BRIEFINGS IN BIOINFORMATICS. VOL 8. NO 3. 141–149
Advance Access publication May 14, 2007
doi:10.1093/bib/bbm013

Knowledge networks in the age of the Semantic Web

Eric Neumann and Larry Prusak

Submitted: 13th February 2007; revised: 13th April 2007

Abstract

The Web has become the major medium for various communities to share their knowledge. To this end, it provides an optimal environment for knowledge networks. The web offers global connectivity that is virtually instantaneous, and whose resources and documents can easily be indexed for easy searching. In the coupled realms of biomedical research and healthcare, this has become especially important where today many thousands of communities already exist that connect across academia, hospitals and industry. These communities also rely on several forms of knowledge assets, including publications, experimental data, domain-specific vocabularies and policies. Web-based communities will be one of the earlier beneficiaries of the emerging Semantic Web. With the new standards and technologies of the Semantic Web, effective utilization of knowledge networks will expand profoundly, fostering new levels of innovation and knowledge.

Keywords: *knowledge networks; Semantic Web; communities of practice; life sciences; social web; working knowledge*

INTRODUCTION

We are all today very connected by the digital world; our lives are interwoven with the applications of computers and the Internet, at home, at work and in our community. These technologies are used both in the everyday exchange of friendly messaging, as well as in more goal-oriented tasks by coordinated groups, who understand the critical value of the knowledge contained within the group. This latter category of users is of particular interest since it has far reaching consequences for our society, and has the potential of advancing us as much as the invention of the printed word had done.

Knowledge, as defined in this article, is the human resource that enables the capability to take action in uncertain and novel situations. It is always contextual and local yet may rely on artifacts and 'intermediate products' that are persistent and media based. Consequently the capability that results from individual and collective knowledge is not a compendium of compiled facts (as in a database), but the potential for decisive and influential actions that serve a larger community or organizational need. Yet, as we shall see, the growth, use and flow of knowledge in many practices will depend on both technological and social modalities. Although often misunderstood, information is not knowledge; it can be better understood as a system comprised of a message, a sender, multiple receivers and some coding that allows it to be communicated (Information Theory, [1]). Information becomes knowledge when it is absorbed and socialized by an individual or group. It then becomes part of the person's knowledge resource base. It can be said that knowledge is what gives information practical meaning.

To take a particular example, companies today have made the transition to a paperless (or more accurately, paper as secondary) office, where communication and protocol are maintained and accessed via digital media and networks. Examples include email, text documents, presentation documents, financial spreadsheets, intranet web-sites and corporate databases. Most business processes and documents reside as digital constructs that can be

Corresponding author. Eric Neumann, 14 Colony Road, Lexington, MA 02420, USA. E-mail: eneumann@teranode.com

Eric K. Neumann is Senior Director of Product Strategy with Teranode Corporation, and is also founder and co-chair for the W3C Semantic Web Healthcare and Life Science Interest Group (HCLSIG) that is bringing together industry leaders to inform them to the benefits of semantic web technologies.

Larry Prusak is a researcher and consultant and works with many national and international organizations on knowledge utilization. He currently co-directs 'Working Knowledge', a knowledge research program at Babson College, where he is a Distinguished Scholar in Residence.

readily obtained by employees via a company's intranet. In such organizations, the role of knowledge workers must include their ability to navigate, utilize and contribute to this network asset space (NAS). Indeed, web-based navigation of these spaces has hit a new level with the rise of powerful search engines that anticipate some of the contextual relevance of which knowledge workers need to take advantage [2]. Yet an identified limitation of these search tools is their neglect of the semantics contained within the content. Semantics is tied to the associated meaning structure a community has around its knowledge, such as 'genes have to be encoded within an organism's genome'. This approach requires the challenging task of harmonizing meaning around a common set of semantics, which would then enable semantic interoperability across many applications.

The issue of semantics has recently become center stage in the latest advancement of the World Wide Web, and over the last 10 years the web has acquired new enhancements that support the embedding of semantics in both documents and other information resources on the web [3]. Although this allows the formal codification of information, we believe that this will not lead to the codification of knowledge. Rather it will produce a larger benefit: the facilitation of community knowledge networks and the exposing (and thereby freeing) of contextual artifacts of various knowledge processes. This will help prevent what has been previously described as 'knowledge cul-de-sacs', the misplacement and mislabeling of important information and the links to the knowledge experts [4].

A great diversity of knowledge communities exist (e.g. software engineering, religions, organic living, archeologists, astronomy). In this discussion, we will focus on the scientific research communities that span basic life science research, drug R&D, and health care clinics. These constitute a complex set of active communities that interact together in a rich network of knowledge exchanges. For example, biomedical research is funded in disease areas driven by unmet medical needs. The success of such supported projects depends on available knowledge and the application of the scientific process; this eventually leads to new hypotheses that are tested through experimentation, and then are applicable to future research. In this way new biomedical knowledge is gained that is 'actionable', to be used in (i) determining the cause of a disease, (ii) identifying potential disease treatments, (iii) diagnosing diseases and staging them accurately, (iv) providing resources that help in the prevention and management of a disease and (v) seeking the best line of healthcare for treatment of the disease. The proper creation and use of knowledge can help fill various innovation gaps in healthcare and pharmaceutics, finding more optimal solutions for medical and industrial processes and thereby reducing inefficiencies in systems that burden our society.

BASIC THEMES

We see three main elements as keys to the changes that will support greater knowledge recombining between communities and the unlocking of digital assets.

Distributed (global) innovative communities of practice

At the center of all knowledge practices is the community, a living entity that is defined by the purposeful interactions of its participants, and whose definition evolves throughout its lifecycle. Their goals can only be attained by the collective development and sharing of their knowledge. The community defines the knowledge and the knowledge binds the community together. Successful communities eventually evolve roles and practices that work in synchrony with the local knowledge and leverage it effectively to solve problems or deeper community objectives (e.g. the creation of an open source of tools or the search for truth). As the practice defines these communities rather than a formal structure, these are referred to as communities of practice.

Curated knowledge assets and by-products

The non-human things that form a community's knowledge resources come in two major forms: those requiring human interpretation for proper cataloging and application (aka unstructured information); and those with some level of regular formatting (data columns, relations, or tags) whose structure imposes a meaning onto the content (e.g. column headers or fields that are understood only by select members). The assets can be made available through a range of interfaces: web pages, portals, newsfeeds, databases, knowledge bases and semantics models (data models, ontologies). Even unstructured or 'image' resources often can have metadata associated that help describe who created the resources, when, and what kind of data they contain (i.e. text, audio, or video).

Technology-based environments that facilitate community activities

A distinction is made here between knowledge reasoning technologies (derived from the Artificial Intelligence goals) and technologies that enable knowledge networks. Both can utilize knowledge representations, but the latter directly supports community functions. In this article, we will focus entirely on the latter, and avoid arguments regarding whether or not complete knowledge can be represented in knowledge bases. The web and the software that enables it are in most cases tools designed for people and social organizations.

COMMUNITIES OF PRACTICE

The first component is concerned with the social aspect of knowledge, but it is not an *ad hoc* organization, nor does it exist without a structure. Such communities of practices, especially in the life sciences, rely on what their members identify are common ideas, leading them to define what the core concepts and premises are. The core of any true community is that it is self-organizing and self-monitoring. It is held together by common tasks, passion of a subject, or situations that demand commonality (such as that experienced by parents of sick children and their support groups.) The social norms that develop in school communities are as strong as, if not stronger, then more formal organizational rules and give these entities their particular force and appeal. The collective knowledge can only be utilized if they are sufficiently described using a common basis of meanings or *semantics* held by the community's members, such as what is meant be *effective* and *supported* as in: '*Treatment of Parkinson's Disease by L-DOPA is not effective, therefore Stem Cell research needs to be supported*'. Defining common semantics is guided by the group and reinforces group membership.

Examples of communities of practices include the open-source code initiatives (www.linux.org), research data consortia (www.openmicroscopy.org), ontology working groups (www.biopax.org) and disease-support groups. Communities that actively contribute and compile knowledge as part of objectives create what are known as 'Commons'; these resources are openly available for public use (though there may be restrictions on their commercial use), but are not always in the public domain. Examples include Science Commons (www.sciencecommons.org) and the Alzforum (www.alzforum.com).

Communities of practice can become formalized and transform into consortia and alliances, but not every communities of practice necessarily needs to become formalized. Many continue to serve important functions as informal communities.

Negotiation of meaning

Knowledge is not simply produced and exchanged; it must be presented in a way that others can comprehend and work with it. If the word 'meaning' is a close synonym for knowledge then the negotiation of this very meaning is at the core of working with knowledge. This negotiation requires a series of interactions among the members that effectively define the necessary meaning around various concepts and words. The meaning around these ideas can only be established in a community context whose members are involved in similar activities and share similar values. Wenger [5] describes this process as 'Negotiation of Meaning'; the production of meanings 'that extend, redirect, dismiss, reinterpret, modify or confirm . . . the histories of meanings of which they are a part'. This is strongly illustrated in the development of the Gene Ontology (GO) by the GO community (www.geneontology.org).

Negotiation of Meaning is the process by which semantic meaning is established in a given domain, resulting in the common usage of terms and ideas by the community to further explore new ideas. Negotiation of Meaning is also very much tied to the notion of semantic interoperability as applied to data models, since the creation and use of data models depends on agreed common usage. This also suggests that the same word can have different meanings depending on the community. Indeed this can be seen around the various definitions of 'gene', which depends on whether we are bioinformaticists (gene $\sim$ sequence), human geneticists (gene $\sim$ disease inheritance) or plant geneticists (gene $\sim$ crop attributes). Hence Negotiation of Meaning establishes semantic models for different communities of practice that are not uniquely linked to terminology (and why today's search engines based on word spelling are of limited use by communities of practices).

This, of course, has an impact on how information is represented, and explains why information systems and databases are locally useful, but cross-community brittle. As Moen [6] exclaims: 'Within a community or domain, relative homogeneity reduces interoperability challenges. Heterogeneity increases as one moves outside of a focal community/domain, and interoperability is likely [to be] more costly and difficult to achieve.' This speaks directly to all web-based assets (see further) and sources of data, and why a term like *skin rash* can either mean general red bumps on the skin or something more specific (seborrheic dermatitis) depending on the clinician. The direct consequence is that there can never be an over-arching semantic model that will work for all individuals (or even across all biology). Rather there will be a distributed network of many semantic constructs that may occasionally overlap each other on common agreed points (e.g. '*all organisms contain a set of inherited genes*').

Negotiation of Meaning is specifically useful for forming common (controlled) vocabularies, and more specifically, ontologies. These are the coded artifacts of a community's working practice, whose creation required in-depth negotiation by the community. This is the case for the GO, where scores of geneticists have labored to create a consistent system of biological descriptors. The artifacts represent a text-coded expression of the community's views of comprehending genetic and molecular functions. The very act of negotiating and coding the meaning reifies and enforces an understanding between community members and allows them to converge on meanings in new areas. As we shall see later, this has a direct bearing on how communities are able to work together over the Internet.

Provenance and ownership

People generally do not share their knowledge without some for of incentive to do so. These incentives include reputation, reciprocity, altruistic, and can of course be used to establish trust and a sense of belonging. At the minimum, knowledge sharers need to be acknowledged for their contributions, along with assurance that their work will not be misused or misappropriated. Guaranteeing to record the provenance of any contribution is therefore an essential part of a community of practice, helping strengthen the connection each has to the community by virtue of others valuing their contributions. It also ensures that the quality or validity of a contribution is maximized, and in cases where the contribution is ongoing (curation projects, expanding resources), it motivates them to ensure that the resources will be sufficiently maintained over time, imbuing a form of robustness onto a community.

Group identity and common knowledge

A group is often identified as much by its collective objectives as by its composition; this translates into the desire to form what is known as *common knowledge*. This can be illustrated by the following example: if Tom, Angela and Kim all know that 'avian flu is transmitted by birds' and each knows the other knows this as well, (definition of common knowledge), then it is also common knowledge that 'Tom also knows that Angela and Kim both know this same fact'; moreover, it follows that Angela knows, that Tom knows, that Kim knows the fact. In fact, this 'who knows what' series is infinite for any combination of players that have common knowledge: T knows (A knows(... (K knows⇒'Avian flu ...')...). In other words, the entire group knows at once that all deducible knowledge on a common subject is deducible by all the members, and that even this fact is also commonly known. This establishes a criterion for group identity (*we all know this*) and helps them realize what they can do together (we know what knowledge can be leveraged); hence the knowledge defines the group as much as the group defines the knowledge. In earlier civilizations, this manifested itself as secrets societies, whose secrets were all known to only its members. Today it is more about managing growing amounts of complex knowledge that has practical application, and making sure the entire community is 'up to speed' with the knowledge. A contemporary illustration of common knowledge is the use of the 'cc' field in email messages, used to ensure that a group of individuals have the same knowledge about a particular message (i.e. reducing the chance of misinterpretations and misquoting); even the sociology of who is included in a cc is quite intricate. Common knowledge therefore is a key element of any communities of practice, and its recognition and continuous reaffirmation (towards educating new members) plays an important, dynamic role for any community.

Standing on shoulders

In the biomedical ecosystem, it is broadly recognized that isolated groups or 'knowledge silos' will not be successful or innovative, as compared to those groups that share knowledge. In science, there is a strong requirement to be '*boundaryless*' [7] while still retaining the values that adhere to more local community structures, such as identification with colleagues close by and an esprit-de-corps that should not be overlooked. Knowledge is closely linked to emotions (see Descartes' Error by Damasio [8]) and emotions are more likely to be reinforced locally. What is crucial is that the knowledge is readily available for re-use by other groups, and the knowledge is used as the basis for further knowledge growth, resulting in an ever-growing network of new insights and hypotheses (a partial manifestation of this is the network of citations). In fact, most recent advances in biomedical sciences have an intricate dependency both on technology and on biological theory that comes from multiple communities. Each discovery stands on the shoulders of previous successes (The original quote is by Sir Isaac Newton: 'If I have seen further it is by standing on the shoulders of giants . . .'). This group-knowledge is like an evolving web, and is becoming more and more realized as such through the augmenting knowledge networks that exist on the web.

KNOWLEDGE NETWORKS AND ASSETS

Basic connectivity: mailing lists

This form of communication has been around for many years, offering an easy to use mechanism for people to communicate with each other, in real-time or close to real-time. Email is directly derived from the more traditional 'exchanging of letters' modus operandi, though now mail can be broadcast, via mailing lists, to many thousands of individuals. This allows many to receive the messages (since message copying is virtually free), but it does severely restrict how a list of '*n*' participants can contribute simultaneously since every response has the potential for $n-1$ more responses, and so forth (typically only a few active individuals respond to a specific topic). Often this results in misunderstandings that erupt emotionally as 'flame wars'. Nonetheless it does allow rapid dissemination of topics, and helps establish focus and common knowledge. On the other hand, it is not an effective way for cataloging various topical threads since message do not easily sort themselves by topic.

Collective knowledge: blogs, wikis and tagging

The next level beyond email and instant messaging is the formation of collective spaces such as blogs and *wikis*, whose rise illustrates the continuing evolution of the web in a social context. Web logs or *blogs* are a growing social phenomenon [e.g. mySpace (http://www.myspace.com), blogger (http://googleblog.blogspot.com)] whereby a person publishes their ongoing thoughts and opinions via an updated web site. But while blogs are usually personalized, wikis are community oriented and reflect complex group interactions. Here the group aligns its actions to develop a space to hold and display their collective knowledge (text, images and additional links). An impressive illustration of this can be seen with the popular and richly endowed Wikipedia (www.wikipedia.org), whose content aims to reflect authoritative views and whose quality is comparable to more traditional encyclopedic sources [9]. It is this visibility of communal contributions that attracts and engages authors on an unprecedented scale from all over the world. Within a wiki community, each page holds a piece of a subject story. Members of this community organically manage the story, correcting any inconsistencies and augmenting it as new material is uncovered. Wikis have sprung up around hundreds of thousands of different communities, and appear to be gaining in stature as the primary Internet collective space for communities of practice.

Another example of a collective space is the ability to 'tag' publication (Nature's connotea.org) or photo contributions (www.flickr.com), so that others can find these and related resources, as well as who identified and tagged these, through the tag associations. Tags are usually not part of controlled vocabularies, but are simple forms that can be created, shared and re-used as *ad hoc* 'categories for bookmarks', similar to *keywords*. They are dynamic and local, and form what have been termed *folksonomies*: a basis of simple vocabularies that arise within a given community. However, their strength of being easy to create is also their weakness in enterprise systems, where some degree of codification needs to be defined ahead of time and used in basic operations that require some level of compliance. Nonetheless, they allow for rapid terminological sharing and equilibration within communities of practices, and could support the early stages of more rigorous Negotiation of Meaning.

A missing ingredient in the collective space phenomenon is the need to guarantee persistence of documents and links; knowing that something seen one day will be there in a month or year is fundamental to most business contracts. In enterprise environments, this will need to be addressed before tagging systems become mainstream within organizations [In contrast, an Resource Description Framework (RDF) document containing the linked tags would require no additional database server].

Managing digital assets

As illustrated earlier, digital assets come in many forms, though research informaticists often presume databases are a primary source, capturing both experimental data and analytical results. However, even data should not be treated solely as authorless input, since there is a premium on it when generated by good experimental design and curated properly by knowledgeable experts. Beyond the creation and sharing of digital assets, there are issues that play a major role in asset use:

- What do I trust and by whom?
- What do I share?
- How do I get recognition and credit?
- How does my work add onto others?

Each of these items needs appropriate mechanisms for being supported and ensured over web-based communities, where we are already aware of the difficulties of knowing where an email has originated from with certainty. This implies that assets, such as wikis, data tables and thesauri should never be simply placed in some web page or network asset space, but should always contain specified links to where they originated from, when, and what conditions are associated with them. The specifications developed by Creative Commons (http://creativecommons.org/licenses/) address many of these needs, and offer a means to making assets more trustable (digital signatures) and contextualized (policy conditions attached to resource URI). These will be essential in transforming enterprise spaces into successful knowledge communities.

TECHNOLOGY

We now provide an overview of essential technologies that we believe contribute in a significant way to the dynamic interactions of groups and their assets. This listing is not exhaustive, but attempts to spotlight more recent technologies that are showing promising social impact.

The World Wide Web

Central to most of the technologies is the World Wide Web, since it defines the protocol for storing and accessing web-based documents, the representation of web pages, and the configuration of web servers that handle the pages using the protocols. The web is open and extensible, and most of the activities and systems defined above are all taking place through the web, including corporate intranets.

The web is not just about data and documents; it's mainly about people putting their assets on the web and being acknowledged for it. Discovering communities and network asset spaces on the web is as important as finding knowledge experts in physical space. The web is changing rapidly so that more things are discoverable and accessible to more people, and in many different representations.

Search engines

As the web has grown, in has become necessary to provide systems that index the location of content and update this regularly as the web changes. Though some of these have impressive capabilities, they all lack the ability to associate people or communities to ideas and knowledge. They serve as a first level web-discovery tool able to index word phrases and map them to locations. However, they always require additional deduction and action by the user to refine the returned ranked list in to a more desired outcome. Nonetheless, those search tools that perform page-ranking [10] take advantage of a rudimentary social phenomenon on the web: the identification and embedding of authoritative and relevant links into one's web pages in an effort to provide visitors with more useful information, in the hope that their pages will also be referenced. This yields a form of social network that can be analyzed by search engines and other agents. This is still an emerging field, and systems that utilize social networks, semantics and content meaning are still years away.

Social network systems

Since most of this technology is in support of social constructs, the ability to represent, manage and display social networks is going to be powerful. Social networks are used to provide information gained through trust-networks, whether one is looking for experts, job positions or just friend-activities networks. Commercial systems, such as Linked-In, exist to capitalize on the demands for this social access, and open systems such as Friend-of-a-Friend (FOAF, based on Semantic Web standards) are also available and are already in wide-scale use by many in the life science community and related industries. In the sciences this translates into who is working on a project, is part of a lab, or is referenced in areas of continuing research.

Technologically, a wiki requires only a simple to-use, freely available piece of software that enables the dynamic editing of pages via any standard web browser (Interestingly, the original tools written for the web by Tim Berners Lee allowed direct editing of HTML pages via a combination browser-editor; wikis are a curious retrograde shift using another intermediate form). Tags are simply string objects, which serve as a mnemonic around some concept such as 'beta-blocker' or 'avian flu' and which can be linked to any web-based resource. All of these systems are either accessible by users with accounts or open to the public. Those with accounts can track modifications and embed provenance information into the web content and tags.

Newsfeeds

Resources and assets can be 'pushed' to users who subscribe to channels oriented around topics of interest. The most prominent form is the *newsfeed*, which are increasingly used in scientific publications such as *Nature* (www.nature.com/rss). Several newsfeed formats exist for, but only RSS1.0 is based on the Semantic Web RDF standard. This by itself is not a social tool, but if different feed-channels can be set up as topics by community members, other members can readily subscribe to them, and a dynamic social network is created.

Semantic Web standards

A new set of recommendations developed by the W3C [11, 12] is being promoted to enable the embedding of semantic relations within and between web documents. It extends the original 'link' form (known as *href*) to incorporate the reason the web link is there in the first place, e.g. is it a 'thing', a supportive reference, a standard definition or some alternative model. It also allows one to define any web resource by types, e.g. *genes*, *diseases*, *etc*. The consequence is that data and propositions (well-structured statements) can now be added explicitly to the make-up of the web, improving clarity, enabling consistency in re-use and enhancing search and cataloging mechanisms.

Semantic relations and links do not need formal standard bodies to become established; they can be readily defined by small groups or even individuals. In this regard they are not too different from tags, but are more powerful since they are URI-valid web entities, and therefore, can be offered to the whole world through the web [2]. Contrary to many opinions [13], the Semantic Web vision for the most part is really about a large, distributed collection of bottom-up semantic definitions, driven autonomously by individual communities. Semantic links also allow the association of authorship to ideas and relations, something tags do not do on their own.

These specifications (based on RDF) have the added affect of allowing structured information (scientific data) to be referenced on the web, and include explicit references to their content type, whether they reside in databases or tabular files. In addition, statements (including beliefs and hypotheses) can be constructed about any combination of facts and published on the web. This is referred to as 'exposing the data', and it allows participants to not only make mention of data and experimental results, but to concisely refer to its content in the context of a supportive argument or a refutation. This would have the effect of reducing spurious arguments not supported by references. The data can also be reused in various combinations with other data in ways that preserve its meaning but bring to surface new insights—this is referred to as Recombinant Data [14], and could dramatically propel the effective use of data in science. For example, a semantic description of how an off-target gene gives rise to a side effect when a new drug is applied can be recombined with the annotated data of that gene as well as the annotated data around the new drug. This bundle as a whole can be annotated as being the proposed hypothesis by this individual researcher, and would be retrievable as such.

More interestingly at the community level, members are able to collectively annotate scientific facts or concepts over the web on the basis of insights, proposed hypotheses, alternative views, or the disproof of a theory. Common vocabularies defined through community-based 'Negotiation of Meaning' can be readily used in the annotated links to make their statements clearer and comparable. Evidence as well can be linked to these statements and made instantly visible to the community. These capabilities together support scientific discourse and promote effective knowledge exchange in a research community.

Once vocabularies and their dependencies have reached an agreed structured level (implied in Wenger's second stage of communities of practice), the elements can be represented as a formal ontology. This does not mean all negotiated vocabulary need become ontologies, but only when their broad use needs to be constrained for reasons of conciseness or regulation (as in clinical practice and drug regulation). By using the W3C's Web Ontology Language (OWL) specification, controlled vocabularies can be offered and utilized throughout the web.

Connected knowledge

By choosing to represent scientific information ('*the BRCA1 gene contributes to incidences of Breast Cancer*'), processes ('*Radiation therapy can reduce Liver Tumor growth*'), and policies ('*Men over 50 should be routinely screened for Prostate Cancer*') using ontologies, many functions of biomedical research communities can be normalized and optimized. We do not go as far to suggest that (all) community knowledge can itself be represented formally; more appropriately it is a means to embed concise, discoverable markers (defined as part of Negotiation of Meaning) into additional knowledge artifacts (published papers and data sets). When accessed and interpreted by community members, this web of assets can become reified as actionable knowledge, which can be guaranteed to be current and authoritative by virtue of its referential nature. Finally, these assets can be offered as a basis (knowledge map) to other communities requiring a subset of their elements for their own pursuits (e.g. human genomics → human inherited diseases).

The key issue here remains: how do these formal structures come about in early stages of the community life cycle? We believe this requires a dynamic discourse in the community that is not constrained up by formal rules, but rather is supported by the ability to track ideas and points-of-view with full knowledge of their provenance (source). The crux of our premise is that the bottom-up model offered by the Semantic Web will go a long way to making community activities be productive.

ENTERPRISE KNOWLEDGE

Our description of communities has up to this point referred mainly to informal groups that form in response to a need by knowledge workers. However, these can easily (and often do) include members from formal organizations, and the communities may even provide knowledge solutions to problems that exist within organizations or corporations. This is why communities of practices are an important part of any company that values knowledge workers, and offers them a venue for improving their knowledge and skills outside of any formal corporate training program.

Communities can form not only outside but also within organizations, having an existence parallel to the organizational structure. The incentive for being part of a communities of practice in a company is that it allows people to both contribute ideas and learn from others in a relatively safe (trusted) environment. Since almost everyone wants to be recognized as offering value to the organization, there will always be a supply of knowledge contributors, who are often the younger workers wishing to make a name for themselves.

An integral requirement for these communities to thrive is the ability to support the association of ideas and solutions to the workers who donated them in a clearly visible and non-corruptible way. In this way the main objective of getting more colleagues to seek out the knowledge source personally is achieved.

CONCLUSION

The world is evolving into a globally connected set of communities, not just around business markets and public services, but also around knowledge communities. The Internet and the web are becoming key spaces for individuals, groups, as well as organizations. These spaces allow members of the academic, industrial and healthcare provider domains to interact with one another and exchange knowledge and information. This is increasingly becoming an integral part of their work style, and provides an essential knowledge development and sharing function that is not provided by the formal organizations.

In many cases, innovations can be shown to originate from such communities. One case is the rapid research and adoption of the very promising 'RNAi technology', where a network of RNAi researchers emerged each working on different species in the 90's (seminal paper Fire *et al*. 1998 [15]), and exchanging RNA sequence information and experimental results. Within a matter of a few years RNAi quickly was applied to therapies and drug screening projects in pharmaceuticals, and with these apparent applications, Mello and Fire received the 2006 Nobel Prize for Medicine— a remarkable short time from discovery to award. Due to the accelerated adoption of theory and technology, global genome-wide screening initiatives have begun, some as alliances between industry and academia [16], and some as public consortia (TRC, [17]), both formal and not informal communities of practices anymore.

Other examples of rapid innovation are seen in communities such as the public genome sequencing initiative, the Gene Ontology Consortium (initiated by geneticists, not computer engineers), the Biomedical Informatics Research Network (BIRN [18]), and the Alliance for Cell Signaling (AFCS [19]). Each of these relies heavily on the instantaneous and dynamic nature of the web to keep its community connected and functioning.

It is well recognized that knowledge communities are an essential part of the scientific, pharmaceutical and medical ecosystem. It is also clear that the web is increasingly becoming the main medium for knowledge networks. Advances and successful innovations are growing in number and complexity as a direct consequence of the knowledge being web discoverable and accessible. Therefore, the more effective the knowledge can be presented and used by communities, the more successful these communities will be. Being visible, having consistency, displaying associations, relevance and dependencies, supporting provenance and maintaining quality are all things the Semantic Web is capable of improving.

Individuals and groups develop knowledge as social beings. Indeed, knowledge itself is a profoundly social 'thing', as Wittgenstein and other philosophers have argued. Any tools that enable the social being of knowledge to develop and flourish is by its nature a great boon to all who believe that knowledge development is a critical activity for all mankind. The Semantic Web has the capability of being just such an enabler. The very act of making meaning more explicit and helping to better understand common meanings will prove, we believe, to be a great technical step forward in the ways we understand and work with knowledge.

Key Points

- Knowledge networks require the convergence of three areas: communities of practice, knowledge assets, and technology.
- Social dynamics and the desire for recognition are the main drivers for generating successful knowledge networks.
- The Semantic Web paradigm will prove its utility by supporting knowledge communities.

References

1. Shannon CE. A mathematical theory of communication. *Bell Sys Technical J* 1948;**27:**379–423, 623–56.
2. Prusak L, Davenport T. *Working Knowledge*. Harvard Business School. Press: Cambridge, MA, 1998.
3. Berners-Lee T, Hendler J, Lassila, O. *The Semantic Web*. Scientific American, May 2001, 29–37.
4. Neumann E. *Finding the Critical Path: Applying the Semantic Web to Drug Discovery and Development*. Drug Discovery World. RJ Communications & Media Ltd, 2005, 25–33.
5. Wenger E. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, 1999, 52–54.
6. Moen W.E. Mapping the interoperability landscape for networked information retrieval. In: *Proceedings of First ACM/IEEE-CS Joint Conference on Digital Libraries*, Roanoke, VA. ACM Press New York, NY, June 24–28, 2001: 50–2.
7. Ashkenas RA, Ulrich D, Prahalad CK, Jick T. *The Boundaryless Organization: Breaking the Chains of Organizational Structure*. Jossey-Bass Publishers, 1996, 111–190.
8. Antonio D. *Descarte's Error*. Vintage Press, 2006; 104–105.
9. Giles J. Internet encyclopaedias go head to head. *Nature* 2005;**438:**900–1.
10. Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine. In: *Proceedings of the seventh international conference on World Wide Web*. Elsevier Science Publishers, 7, 1998; 107–17.
11. http://www.w3.org/RDF/(April 2007, date last accessed).
12. http://www.w3.org/2004/OWL/(April 2007, date last accessed).
13. http://www.shirky.com/writings/semantic_syllogism.html (March 2007, date last accessed).
14. http://www.w3.org/2006/Talks/0404-bioit-em/ (April 2007, date last accessed).
15. Fire A, Xu S, Montgomery MK, *et al*. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 1998;**391:**806–11.
16. http://www.rnaiglobal.org/initiative_pr_0706.pdf (April 2007, date last accessed).
17. http://www.broad.mit.edu/cgi-bin/news/display_news.cgi?id = 421 (April 2007, date last accessed).
18. Biomedical Informatics Research Network. http://www.nbirn.net/ (April 2007, date last accessed).
19. The Alliance for Cellular Signaling. http://www.afcs.org/ (April 2007, date last accessed).