows the components that define the enterprise knowledge network. These three exhaustive but nonexclusive categories — tools, tasks, and sources — must be integrated at both the architectural and content levels. *Tools* are specific pieces of technology that facilitate execution of one or more *tasks* associated with applying knowledge. *Sources* feed raw data and information into the knowledge management system. Input data include distributed search and retrieval results, multimedia files, and transaction reports.

Multimedia information allows knowledge management systems to capture noncodifiable informal content that would otherwise be lost. It can also overcome some language barriers, which is especially helpful in transnational or cross-cultural project teams. For example, a multimedia clip of a ball-bearing mechanism in an automobile can convey information easier, faster, and at a lower cost than a written explanation.

The tools identified in Figure 1 can be categorized according to the specific knowledge-networking tasks they support. These tools contain components that serve generic functions:

- *Information mapping* components such as organizational memory systems link and map

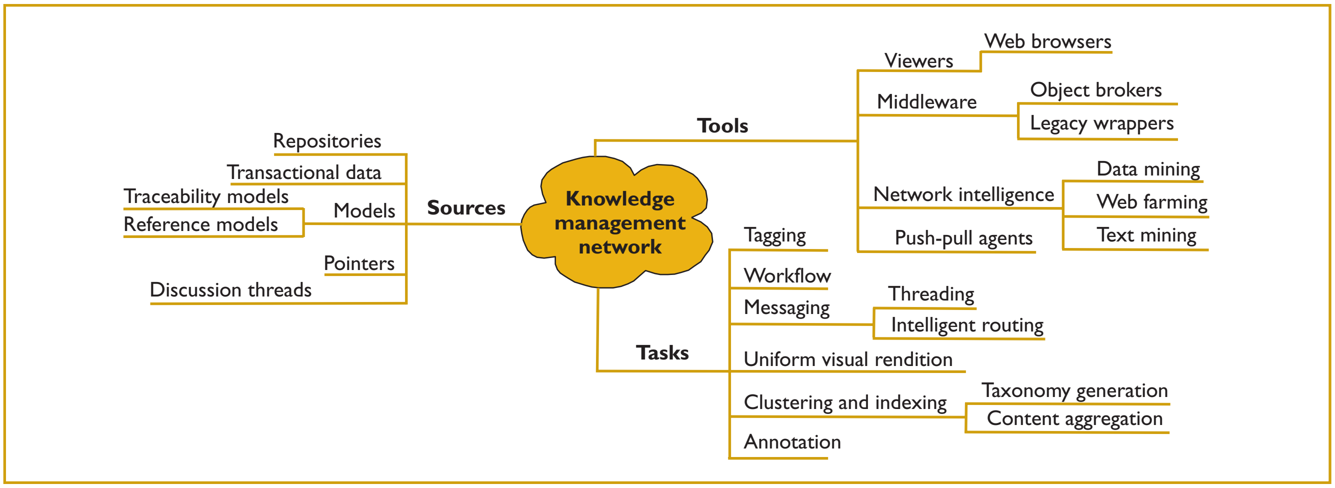

Figure 1. Enterprise knowledge management network components. Tools facilitate specific tasks. Sources feed raw data and information into the knowledge management system.

the flow of information that might later be converted to knowledge across the enterprise, and help create repositories with context.

- *Information and knowledge exchange* components such as annotation procedures that help create and share context by allowing people and systems to exchange explicit and tacit knowledge.[7] Subcomponents include collaboration tools and legacy-integrating middleware. On the tacit end are context-addition mechanisms (such as annotation), community networks, and shared applications on a network.
- *Knowledge flow* components provide paths for knowledge across organizations. Intranets and extranets provide paths for explicated knowledge; group support mechanisms and collaborative platforms provide paths for both explicated content and tacit context; and knowledge pointers direct users to the sources of tacit knowledge. Creating, sharing, and applying knowledge inherently involves collaboration. For example, brainstorming sessions, problem solving, and strategy-planning meetings are usually interactive and often involve people from different locations and functional backgrounds.
- *Intelligent agent and network mining* components include tools such as intelligent decision support systems, search engines, push- and pull-based intelligent agents, and automatic indexing tools. Filtering, editing, searching, and organizing (collectively called packaging) are essential to knowledge management system design. Packaging knowledge ensures that information is used to address actual business issues. Search tools integrate tacit knowledge and knowledge in databases, documents, and informal media. Without powerful search and intelligent retrieval tools to support meta-information creation, a knowledge management system is as limited as a traditional information system.

Middleware and Integration

Most contemporary approaches to knowledge management system design rely on a Web interface because it offers unrestricted possibilities for uniformly accessing knowledge across a variety of mobile and stationary platforms. Even within Web-based knowledge management systems, new information is generated daily and needs to be integrated with existing data. Knowledge management components must be seamlessly integrated so that adding new content to the repository is as efficient and nonintrusive as possible.

Integrating explicit knowledge distributed across incompatible platforms, mainframe legacy systems, and proprietary or Web-unfriendly applications is challenging. Several mechanisms for achieving such integration are worth mention.

Legacy integration. Legacy integration mechanisms connect legacy data with existing and new systems. A "wrapper" provides connectivity between old and new data formats, often through a Web front end. Although this approach is common in systems integration, it has not been adapted to knowledge management systems.

Wrapper complexity depends on the structural complexity of the data sources to be integrated. Several scripting languages can be used to integrate data sources such as those on mainframes. For example, Knowledge Query Modeling Language (KQML) supports knowledge sharing among

Related Work in Knowledge Networking

Effective knowledge management must integrate, rather than separate, technology-focused research and process-focused social sciences. Here we review related work and identify some promising research directions.

Intelligent Information Integration

Intelligent integration of knowledge — going beyond integrating databases — increases the value of the information accessed from multiple sources. Some current systems act as intelligent intermediaries between users and various sources. These systems include

- *facilitators* that search for potential sources of information and mechanisms for accessing them;
- *query processors* that reformulate user requests to improve the prospects of accessing relevant information;
- *mediators* that combine and summarize information from a variety of sources; and
- *data miners* that discover interesting patterns.

Whereas facilitators rely on automating mediation tasks to the extent feasible, mediators typically require human intervention when changes to resources, user requirements, or process-specific knowledge occur.

Mediator Architectures

A mediator architecture provides intermediary services that link data sources and application programs so that information flows across heterogeneous sources are facilitated without the need to integrate the base data sources. The Stanford-IBM Manager of Multiple Information Sources (TSIMMIS) system[1] uses a mediator architecture to integrate data drawn from multiple and often heterogeneous sources. Unlike other mediator systems that focus on the information content of the sources and their relationships to the integrated view users want, TSIMMIS relies on understanding the capabilities of the sources so that only queries with valid answers are generated.

The Information Manifold[2] project is similar in spirit. It uses source descriptions to prune the set of information sources for a given user query. However, whereas TSIMMIS uses query templates to describe source capabilities and generates an efficient plan for satisfying user requests, Information Manifold uses binding patterns associated with views exported by the sources. By using source query capability descriptions that can describe an infinite set of queries, TSIMMIS offers very high flexibility.

Garlic[3] takes an approach comparable to TSIMMIS**.** The project integrates multimedia information systems (including application-specific data such as CAD drawings, medical objects, and maps) from both database and nondatabase sources using an object-oriented data model and an object-oriented dialect of SQL. A middleware layer allows efficient query processing and data access.

In contrast, the Hermes[4] project develops a general, declarative language for defining a mediator. Here, heterogeneous information sources are treated as domains capable of executing certain functions with prespecified input and output parameters. Though this approach is restricted to integration through a limited set of parameterized calls, it is generic enough to integrate information from a wide variety of sources such as databases, software packages, and reasoning paradigms (for example, path planning and terrain reasoning).

XML / RDF in Integration

The Mediation of Information using XML[5] (MIX) project views the Web as a distributed database and XML (or its modifications and extensions) as its data model. Because many Web sources are likely to export XML views of their data, the semantic descriptions of their content, and interface descriptions (as XML queries), project members have developed the XMAS query language to integrate information from these sources. When data repositories are not converted to XML, this approach provides wrapping technologies for translating XMAS queries or commands that are understood by the underlying systems.

The On2broker[6] provides brokering services to access semistructured information on the Web based on HTML, XML, and RDF content descriptions. This approach relies on ontologies to make the semantics of Web documents explicit and uses semantic information to answer queries. It provides an interface for formulating queries, an information agent for collecting required knowledge from the Web, an inference engine for deriving appropriate answers to queries, and a database for caching semantic annotations.

Decker et al.[7] discuss the role of XML and RDF in creating a "semantic Web" in which knowledge can be stored in a machine-processible form. They argue that XML is unlikely to be an effective

continued on p. 36

applications, especially in cooperative problem solving.[8] Wrappers can also be written in Perl or Python, or with wrapper tools like those in the Araneus toolkit (Minerva+Editor)[9] and the Stanford-IBM Manager of Multiple Information Sources (TSIMMIS) project.[10]

Message brokers. Traditional middleware such as message queuing, transaction processing, and publish-subscribe engines solve only part of the knowledge management system-integration problem. Message brokers, described as the "middleware of middleware,"[11] provide a central integration point that seamlessly works with approaches like XML, RDF, and knowledge servers by intelligently coordinating routing, reformatting, and traffic flow. Message brokers are based on asynchronous store-and-forward logic that uses abstraction techniques to provide any-to-any connectivity (as opposed to one-to-many or many-to-many techniques used in

Related Work in Knowledge Networking (continued)

continued from p. 35

tool for semantic interoperability and propose a method for encoding ontology representations into RDF. We, however, note that market forces behind XML may pose challenges to the successful adoption of this approach.

Collaborating Agents

In the InfoSleuth project,[8] collaborating agents gather knowledge from a changing set of databases and semistructured text repositories across the Internet. Broker agents process user requests made against a domain ontology by finding other agents that can satisfy them. Resource agents support the mapping between the user queries and the information resources. Numerous other agents perform tasks like data aggregation, event detection, and query distribution.

Digital Libraries

Projects under the NSF/ARPA/NASA Digital Library Initiative (DLI)[9] also address the problems of accessing and using distributed information resources. Many efforts under this initiative integrate text, graphics, and images from sources that are typically heterogeneous, redundant, and without common structure. The projects demonstrate approaches for managing large amounts of heterogeneous information, and thereby facilitate search and manipulation by organizing it into interoperable and coherent sources.

For example, the Interspace project[10] aims to evolve the Internet into a virtual hyperspace using intelligent search and display of structured documents. Using automatic thesaurus-generation techniques, concept spaces are created as graphs of domain-specific concepts and their weighted co-occurrence relationships. Merging these concept spaces and providing traversal paths across different concept spaces facilitates large-scale information retrieval and integration.

Organizational Memory Systems

REMAP[11] supports knowledge networking among distributed collaborative development team members by capturing organizational knowledge including some noncodifiable tacit knowledge. The system provides a Web-based group-support environment in which participants can deliberate and make product development decisions. A distributed multimedia annotation system facilitates thick descriptions (using rich media such as video and audio) of knowledge, and a reason-maintenance system detects and resolves conflicts to maintain captured knowledge consistency. REMAP is integrated with collaboration tools like Microsoft's NetMeeting to facilitate synchronous knowledge exchange as well as with tools used in the normal work environment, such as Rational Rose and Microsoft Office, to help capture and use knowledge without loss of context.

The Stars project[12] takes a similar but more formal approach to representing the different perspectives of design team members and provides mechanisms for managing conflict and integrating knowledge.

References

1. H. Garcia-Molina et al., "The TSIMMIS Approach to Mediation: Data Models and Languages," *J. Intelligent Information Systems*, vol. 8, 1997, pp. 117-132.
2. A. Levi, A. Rajaraman, and J. Oedille, "Querying Heterogeneous Information Sources Using Source Descriptions," *Proc. Very Large Data Bases Conf.*, Morgan Kaufmann, San Francisco, Calif., 1996.
3. M.J. Carey et al., "Towards Heterogenous Multimedia Information Systems: The Garlic Approach," *Proc. Fifth Int'l Workshop Research Issues in Data Eng.* (RIDE), IEEE CS Press, Los Alamitos, Calif., 1995, pp. 124-131.
4. A. Brink, S. Marcus, and V.S. Subramaniam, "Heterogeneous Multimedia Reasoning," *Computer*, vol. 29, no. 9, Sept. 1995, pp. 33-39.
5. C. Baru et al., "XML-Based Information Mediation with MIX," *Proc. ACM SIGMOD*, ACM Press, New York, 1999, pp. 597-599.
6. D. Fensel et al., "On2broker: Semantic-Based Access to Information Sources at the WWW," *Proc. World Conf. WWW and Internet* (WebNet), Assoc. for Advancement of Computing Education, Norfolk, Va., 1999, pp. 366-371.
7. S. Decker et al., "The Semantic Web: The Roles of XML and RDF," *IEEE Internet Computing*, vol. 4, no. 5, Sept.-Oct. 2000, pp. 63-74.
8. J. Fowler et al., "Agent-Based Semantic Interoperability in InfoSleuth," *SIGMOD Record*, vol. 28, no. 1, 1999, pp. 60-67.
9. B. Schatz and H. Chen, "Building Large Scale Digital Libraries," *Computer*, vol. 29, no. 5, May, 1996, pp. 22-26.
10. H. Chen et al., "A Parallel Computing Approach to Creating Engineering Concept Spaces for Semantic Retrieval: The Illinois Digital Library Initiative Project," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, Aug. 1996, pp. 771-782.
11. B. Ramesh and A. Tiwana, "Supporting Collaborative Process Knowledge Management in New Product Development Teams," *Decision Support Systems*, vol. 27, nos. 1-2, 1999, pp. 213-235.
12. S.C.-Y. Lu and J. Cai, "STARS: A Socio-Technical Framework for Integrating Design Knowledge over the Internet," *IEEE Internet Computing*, vol. 4, no. 5, Sept.-Oct. 2000, pp. 54-62.

traditional middleware) among diverse components and subsystems. Although several commercial initiatives have emerged, no set of accepted standards exists in the middleware industry.

Extensible Markup Language. XML was designed with two objectives:

- to increase the degree of control over how documents are presented on the Web, and
- to explicate the standards for exchanging information that is structured for further processing.

Quite aptly, XML has been described as "a philosophy of data identification and a host of languages and specifications for functions related to data."[12] XML extends the concept of documents beyond memos and reports to virtually any information that can be platform-independently aggregated across distributed systems, and thus can poten-

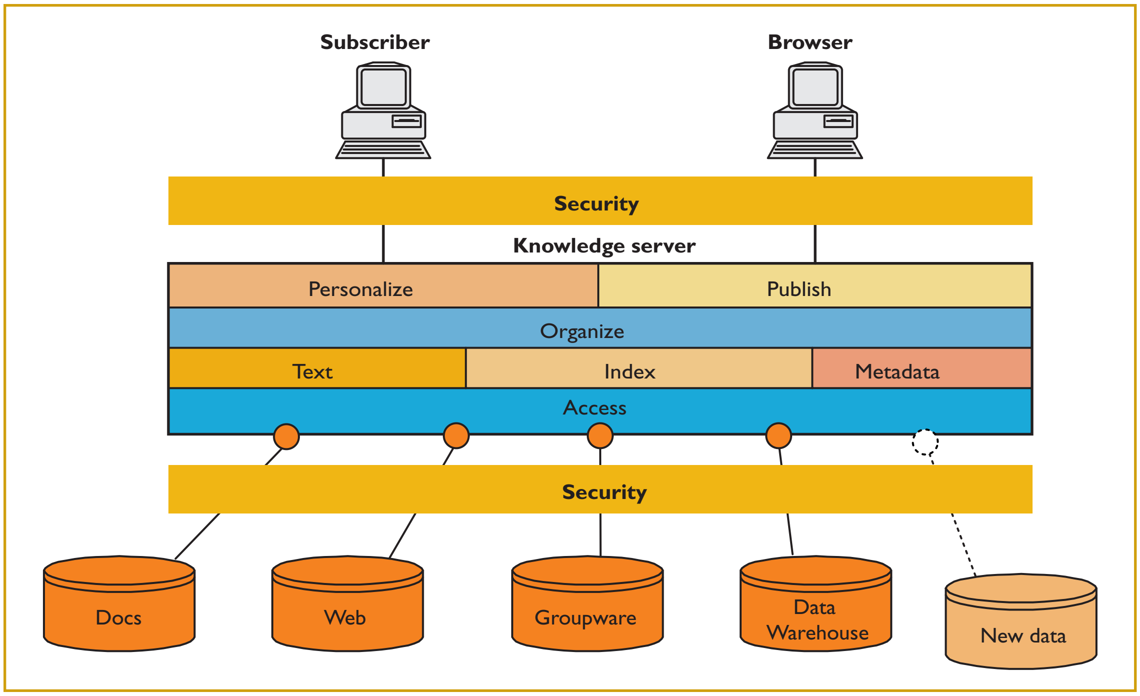

Figure 2. A knowledge server and remote data sources. Knowledge servers facilitate integration, organization, and aggregation of spatially distributed explicit content and tacit knowledge pointers.

tially integrate knowledge management systems into myriad e-business applications.

Using customizable application-specific tags that describe a document's structure and the method for displaying it, XML makes it easier to search and retrieve documents containing explicit knowledge and to device-independently format and track documents. Intelligent agent-based retrieval technologies can efficiently search highly distributed repositories of knowledge stored in XML documents.

Resource Description Framework. RDF is an emerging standard for processing metadata. It provides interoperability between applications that exchange machine-understandable information on the Web.[13] RDF's broad goal is to define a mechanism for describing resources that makes no a priori assumptions about a particular application domain or the associated semantics. In knowledge management systems, RDF can automate resource discovery (explicit or codified knowledge), cataloging, and content rating. RDF draws upon knowledge-integration concepts from several domains including structured document composition (for example, SGML), object-oriented modeling, and knowledge engineering.

Knowledge servers. A knowledge server is a Web server or set of servers that acts as a knowledge aggregation hub. Figure 2 shows how knowledge servers mediate aggregation of explicit knowledge from myriad sources. In a knowledge server, plug-in components or *accessors* periodically poll remote data sources for new content.[1] The knowledge server creates a reference, or *card*, for each new document to capture key metadata such as author, subject, and title. Each card maintains a link to the original content, which the knowledge server indexes in a text search engine. The knowledge server uses each card's text index and metadata properties to organize them in a hierarchy of administrator-defined categories that users can view through a Web browser. In addition, knowledge servers direct users to sources of tacit knowledge when the limits of codification are reached. This is often implemented in the form of skill and resource directories, which can be automatically generated by associating content with its producers, authors, or collaborators.

Attributes of a Web-Based System

Based on the architecture described in the previous section, we have identified several desirable attributes of a knowledge management system. Here, we describe how Web-based protocols can support them.

- *Platform independence and portability.* As a platform-free application, the Web lets knowledge management systems interoperate across diverse platforms and operating system environments within the enterprise.

- *Robust access.* If Web-centric protocols are used as the basis for enterprise-level integration, the knowledge management system can take advantage of the Internet's global network backbone, which offers multiple, redundant, and robust paths for moving sensitive data reliably — even if networks fail at multiple points.
- *Integration with legacy and existing systems.* A collaborative knowledge-sharing platform can use wrappers to access legacy data through Web browsers, and HTTP and associated open Web protocols to integrate this data across the enterprise.
- *Security.* When used as the primary access mechanism, Web browsers with built-in encryption provide a layer of security; therefore, a Web-based platform minimizes long-term security costs by tightly integrating virtual network tunnels with the existing TCP/IP network.
- *Scalability.* The Internet offers a convenient platform for building systems using established methods for managing Web server scalability.
- *Distributed connectivity, flexibility, and customizability.* Distributed resources and databases can be interconnected cost-effectively and reliably using virtual networks tunneled within the Internet, and Web browsers provide a ubiquitous interface that is easily customized to support multiple languages, regional preferences, and features across the enterprise.
- *Ubiquitous, consistent, intuitive client interface.* A Web interface allows a variety of applications to run from almost any platform and provides a universal and lightweight client through which end users can run applications and access repositories without switching to an unfamiliar operating environment.

Web-based integration facilitates simultaneous realization of these attributes in an efficient and cost-effective way.

Challenges and Opportunities

Many of the recent technical advances in knowledge networking have concentrated on managing, exchanging, and integrating explicit knowledge. We believe, however, that these advances are but the tip of the iceberg for technology-facilitated knowledge networking's potential. Attempts to simultaneously integrate tacit and explicit knowledge have been sparse,[14] and the few attempts to specify organizational-level architectures for tacit-explicit knowledge conversion and reconversion have been at a high level of abstraction.[15] Interdisciplinary collaboration can potentially yield technical advances by allowing technology-focused researchers to address — and perhaps incrementally solve — issues that social science researchers can only identify.

Recently published reviews of social, cognitive, managerial, technical, and interdisciplinary approaches provide a good starting point for exploring the integration of these rich research traditions and approaches to creating knowledge networks. Specifically, the technical community can apply principles that facilitate the transfer of tacit knowledge in settings such as finance, electronic publishing, and distributed and collaborative design to future research on knowledge networking.

Our ongoing work is exploring how knowledge management technologies impact the performance of temporary, distributed teams executing projects on near-vertical learning curves. We are also developing nonintrusive mechanisms to facilitate knowledge capture, transfer, and use in complex organizational problem-solving situations such as complex systems development, collaborative design, and virtual teamwork.

Acknowledgments

This research was funded in part by a grant from the U.S. National Science Foundation, the U.S. Office of Naval Research, and the Department of Defense DAU EARP project.

References

1. A. Tiwana, *The Knowledge Management Toolkit: Practical Techniques for Building a Knowledge Management System*, Prentice-Hall, Upper Saddle River, N.J., 2000.
2. S. Heijst et al., "The Lessons Learned Cycle," in *Information Technology for Knowledge Management*, U. Borghoff and R. Pareschi, eds., Springer-Verlag, Berlin, 1998, pp. 17-34.
3. F. Maurer and W. Regli, "Weaving the Web of Reason," *IEEE Internet Computing*, vol. 4, no. 5, Sept.-Oct., 2000, pp. 50-53.
4. P. Saviotti, "On the Dynamics of Appropriability of Tacit and of Codified Knowledge," *Research Policy*, vol. 26, 1998, pp. 843-856.
5. R. Fielding et al., "Web-Based Development of Complex Information Products," *Comm. ACM*, vol. 41, 1998, pp. 84-92.
6. F. Maurer et al., "Merging Project Planning and Web-Enabled Dynamic Workflow Technologies," *IEEE Internet Computing*, vol. 4, no. 3, May-June 2000, pp. 65-74.
7. C.W. Choo, *The Knowing Organization*, Cambridge Univer-

sity Press, New York, 1998.
8. Q. He, K.P. Sycara, and T.W. Finin, "Personal Security Agent: KQML-Based PKI," *Proc. Second Int'l Conf. Autonomous Agents*, ACM Press, New York, 1998.
9. G. Mecca, P. Merialdo, and P. Atzeni, "ARANEUS in the Era of XML," *IEEE Data Eng. Bulletin*, 1999.
10. J. Hammer et al., "Template-Based Wrappers in the TSIMMIS System," *Proc. SIGMOD*, ACM Press, New York, 1997, pp. 532-535.
11. D. Linthicum, *Enterprise Application Integration*, Addison Wesley Longman, Reading, Mass., 2000.
12. T. Usdin and T. Graham, "XML: Not a Silver Bullet, but a Great Pipe Wrench," *ACM Standard View*, vol. 6, 1998, pp. 125-132.
13. O. Lassila and R. Swick, "Resource Description Framework (RDF) Model and Syntax Specification," World Wide Web Consortium, Feb. 1999; available online at http://www.w3.org/TR/REC-rdf-syntax/.
14. P.N. Robillard, "The Role of Knowledge in Software Development," *Comm. ACM*, vol. 42, 1999, pp. 87-92.
15. M. Zack, "An Architecture for Managing Codified Knowledge," *Sloan Management Rev.*, 1999, pp. 45-48.

Amrit Tiwana is finishing up his doctoral work at the Robinson College of Business at Georgia State University. In August 2001, he begins his appointment as an assistant professor at Goizueta Business School at Emory University. His research interests include knowledge management in innovative e-business projects and knowledge integration in business networks. He is a member of the Academy of Management, the IEEE, and the ACM.

Balasubramaniam Ramesh is associate professor and director of research and sponsored programs in the Computer Information Systems Department at Georgia State University. His research interests include supporting complex organizational processes such as requirements engineering and traceability with decision support and knowledge systems, data mining, and e-services. His work has been supported by several major funding agencies such as the U.S. Office of Naval Research and the U.S. National Science Foundation, and adopted widely in practice. He is a member of the ACM, the IEEE, and AAAI.

Readers can contact the authors at the Goizueta Business School, Emory University, 1300 Clifton Road, Atlanta, GA 30328, atiwana@acm.org, and Computer Information Systems Department, 35 Broad Street, Georgia State University, Atlanta, GA 30303; bramesh@gsu.edu.