Collection, Evaluation, and Diffusion of Information in Online Communities

by

Evangelos Mamas

Submitted to the System Design and Management Program
in partial fulfillment of the requirements for the degree of

Master of Science in Engineering and Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2005
[February 2006]

Signature of Author .....
Evangelos Mamas
System Design and Management Program
September 7, 2005

Certified by .....
Eric von Hippel
Professor, Sloan School of Management
Thesis Supervisor

Collection, Evaluation, and Diffusion of Information in Online Communities

by

Evangelos Mamas

Submitted to the System Design and Management Program
on September 7, 2005, in partial fulfillment of the
requirements for the degree of
Master of Science in Engineering and Management

Abstract

In this thesis, I study a number of online information communities to understand how open practices are currently used in supporting community functions. By examining how communities operate, I hope to provide individuals and corporations interested in creating such communities with a good starting point. The communities examined, some corporate and some user sponsored, share different types of information and have different intents. I analyzed the communities in terms of their openness in the following three major community functions: information collection, information evaluation, and information diffusion. The benefits and challenges of open practices are discussed and the tradeoffs involved in selecting the most appropriate practice for each major community function are presented.

Thesis Supervisor: Eric von Hippel
Title: Professor, Sloan School of Management

Acknowledgments

I am very grateful to Professor Eric von Hippel not only for guiding me throughout the thesis process but also for his patience and understanding during the past few terms. I would like to express my gratitude to MIT and Sloan for providing an amazing learning opportunity. My experience here has been great and my unwillingness to finish my thesis is a proof of that (or so I like to think!).

I would like to thank my parents and my sister for their continuing support and encouragement during the program and my stay here in Boston. Last but not least, I would like to thank Angela for being there for me throughout the past year, for providing the motivation to continue the work, and for taking the time to proof read my thesis!

Contents

List of Figures

Chapter 1

Introduction

The revolution in information and communication technologies during the past two decades has enabled a global information network known as the Internet. The Internet has drastically changed the boundaries of communities by enabling people from all around the world to interact in ways that were previously impractical if not impossible. One common use of the Internet is to extend the interactions between social groups with preexisting ties such as family, friends and even corporate networks. However, what is more astonishing is the amount of interaction between people without pre-existing ties. In fact, a large number of online communities (communities based on the Internet) with diverse goals have been formed between people with similar interests[3].

In recent years, many corporations have realized the business value that online communities can add and as a result have started building and supporting their own online communities. Chan and Lee report that online user communities have helped corporations leverage the creativity of customers in all phases on new product development[4]. They classify communities into five types based on the characteristics of the members and the interaction levels: virtual customer communities, beta testing volunteer corps, user content collaboration innovation community, user development community, user product collaboration innovation community. While they observed a significant difference between knowledge creation, sharing, and diffusion

in the various community types they also found that user product collaboration innovation communities can help in all phases of new product development. Dahan and Srinivasan describe in their work how virtual reality tools can be used by companies to enable internet-based testing of new product concepts[7]. Satish describes how corporations across industries including Cisco, Bang & Olufsen, Fiat, and Microsoft have established virtual customer communities to improve their new product development processes[17]. Given the potential value customers and lead users can deliver to corporations, it should come as no surprise that corporations have been actively involved in establishing online communities. Wenger's work on Communities of Practice (COP), has helped raise awareness on the business benefits of supporting "groups of people informally bound together by shared expertise and passion for a joint enterprise"[26]. In particular, he claims that COPs can add value to corporations by generating new lines of business, solving problems, spreading best practices, developing people's skills, and helping to recruit talent. With the use of new information technologies, COPs become no different than online information communities sponsored by corporations to help accomplish certain business objectives. Yet another way for corporations to get value from online communities is in the area of market research. According to Kozinets, internet ethnography methods can be applied on internet discussion forums and communities where people discuss the value of various products[16]. This is a non-invasive way to learn how well products are perceived and how they should be redesigned in the future. Individuals themselves also benefit by sharing information in communities. Von Hippel has studied how both online communities such as the Apache Open Source Software and more traditional communities like those in high performance windsurfing can be a great source of innovation[25]. He introduces the notion of "agency cost" which is the cost that manufactures have to pay in order to learn what users really want. As it turns out, users are motivated to freely reveal their innovations when their benefits outweigh their costs. He concludes that when the following three conditions are met user communities engaged in innovation can flourish: i) at least some users have incentives to innovate, ii) at least some users have incentives and means to reveal their innovations, and iii) user

led-diffusion can compete with commercial production and distribution. As more and more user communities equipped with valuable knowledge emerge, corporations need to rethink not only how they develop new products but also how they collaborate with their customers and even competitors. Chesbrough, in his work on Open Innovation, suggests that corporations use external channels to obtain ideas and to market their products[6]. The concept of using open practises in online communities sponsored by corporations and/or users in their effort to create and deliver value to their sponsors has been a major motivation for this work.

The purpose of this work is to explore and understand how online information communities can be supported by open practices. I have identified a number of online information communities that differ in the type of information they share, the nature of their sponsorship and their intent (See Appendix A). In terms of the type of information, I have selected online communities that share information in the form of software toolkits (alphaWorks), technical articles (developerWorks), encyclopedic entries (Wikipedia), news articles (Slashdot), design documents (ThinkCycle), online discussion forums (HPForums), and even crater recognition results from pictures of Mars (Clickworkers). The diversity in the type of information that these communities share is intentional in the hope to identify practices that hold for a broad range of online information communities. In terms of the nature of sponsorship, I make the distinction between communities that are user sponsored and communities that are corporate sponsored. While the specific intent of each of the communities is unique, I choose to differentiate between communities that serve a business objective and communities that solely serve the public good. In this broad sense intent is aligned with sponsorship. In the chapters that follow, I examine all communities in terms of the following three major community functions: Collection, Evaluation, and Diffusion. Information Collection is the function of growing the content of the community by gathering information from its members. This function focuses on the openness of the collection practice in terms of who can contribute to the community and what their incentives are. Information Evaluation is the function of screening and assessing

the quality of the information. More specifically, I consider how open the screening process and the information quality mechanisms are. Information Diffusion is the function of sharing the information with the community and considers the intellectual property concerns and how they relate to the intent of the community.

Chapter 2

Information Collection

The focus of this chapter is on the information collection practices used in online user communities. Online communities are examined with respect to the intent of their sponsors and categorized according to the degree of openness they employ in collecting information from their members. The incentives to contribute are further examined in the context of each community. Finally, a number of practices that support and encourage collection of information are presented to guide community builders in growing their communities.

2.1 Collection Defined

Information collection is the most fundamental function of an online community and is essential to the existence of the community itself. In this work, information collection is defined as the set of practices that enable and support the function of information accumulation in online communities. At the core of this definition is the act of contributing information relevant to the purpose of the community by members of the community, or information producers, with diverse motives. The definition goes beyond the mere act of contributing information, to include the organization of communities to enable members to contribute information, the information producer incentives to share the information with the community, and the practices that encourage information contribution.

2.2 Understanding Communities

In this section, communities are examined with respect to the following three criteria: i) the type of information collected, ii) the intent of the sponsors who established and support the communities, and iii) the degree of openness in terms of information collection.

2.2.1 Central and Peripheral Information

When examining online communities it is frequently possible to make the distinction between central and peripheral information. Central information is any information that is at the heart of the community's intent. On the other hand, peripheral information is information supporting the central information. In Wikipedia for example, the central contribution comes in the form of an "encyclopedic entry" while peripheral information comes in the form of discussion threads around the given entry. In alphaWorks, the central contribution comes in the form of software toolkits, while the peripheral contributions come in the form of questions and comments regarding the use of the toolkits. In Figure 2-1, I summarize the central and peripheral information types for each community examined in this work. It should be noted that the distinction between the central and peripheral information is not always possible. In the case of HPForums both central and peripheral information are in the form of discussion forum messages and cannot be distinguished. In other cases, such as Clickworkers, peripheral information is not collected at all.

A simple way to think about central vs. peripheral information contributions is to consider central as the information that attracts members to the community and peripheral as the information that engages people to the community. In his work on the role of peripheral members in online communities, Zhang concludes that the number of peripheral members in an online community is typically much higher than that of central members[29]. While peripheral members may appear to be free-riding, in reality they may collectively contribute as much as the central members. In some of

Community	Central information	Peripheral information
alphaWorks	Software toolkits	Discussion messages, Ratings of technologies
developerWorks	Technology articles	Discussion messages and feedback on product features
HPForums	Discussion messages	Discussion messages
Clickworkers	Identification of craters	None
Slashdot	News stories	Discussion messages and ratings of news stories
ThinkCycle	Problem statements and design concepts	Discussion messages and feedback on concepts
Wikipedia	Encyclopedic entries	Discussion messages

Figure 2-1: Central and Peripheral Information Types

the communities examined in this work, one can observe that peripheral contributions can increase the perceived value of the central information and can help grow the community. One of the key features of Slashdot for example, is that members of the community get to read not only the news stories but most importantly the reactions of other community members to these news stories.

2.2.2 Community Intent and Sponsorship

The intent of online information communities is directly linked to the type of sponsorship supporting the operation of the community. I choose to make the distinction between user-sponsored and corporate-sponsored communities. The intent of user-sponsored communities is aligned with serving the needs of the community members. On the other hand, the intent of corporate-sponsored communities is to support the community and its members in order to address a business objective.

Traditionally online communities have been formed by enthusiasts with a need to share knowledge on a topic of common interest. The founders of these communities are typically individuals with the know-how required to build and operate the tech-

nology necessary to support the community itself. The focus of these communities can range from providing news stories for technology enthusiasts, as in Slashdot, to sharing software toolkits as in the open source communities, to collaboratively creating free encyclopedia for the masses as in Wikipedia. What is common in all these cases is that these communities have been founded by users to serve users without any explicit expectation of achieving monetary benefits from the community. While the founders and users of these communities enjoy other benefits the communities themselves were founded in the spirit of enabling users to help each other.

It is important to note that as online communities grow in terms of members so do their financial needs. Depending on the community interactions, the cost associated with maintaining online communities will increase as the number of members and the amount of information grows. Some of the user-supported communities, such as Slashdot, have chosen to support their activities by introducing advertising in the content and by offering members the option to purchase subscriptions which are free of advertising. On the other hand, other communities such as Wikipedia have decided to grow by obtaining a not-for-profit status and seeking donation from private sources as well as grants from government sources all over the world. For the purpose of this work, these communities are considered user-sponsored communities because their intent is simply to serve the users of the community.

Corporate-sponsored online communities on the other hand, have been emerging in the recent years as more corporations started realizing their potential and understanding how to create and operate them. Typically, the intent of corporate sponsored communities is aligned with a business objective. As such, the benefits of the community sponsor need to be well understood and the appropriate funding to support the community is allocated by the sponsoring corporation. Examples of corporate-sponsored communities include alphaWorks from IBM and HPForums from Hewlett-Packard. alphaWorks is an online community used by IBM to identify the potential of emerging technologies that are developed within the company. HPFo-

rums is a user support forum that enables HP customers to support each other on using products from HP and in the process enables HP to improve its own products. It should be noted that even though corporate sponsored communities are aligned with business objectives, this does not and should not prevent them from serving the rest of the community members.

In Figure 2-2, I describe the intent of each online information community I examined in this work and I indicate the type of sponsorship for each.

Community	**Community Intent**	**Sponsorship Type**
alphaWorks	Identify the potential of IBM emerging technologies	Corporate
developerWorks	Support and promote IBM products and technologies	Corporate
HPForums	Support HP products	Corporate
Clickworkers	Use volunteers to perform routine scientific analysis	Corporate
Slashdot	Share interesting news stories on a daily basis	User
ThinkCycle	Enable collaborative design to solve community challenges	User
Wikipedia	Create a free and reliable encyclopedia	User

Figure 2-2: Community Intent and Type of Sponsorship

2.2.3 Degree of Openness

The term open community has been widely used in the management literature to encapsulate a number of concepts. Reagle defines an open community as one that demonstrates: 1) open products, transparency, integrity, non-discrimination, and non-interference[15]. For the purpose of information collection this definition is overdetermined. I shall define openness as it relates to the ease of a user to contribute to the community. In this work an open online community is one that allows anyone

to contribute central and peripheral information. On the other hand, the smaller the set of people that can contribute the less open, or more closed, an online community is. In later chapters, I will revisit the degree of openness and examine it in terms of evaluation and diffusion of information.

A community with the highest degree of openness, or an open information collection model, allows any user to contribute information that is both central and peripheral in nature. On the other hand, a community with a closed collection model accepts contributions from a selected group of people. It should be made clear that open collection of information does not necessarily translate to open diffusion of information due to the concerns of information quality and intellectual property ownership. The process of "screening" the information applies to both open and closed communities and will be discussed in depth in the following chapter.

Open communities have the potential to engage a large number of members and can grow very fast. The open nature of the community however increases the needs for sophisticated quality control mechanisms to ensure that the collected information is relevant to the community and that it does not violate community policies and copyrights. The Slashdot community for example, collects news stories from all members and each such story has the potential to become central information by becoming available on the main web pages.

On the other hand, closed communities can be used to simplify management and to control intellectual property ownership. The alphaWorks community, used to identify the emerging technologies of the future and capture economic rents from them, requires that the sponsoring company, IBM in this case, own the intellectual property related to all central contributions, or software toolkits. Closed communities may run into the problem of maintaining a critical mass of interesting information without which any online community is doomed to fail.

2.2.4 Aligning Degree of Openness with Type of Sponsorship

In this section, I examine communities in terms of the type of sponsorship and the degree of openness. As can be seen in Figure 2-3, it is possible for corporations to create and sponsor successful online information communities that use either open or closed collection models. In particular, communities like developerWorks, HPForums, and Clickworkers are used to share knowledge within the community and benefit greatly from allowing everyone to contribute. On the other hand, communities like alphaWorks choose to accept only the content that the sponsoring company creates because that is the only content IBM can generate value from. It should be noted that information collection is closed with respect to the central information, software toolkits in the case of alphaWorks; it remains open for peripheral information like feedback on the toolkits. As expected, all communities examined implement an open collection model for peripheral information because the feedback of the community is always valuable. Looking at the user sponsored communities, one can observe that all the ones examined use an open collection model. However, open communities with a closed collection model for central information can exist if the diffusion of information is also closed. In such a case these communities can be considered private that are supported by the users and are only accessible by the supporting members.

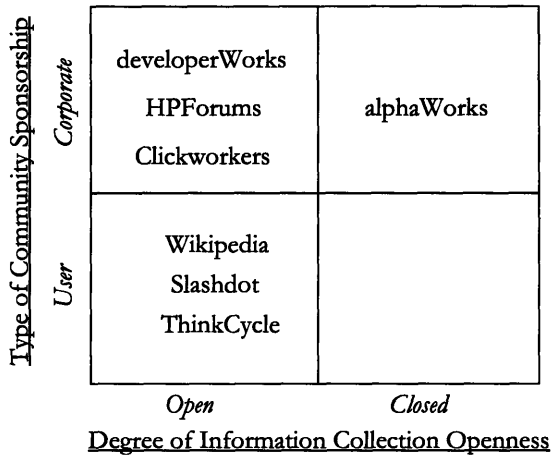

Figure 2-3: Type of Sponsorship vs. Degree of Openness

2.3 Contribution Incentives

In this section we examine what motivates people to contribute information to online communities. The motivations of individuals to contribute to online communities have been covered extensively in management research literature using a variety of underlying theories. Tzouris, in his master thesis on motivation in software communities, examined 36 articles and found that 33.3% used theory of communities, 27.7% used motivational psychology, 13.8% used governance structure, 13.8% used economics, and 11% used the gift economy or other theories[23]. Hemetsberger in her work on virtual communities identified a set of key elements and related concepts and theories to explain exchange processes in virtual communities[9]. They key elements include intrinsic motivation, extrinsic rewards, meaning of exchange, gift-giving, common goals and values, and communal relationships. Butler et al. in their work on community building define four types of benefits to explain what motivates people to participate in online communities[3]. The benefits types can be grouped into personal that benefit the contributor directly and altruistic that benefit others in the community. The personal benefits are further categorized in informational, social, and visibility benefits. Informational benefits include gaining access to otherwise inaccessible information and improving ones abilities by leveraging the information they acquire. Social benefits include the building of social ties that provide friendship and support. Visibility benefits enable people to establish an online reputation that could be eventually linked to more direct economic of professional benefits. Finally, altruistic benefits include the personal satisfaction of contributing to the community, helping others and working towards a common goal.

In this work, I use the four types of benefits introduced by Butler et al. and apply them to each community studied. The results presented in Figure 2-4, examine the benefits of contributors of central information. Overall, it appears that contributors of central information value highly the ability to receive feedback on their contributions and are also motivated by visibility and altruistic benefits. Contributors of peripheral information appear to be more motivated by altruistic and visibility benefits. One

important benefit, not captured by the four types, is the direct economic and task involvement benefits. In the case of direct economic benefits, the developerWorks community provides competitive monetary rewards to authors of published articles. When the information producers have many channels for distributing their work, corporate-sponsored communities can successfully use monetary rewards to collect information. However, using this type of motivation by itself is not recommended. It should be coupled with visibility benefits that motivate contributors to gain high visibility by contributing central information to very popular and highly respected communities. Surprisingly corporate-sponsored communities such as Clickworkers demonstrate that information collection can be motivated by task involvement benefits alone. In this community users are willing to play the "game" of identifying craters in digital images mainly because the activity itself is extremely enjoyable.

Community	Informational Benefits	Social Benefits	Visibility Benefits	Altruistic Benefits
alphaWorks	Receive feedback, identify new uses for technology	None	Reputation, professional advancement	Promote IBM, enable community to innovate
developerWorks	None	None	Global exposure in a popular site, professional advancement	Help others learn
HPForums	Improve knowledge on topic	Socialize with highly skilled members	Build community status, professional advancement	Help others solve problems
Clickworkers	None	None	None	Support NASA research
Slashdot	Receive feedback, learn from experts opinions	Socialize with highly skilled members, define friends and foes	Build community status	Help community to stay current on technology news
ThinkCycle	Receive solutions to challenges and feedback on solutions	Socialize with members with same interests	Reputation, professional advancement	Help people in need with innovative designs
Wikipedia	Receive feedback on the neutrality of your information	Socialize with members	None	Share knowledge, Create a free encyclopedia for everyone

Figure 2-4: Information Contributors Benefits

2.4 Practices that support contributions

A number of practices that support information collection in online communities and encourage members to contribute have been observed in the communities examined. The practices listed below do not serve as a complete reference; rather they are a starting point in helping builders of both user and corporate sponsored online communities ease the task of information collection.

- Simplicity of the collection process is critical to encourage contribution. If the cost associated with contributing information is high then individuals will be less willing to contribute. Most modern information systems such as Wikis, Blogs, and phpBB forums used to support communities, enable members to easily contribute content, and provide them with instant gratification. This is true for all communities examined.

- Modularity of the contributions is another way to encourage contribution. Community interactions should be designed to be highly modular to enable contributors to do as much or as little as they want while enabling them to link their contributions to the rest of the information. Wikipedia, for example, allows users to create entry stubs without much content in the hope that other members will populate them. Wikipedia also allows users to update an entry by making only minor edits if they choose to do so. In open source software development high modularity enables easy division of labor among users and effective coordination of their efforts[22].

- Enjoyment of the contribution itself can be a motivating factor for contributing information to the community as previously described in reference to the Clickworkers community.

All the practices mentioned above, can be used to improve the collection of both central and peripheral information. The practice of peripheral information collection itself is extremely important and should be supported by all communities. Peripheral information collection will enable the majority of the community members, who

are peripheral in nature, to contribute in alternative ways. As previously discussed, collection of peripheral information is highly valued by contributors of central information because this is feedback on their contributions. Peripheral information can be easily organized around central information by using discussion forums, online polls, rating mechanisms, and even file sharing.

Chapter 3

Information Evaluation

This chapter addresses the issue of information evaluation in online user communities. First, I define evaluation as a function comprised of the screening and rating sub-functions. I explain the differences between these two functions and I summarize the screening and rating models used in the examined communities. Each sub-function is further examined in more detail and the common practices for each are described. Finally, tradeoffs in selecting the right screening and evaluation practices are presented and explained.

3.1 Evaluation defined

Evaluation of information in online communities in terms of its quality is a critical function for the success and growth of the community. Neus, in his work on management of information quality in virtual communities, describes the significance of information quality and observes that most research focuses on information concepts suitable for mega-sites with millions of users[18]. However, for smaller scale virtual communities it is common to either ignore, over-control, or rely on unwieldy tools for information quality management. All of these approaches could hamper the success of an online information community because of the potentially high transaction costs incurred by the information consumers. As it turns out, low-quality information requires a lot of human attention and as some have observed a explosion in low-quality

information may convert our "information society" to an "attention deficit society". Herbert Simon, Turing Award winner, captures this phenomenon very elegantly in his quote:

> What information consumes is rather obvious: It consumes the attention of its recipient. Hence a wealth of information creates a poverty of attention.

One of the greatest challenges to implementing any type of quality policy is the difficulty to define quality in a way that every community member can agree with. Even if one was able to do so, it would still not work because the community itself will change its views as it acquires new members over time. Joseph Juran, author of many books on quality management, defines quality as the "fitness for use". Phil Crosby, another popular author on the topic, defines quality as "conformance to requirements". Both of these popular definitions provided by experts on the topic of product quality provide us with no direct answer to the question "what is quality". They merely point us to the "right" direction of what quality may be. Juran's definition requires that we define "fit" and "use". Is there an objective way to determine what information is fit and how it will be used within a community? Crosby's definition implies that we define "conformance" and that we are aware of the "requirements". How can one know if all requirements have been gathered and how can conformance be checked?

3.1.1 Information Screening and Rating

In this work, I will not attempt to provide yet another quality definition. My goal is to understand the practices of online communities in trying to enforce what they interpret as quality. In doing so, I believe that it worth separating the information evaluation function into the two distinct sub-functions of "information screening" and "information rating".

Information screening is the function of filtering out information that is not interesting and relevant to the community. This is a one-time event that typically occurs at the time the information is collected and before it becomes available to everyone

in the community. The purpose of screening is aligned with the intent of each community as I will describe in later sections. The degree of openness of the screening function is defined based on who can participate. If any member of the community can participate then the information screening model is considered open, otherwise it is considered closed.

Information rating is the function of associating a piece of information that is already available in the community with an "index of goodness". I use the term goodness to refer to what each community member perceives as good. Typically information rating is performed on an ongoing basis and the rating of any particular information could change at any given time. Again the degree of openness of the rating function depends on who can participate. If any member gets to participate then the information rating model is considered open, otherwise it is considered closed. In this case however, there is another dimension of openness and it relates to who can see the rating results. Open access to results means that everyone has access to the ratings, while closed results means that only a selected group of people can access the ratings. Furthermore, an additional dimension of the rating model is whether it applies to the information or the contributing member.

Figure 3-1, summarizes the screening and rating models used by the communities I studied. One interesting observation is that all the corporate-sponsored communities use a closed screening model while all the user-sponsored communities use either an open screening model or none at all. The screening and rating models and their relation to the intent of the communities will be examined in more depth in the sections that follow.

3.2 Information Screening in Communities

Screening is used in communities to filter information that does not align with the intent of the community. Screening is extremely important in reducing the transaction costs absorbed by the information consumers. Given that consumers have limited time

Community	Screening Model	Rating Model
alphaWorks	Closed	Open participation Closed use Contribution-oriented
developerWorks	Closed	Open participation Closed use Contribution-oriented
HPForums	Closed	Open participation Open use Contribution/user oriented
Clickworkers	Closed	-
Slashdot	Open	Open participation Open use Contribution/user oriented
ThinkCycle	-	-
Wikipedia	-	Open participation Open use Contribution-oriented

Figure 3-1: Information Screening and Rating

and attention to filter information themselves, screening at the community level can help the community maintain a healthy amount of relevant information at a minimal cost its members. A community able to effectively screen incoming information will most likely grow rapidly. However, the risk of too much screening or misalignment with the intent of the community can result in shrinking communities. Given that it is only natural for the community interests to evolve over time, the information screening practices need to be adjusted accordingly to allow for evolution of the community. As Neus observes, over-controlling of quality can result to tunnel vision[18].

3.2.1 Information Screening Practices

The screening practices of the information communities under examination are presented in Figure 3-2. Most of the communities rely on a peer review practice; Clickworkers however is the only exception since it relies on a computer based screening algorithm. The case of Clickworkers is a special one because the nature of the information contributions is such that screening can be performed by using a computer algorithm; this means that quality can be objectively evaluated. More specifically, the contribution is made through a computer program which requires users to identify a crater by clicking on four points that are on the perimeter of the circle. The fourth point, while redundant in identifying a circle, is used to screen how accurate the user is and instantly rejects badly identified circles. For the remaining communities that implement screening we need to consider their peer review practices in light of their intent. A closer look at Figure 3-2 reveals that corporate-sponsored communities rely on a closed peer review practice while user-sponsored communities rely on an open peer review practice.

In the case of alphaWorks, the intent is to assess the potential of the emerging technologies. This can only be accomplished if the technologies available in the community are screened to meet the criteria of emerging, unique, well documented, and easy to use. IBM employees are appointed the task of screening the submissions to meet these criteria. In doing so, alphaWorks maintains its brand as one of the top emerging technologies sites in the software industry. developerWorks promotes and supports IBM technologies and partially relies on unique, well-written, easy to understand articles that can explain to the community how "easy and reliable" the technologies and products are. IBM, the sponsoring company, needs to screen submissions to meet the expected quality standards of the site and also to ensure that publications are unique and closely related to topics that promote the technologies and products that benefit IBM. HPForums supports product users by providing them with a discussion forum to post problems and solutions to HP related issues. The

screening that takes place here is necessary to ensure that discussions are always related to HP products and to filter the use of inappropriate language. Overall, the screening practice of these corporate-sponsored communities is closely aligned with the intent of sponsors and less with the expectations of the users. Slashdot on the other hand, employs an open screening policy by enabling trusted members of the community to screen the information on behalf of the community. This open peer review model requires a trust model to be in place that if well implemented can be aligned with the intent of the community users and sponsors.

Community	Screening Model	Screening Practice	Community Intent	Sponsorship Type
alphaWorks	Closed	Peer review	Identify the potential of IBM emerging technologies	Corporate
developerWorks	Closed	Peer review	Support and promote IBM products and technologies	Corporate
HPForums	Closed	Peer review	Support HP products	Corporate
Clickworkers	Closed	Computer algorithm	Use volunteers to perform routine scientific analysis	Corporate
Slashdot	Open	Peer review	Share interesting news stories on a daily basis	User
ThinkCycle	-	-	Enable collaborative design to solve community challenges	User
Wikipedia	-	-	Create a free and reliable encyclopedia	User

Figure 3-2: Information Screening Practices

3.2.2 Screening Practices Tradeoffs Considered

Deciding between an open vs. closed screening model requires one to consider the issues of enforceability and people selection. With respect to enforceability it is important to understand that the intent of the sponsor is more likely to be enforced by an individual that does so as part of their paid job. On the other hand, it is expected of people who serve the community to screen information based solely on the appropriateness for the community. This increases the risk of exposing company-sponsored communities to bad publicity. If one decides to use an open model, then a trust mechanism needs to be developed to enable any community member to screen

information. In Slashdot, this is implemented by a hierarchy of moderators that screen news stories submission and route them to the top of the hierarchy for the final decision. Moderators are assigned "karma" based on how well their contributions have been received by the community. The trust model itself, introduces small but measurable transaction costs to the community because it relies on constant member feedback. Another tradeoff worth considering, is that closed models do not scale well with community growth, while open models can share the work using a "divide and conquer" approach. The pace of the community is also important in deciding which model to use. Slashdot and Wikipedia are fast paced communities and can be screening information around the clock. developerWorks on the other hand, relies on a closed peer review practice which typically translates to slower screening speeds.

3.3 Information Rating in Communities

Information Rating can help rate the information that is already available in the community. Information rating is a collaborative approach to identifying how good the information is to the eyes of the community members as a whole. The cost of rating information is spread to all the community members that choose to participate and is typically very low because the rating activity is usually simple and quick. The benefit of rated information is that when the rating results are made openly available they assist community member in consuming better information with lower transaction costs. However, excessive information rating may also lead to tunnel vision because of strong feedback loops: community members will tend to consume and rate only highly rated information.

3.4 Information Rating Practices

The rating practices of the information communities under examination are presented in Figure 3-3. Rating practices are, and should always be, optional in order to keep transaction costs low. In its simplest form, rating enables information consumers to

specify how good the information is and to provide textual feedback on the information. This approach is easy to implement and scales very well relative to the size of the community and the amount of information collected. A more complex rating approach is to use algorithms that rate the users and/or the contributions based on page views, number of downloads, and amount of activity in terms of peripheral information. alphaWorks uses a number of factors including the number of downloads and the number of licensing requests to determine which technologies are worth incorporating into commercial products. An even more surprising, yet very effective, rating practice is the lack of a rating function altogether! Wikipedia does not maintain a rating for each entry. Instead, it eliminates all transactions costs related to collaboration by enabling anyone to directly update the entries with minimal effort[18]. This approach has been criticized for being so open that information quality may suffer as a result of many intentional and unintentional edits of bad quality. However, a recent study reveals that vandalism and inaccuracies in popular pages are typically corrected within two minutes[24]. However, for the practice of Wikipedia to work, a very active and community is required. Not surprisingly, the practice of open rating is used in all the communities we examined (that use a rating policy). After all quality is in the eyes of the beholder and limiting feedback to only a chosen few would be artificially biasing the community.

3.4.1 Rating Practices Tradeoffs Considered

A number of tradeoffs need to be considered when selecting the appropriate rating practice. The first tradeoff comes in deciding between open and closed rating-result models. Communities like alphaWorks and developerWorks choose to collect ratings from all members but do not reveal them. This helps them identify the good content and make decisions that benefit IBM the most. By keeping the rating results closed, they avoid any of the community information from getting tagged as "not good enough". This prevents community members from getting biased towards highly rated information and enables IBM to focus its effort on improving or replacing the less popular information. The downside of using an open rating policy, is that

Community	Rating Model	Rating Practice	Community Intent	Sponsorship Type
alphaWorks	Open participation Closed use Contribution-oriented	Rating technologies	Identify the potential of IBM emerging technologies	Corporate
developerWorks	Open participation Closed use Contribution-oriented	Rating articles	Support and promote IBM products and technologies	Corporate
HPForums	Open participation Open use Contribution/user oriented	Rating responses that link to user status points	Support HP products	Corporate
Clickworkers	-	-	Use volunteers to perform routine scientific analysis	Corporate
Slashdot	Open participation Open use Contribution/user oriented	Rating responses that link to user status points	Share interesting news stories on a daily basis	User
ThinkCycle	-	-	Enable collaborative design to solve community challenges	User
Wikipedia	Open participation Open use Contribution-oriented	Directly overwriting information	Create a free and reliable encyclopedia	User

Figure 3-3: Information Rating Practices

community members are less likely to participate in the feedback rating if they realize that they do not get access to the rating results.

The second tradeoff deals with deciding among rating the information providers, the information, or both. HPForums enables members to rate other members' responses to their questions. It is possible to flag answers as "magical answers" in case they solved the problem at hand. The rating of each response, by the author of the question, is used to calculate the status of each member. This approach helps the community maintain a memory of member contributions and their relative quality. Slashdot uses a similar approach in assigning karma points to members based on the popularity of their contributions. Communities that focus on rating information provide an equal opportunity and reward equally anyone who contributes good information, while communities that focus on rating users favor long term dedicated

members. Extensive use of user ratings may make it extremely hard for new members to gain sufficient status to draw attention to their contributions. Finally, user ratings may lead to community biases towards people and away from information; this may reduce the overall information quality of the community.

The final tradeoff is centered on the means of collecting ratings from the community members. One approach is to directly ask members for their ratings. An alternative approach is to observe the member usage patterns and try to translate them into ratings. The direct approach is likely to yield more accurate results because it is optional; only truly interested members will be inclined to provide feedback. The indirect approach relies on tracking page views, download activities, discussion contribution patterns, and text mining approaches to extrapolate quality metrics on the related information. Some of these metrics can provide a good sense of the popularity of the information. However, the results must be carefully interpreted to determine the true underlying causes. Another potential drawback of this method is that community members may stop participating due to privacy concerns.

Chapter 4

Information Diffusion

The focus of this chapter is on the information diffusion practices used in online communities. The benefits a corporation can enjoy from adopting an open diffusion policy are presented and the challenges in managing costs, reaching critical mass, and selecting the appropriate intellectual property rights practice are discussed. Finally, the life cycle of the information shared by the community and how it impacts the diffusion process is explained.

4.1 Diffusion Defined

Information Diffusion is the last of the three fundamental community functions addressed in this work. I use information diffusion to refer to the set of practices that enable and support the dissemination of information to both community members and the broader Internet audience. The definition encompasses the degree of openness that communities employ to allow members access to the information and make further use of it when applicable. The practices used by communities are affected by the life cycle of the information and should always be aligned with the community intent.

4.2 Benefits of Information Diffusion

The proliferation of the Internet has not only diminished the cost of information collection in communities, but has also affected the cost of information diffusion. In the previous chapter, I discussed how the traditional mechanisms for managing information quality, namely high distributions costs, are no longer in place. This has led to an explosion in the amount of information distributed using the Internet and is contributing to what Neus refereed to as the "attention deficit society"[18]. So why should communities continue to offer their information to the Internet and how can they ensure that their information is truly diffused?

All of the communities I examined have adopted an open access policy. This means that anyone can easily access both the central and peripheral information in the community. The benefits of information diffusion using online communities are numerous. Corporate-sponsored communities, like alphaWorks can enjoy both direct and indirect benefits by spreading their technologies and growing the audience of future product customers. developerWorks and HPForums enjoy indirect benefits by supporting product users address their concerns. All of the above mentioned communities are able to learn from the community members how to build better products. Furthermore, these corporations are well positioned to capture this knowledge, incorporate it into future products, and capture the economic rents in the process. The Clickworkers community benefits directly from getting tasks accomplished at a much lower cost. User-sponsored communities benefit from diffusion because they rely on the users who consume information to perform some of the activities of the community such as contributing and evaluating existing information. Wikipedia for example depends entirely on the community to make new contributions and to update and correct existing ones. Slashdot relies on its large community of readers who submit daily news stories few of which eventually make it to the daily digest. ThinkCycle, relies on open access to encourage members to contribute solutions to the challenges posted from other community members.

4.3 Challenges of Information Diffusion

Open access to information is a necessity for communities that rely on a massive participation of members. However, open access provides no guarantee of community success. A plethora of other challenges stand in the way of communities that want to thrive and survive. The first challenge is the cost of maintaining the community. The cost includes the development of the software infrastructure to support community operation and the ongoing communication costs of hosting the site on a server connected to the Internet. In addition, the cost of maintaining the community may include salaries paid to people who devote their time to the community. In corporate-sponsored communities, a number of people are typically employed to maintain the site; their salaries comprise a big component of the cost structure. While corporate-sponsored communities appear to have plenty financial backing they are constantly under pressure to quantify the benefits of operating the community. User sponsored communities on the other hand, are more likely to rely on voluntary contributions both in time and money from members and government sources. Other approaches, like that of Slashdot, include the use of paid advertisements to subsidize the cost of operating the community. It should be made clear that using advertising increases the transaction costs of information consumers who may eventually decide to stop participating in the community. This happens because advertising can be intrusive by nature and will typically require information consumers to spend more time to access the same amount of information.

The second challenge in information diffusion is that of reaching a critical mass of information that can sustain the community. The theory of network effects suggests that a community with more available information is likely to attract more members. At the same time a community with more members is likely to collect more and better information. This feedback loop can help communities grow very fast but will also translate to stagnation for communities unable to reach critical mass of information. Communities like alphaWorks, that use a closed collection model in conjunction with an open diffusion model are under constant pressure to maintain a healthy number

of members by ensuring that relevant information is always avaialbe. Xerox, a highly innovative company, created and sponsored a community called Alphaavenue. Similar to IBM's alphaWorks, Alphaavenue showcases emerging software technologies in the hope of identifying the ones that would help Xerox capture value by offering better products. However, the Alphavenue community was finally shut down after being unable to maintain a wide offering of popular toolkits. It is important to note that a corporate-sponsored community unable to reach critical mass, is not only costly in terms of development, infrastructure, and salaries; it can become very costly if one considers the brand dilution due to negative publicity. In the case of Xerox, the reputation of an innovative company was being tainted by an unsuccessful on-line community. Another community that has not been able to reach critical mass is ThinkCycle. Limited activity in terms of challenges and solutions posted suggests that ThinkCycle has not been able to attract enough active community members. One of the reasons for this may be that the site has not been kept up to date: the homepage for example advertises a call for papers for a conference that took place 3 years ago!

Yet another significant challenge for information diffusion is that of intellectual property rights. Open access to information is typically not synonymous with open use of information. Open use would imply that the information consumers are not limited in any way to use the information. The recent growth of open source software communities demonstrates that members are likely to contribute if they know that the whole community can enjoy the results of the collaborative efforts of the members. Deciding on the appropriate intellectual property rights approach becomes very critical; selecting the wrong approach may deter community participation. In Figure 4-1, I summarize the practices on intellectual property rights used by each of the communities. One can easily observe that corporate-sponsored communities are more likely to retain or acquire the copyright of the information they distribute. This is mainly done to enable the sponsors to reuse the information to their benefit. User-sponsored communities on the other hand, choose to either leave the rights with

the contributor, or to adopt a more "copyleft" practice like Wikipedia with the GNU Free Documentation License. This license is designed to guarantee attribution to the author and to ensure that any redistribution of the information remains free and under the same license terms.

Community	Intellectual Property Practice	Sponsorship Type
alphaWorks	Software is licensed under IBM or open source licenses	Corporate
developerWorks	IBM pays the author for the perpetual right to use the article. IBM also gets 30 days exclusivity	Corporate
HPForums	HP owns the copyright	Corporate
Clickworkers	None specified	Corporate
Slashdot	Trademarks and copyrights are owned by their respective owners. Comments are owned by the poster	User
ThinkCycle	None specified	User
Wikipedia	GNU Free Documentation License	User

Figure 4-1: Intellectual Property Right Practices

4.4 Information Life Cycle

One important factor in planning for information diffusion is to understand the life cycle of the information. By information life cycle I refer to the following two attributes: i) the expected life of the information and ii) the frequency of changes to the information during its life cycle. Information does not necessarily loose its value but it may become less relevant with respect to the evolving interests of the community. The expected life of information on alphaWorks, depending on the popularity of the technology, will typically be no more than a couple of years. Emerging technologies after all cannot be classified as emerging for a very long time! The approach taken by alphaWorks, is to "retire" the technologies and replace them with new emerging technologies. In general the expected life of the information in a community should

be aligned with the community's intent. In developerWorks, articles are expected to be accessed during the time frame that the technologies and products they address are current. In Slashdot, news stories are archived forever but they are mainly diffused and commented on during the first few days since their posting. The frequency of changes is also important in understanding how the community works and is also an indicator of the expected life of the information. Frequent changes/updates in the central and peripheral information suggest that enough members are interested in it and are motivated to spend time on updating it. Wikipedia is the extreme case of frequent changes and has an infinite expected life for its entries; a constant revision practice leads to entries that are always up to date. In other communities like alphaWorks and ThinkCycle one is able to detect how relevant the information is by observing the frequency of updates from the original contributors and from the peripheral members. In communities with long information life cycles and infrequent changes, it becomes important for the community to identify the right time to bring newer information center stage.

Chapter 5

Conclusions

The major finding of this work is that corporate and user sponsored online information communities are able to grow, and become successful by using a combination of open and closed collection, evaluation, and diffusion practices.

In the area of information collection, it is important that one identifies the central and peripheral information in the community. Collection of central information can use either open or closed practices depending on intellectual property rights and how the information aligns with the intent of the community sponsor. Collection of peripheral information should always be open; many approaches for submitting peripheral information should be made available. Community members can be motivated to contribute information for a variety of reasons; the community sponsors need to understand this and use it to their advantage. Finally, in designing the community interactions it is important to keep in mind that simplicity, modularity, and enjoyment can be of great help in engaging community members.

In terms of information evaluation the following two functions need to be made clear: information screening is a static approach of evaluating the information, while information rating is a dynamic and collaborative approach of determining how good the information is to the eyes of the community that consumes it. Both screening and rating can help reduce the cost of consuming information. Screening can be

done using either an open or closed model but most corporations will favor the closed model because it allows for better alignment with their intent. Rating should be open for participation and can rely on a combination of ratings of users, information, or both. If user ratings are used, mechanisms to prevent favoritism should be put in place to encourage new members to participate. Ratings can also be collected in an automated fashion by monitoring member's interactions; such an approach should be used with caution and the results should be properly interpreted.

In the area of information diffusion it appears that all communities are open in providing access to information. However, communities have different approaches in allowing information consumers to freely use the information. Community sponsors are able to use direct and indirect ways to enjoy the benefits of information diffusion. The challenges include the costs of operating and supporting the community, the ability to reach a critical mass of information and participation, and ownership of intellectual property rights. Closed diffusion models may address the ownership of rights but may also deter members from participating in the community. Finally, the life cycle of the information needs to be understood and the community interactions need to be properly designed to support it.

Appendix A

Online Communities

Each community has its own unique goals, ways to incentivize participants, and systems to support the operation of the community. The following sections provide a brief description of each community that was evaluated during this work. This chapter aims to help the reader become familiar with each community and understand how it relates to this work.

A.1 alphaWorks *"Emerging Technologies"*

alphaWorks[11] is a an online community created by IBM to expose emerging software technologies to anyone outside IBM. The objective of the community is to identify the potential of various technologies that have been developed within IBM. By exposing a number of selected technologies to outsiders and monitoring their reactions IBM is able to better determine how to further develop these technologies[5]. Initially established in 1996[8], alphaWorks, has helped IBM create a community of early technology adopters that are eager to use, test, and even license the emerging technologies. During the years some of these technologies have been incorporated into IBM products while other have been released as open source projects to the community.

The main interaction in this community is the sharing technologies in the form of a software development kit (SDK). At any given time approximately 200 technolo-

gies are available for downloading. The technologies are grouped in broader areas of interest such as Web Services, Data Management and Java technology to name a few. Around each technology a small community is formed via the use of a discussion forum. The forum enables the developers of the technology to get in touch with the early adopters and to further improve it.

When it comes to collecting technologies the alphaWorks community follows a closed model; only IBM employees can submit their technologies to the community. The IBM developers that contribute technologies to alphaWorks are motivated by the recognition they may receive if their tool becomes popular with early adopters. It is also a great channel to receive creative feedback on where to apply the technology and how to improve it.

The evaluation of the technology is done at two levels. First, submissions are reviewed by the alphaWorks team to ensure that they are indeed emerging and that they have a high potential of adoption. One of the goals of the alphaWorks team is to maintain a healthy amount of community visitors. This is accomplished by screening the submissions for originality which helps alphaWorks establish a highly regarded image in the community. Second, once the technologies become available on the site, the community input is used to further screen them. The number of downloads, the activity on the discussion forum, and the adoption of the technology in the wider software community help are metrics used by IBM to assess the potential of each technology. The life cycle of most technologies on the site is limited and is heavily dependent on the popularity of the technology with early adopters. Popular technologies will eventually "graduate" and become part of a product or open source project. Less popular technologies will be "retired" and lose their placement on the site to new ones. This helps alphaWorks to maintain a "fresh" collection of emerging technologies and continue to attract early adopters.

The diffusion of the technologies follows an open model since anyone can down-

load and evaluate them from the website. Depending on the particular technology IBM offers various licensing schemes that enable outsiders to develop new products incorporating the alphaWorks technologies. In such cases, given that alphaWorks is trying to encourage the adoption of these technologies, licenses typically cost less than those of IBM established products and are less restrictive.

A.2 developerWorks *"IBM's resource for developers"*

developerWorks[12] is a technical resource site created by IBM to support software developers that use both IBM and open source products and technologies. By collecting and publishing news, how-to articles, tutorials, and source code examples IBM supports and promotes a number of products and technologies that are important for the company. developerWorks, established in 1999, has won numerous awards including "Best Technical Resource", "Best DevelopersŃews Source", and "Best Website and Developer Network".

The most prominent interaction in the developerWorks website is the sharing of articles on products and technologies. All the content is organized under two broad categories: products resources and technology resources. Every article on the site contains links to products and resources that developers can download to further enhance their learning experience. Furthermore, a discussion forum exists for each article to enable the authors to openly communicate with their readers.

The collection of articles is open and anyone is free to submit an article. The open submission policy helps IBM include users of the technologies and products as authors and gives them a sense of belonging to the developerWorks community. It is not uncommon for users of the various products and technologies to be the best authors of how-to articles; even better than the people who developed the technologies to begin with. The incentives for contributing to developerWorks include monetary rewards at standard published rates, and also provide an opportunity to experts to

share their knowledge and claim some community recognition.

The article submissions at developerWorks are evaluated by a team of editors. Typically a single editor is assigned to each product/technology. The editors maintain a wish list for the content they are interested in including in their respective sections[13]. Anyone interested in submitting an article could quickly get feedback on how suitable it is for developerWorks by simply submitting the article idea. In reviewing articles, editors try to identify original content that is well written and will help developers in understanding and using the various products and technologies.

The developerWorks site is open in diffusing the articles since anyone on the Internet can access them. Authors retain full ownership to their article but grant to developerWorks 30 days' exclusivity, as well as a perpetual license to publish the work[14]. More specifically:

> You will retain full ownership of the article. You hereby grant IBM a perpetual, irrevocable, worldwide, paid up, license to use, perform, publish, reproduce and prepare derivative works of the article in any form, including but not limited to print, CDs and distribution on the Internet, and to allow others to do so. This license is exclusive for the first 30 days from our initial publication, and non exclusive thereafter, provided that any publication of the work by the author or any other licensee of the author must include the notice "First published by IBM developerWorks at http://www.ibm.com/developerWorks/."

A.3 HP Support Forums *"IT resource center forums"*

The HP Support Forums is an online community for peer-to-peer technical support and knowledge sharing[10]. The discussion forums are hosted by HP in an attempt to bring the technical community together to learn from each other on issues related to HP products. In the process, HP is able to learn more about its own products and

use the input to solve existing problem and add new features.

The main form of interaction in this community is a discussion message. The forums are organized in "areas of peer problem solving" representing various groups of HP products such as: "Digital Imaging", "HP-UX", and " "Printers" to name a few. To contribute a message to a discussion registration is necessary. Despite the need for registration the collection model is consider open since anyone with a valid email address can register for free. The overall motivation to participate in sharing knowledge in this community is in the spirit of helping fellow HP product users who are faced with similar issues. In addition, users who post questions are obviously motivated by the desire to solve their own problems. On the other, hand users who post solutions are also motivated to elevate their status within the community by accumulating member points.

The evaluation of the submissions is done using an open model. The authors of questions have the right to grade the various replies using a scale from 1 to 10. The more points assigned to a response the more useful the response was considered by the author of the question. The points awarded to each response become member points for the response author and are used to rank members in the community. Members can reach various status levels within the community based on the points they have collected over time. The status of each member is visually represented using a graphic icon next to their responses. This way a reader of the forums can easily "weigh" the importance of a response based on the author status. In addition to the member points scheme, a number of HP employed moderators are responsible for evaluating the submissions. Their role focuses on identifying content that is offensive, uses vulgar language, contains copyrighted or personal information, or includes advertising.

Diffusion of the support information in the forums is open. Anyone can visit the support forums to read and search the various discussion threads without the need to register. This is in line with the objective of HP which is to provide product support

to their customers.

A.4 Mars Clickworkers

Mars Clickworkers[1] is a NASA experiment in the area of volunteer science. In this experiment, volunteers can visit the Clickworkers site and volunteer to identify craters in digital images of the surface of the plant Mars. The objective of the experiment is to use volunteers to perform routine scientific analysis on large amounts of digital images. By using volunteers for simple scientific analysis tasks, NASA is able to get extremely accurate results[2] practically for free.

The main type of interaction in this community of volunteers is a simple image analysis task. Volunteers contribute by either identifying or classifying craters in images. Training is available to help volunteers understand how to use the system. The provided interface is very simple to use and volunteers are free to participate for as little or as much as they want. The motives for participation include the desire to help the scientific community by donating time and skills and also the personal gratification of "playing the Clickworkers game". The results are compared against known catalogs of craters that have been compiled by scientists. One approach to improve the quality is to cross validate the work of the various clickworkers. The results are made available on the site for viewing but are mainly relevant to the scientific community that studies craters on Mars. In this respect, the diffusion of information is following an open model, since the results are made available to everyone - but of little direct value to the majority of the volunteers themselves.

A.5 Slashdot *"News for Nerds. Stuff that matters"*

SlashDot is an online community created by Rob Malda in 1997 as a way for technology enthusiasts to share interesting news stories that they come across on the

Internet[19]. The objective of the site is to create a daily digest (called the "The Omelette") of the most interesting technology related news stories by enabling readers to submit short abstracts and links to news stories. Since its inception Slashdot has grown to a community of over 890,000 registered users. The popularity of the community and the great number of its readers have caused numerous internet sites to tmeporarily close due to large volumes of internet traffic soon after a news story was posted on Slashdot. The phenomenon in known as the "Slashdot Effect" and has been observed in numerous ocassions and studied in depth[27].

The main interaction in the community is the submissions of a news story. Any reader can submit a story to Slahsdot to be shared with their fellow Slashdotters. A complex moderation scheme is in place that enables a group of editors and any users with enough moderation points (aka karma) to decide which story gets published. A discussion thread linked to each news story enables readers to provide their comments and to rate the stories. Readers can also rate comments by assigning karma points to them. Since almost anyone can become a moderator in Slashdot, a meta-moderation scheme is used to remove bad moderators and reward good moderators. In terms of diffusion of information, Slashdot follows an open model and enables anyone to read both the stories and the related discussions. One feature that facilitates diffusion of stories and comments is the filters that the site implements enabling any user to filter comments from users that are below a certain threshold.

A.6 ThinkCycle *"Open Collaborative Design"*

ThinkCycle[20] is a shared online space supporting a collaborative open source design initiative. The initiative began in March 2001 by a number of graduate students at the MIT Media Lab who wanted to utilize the creative thinkcycles of students and domain experts around the world. The mission is to create an environment where design challenges are posted and communities from all over the world collaborate to design solutions for the challenges.

The main contributions to ThinkCycle come in the form of Challenges and Concepts[21]. Challenges are well-posed problem statements describing the constraints and issues in design for specific problems. Concepts are design approaches to solving the challenges. All contributions are organized in topics and categories. In addition, discussion forums are provided to support the interaction of the community members.

Collection of contributions follows an open model and anyone interested can post challenges, concepts and comments of the various discussion forums. Topic creators are by default editors and can edit, move, and delete contributions as necessary. Active members with expertise in certain areas may request to become topic Editors. However, aside from peer-reviews of the various contributions there is no formal moderations mechanism in place. Diffusion of concepts is mainly open but the access level to the various design files can be controlled by the topic editors. The intellectual property rights that arise from collaborative design projects have not been addressed by ThinkCycle. The concept creators and contributors have control over this issue.

A.7 Wikipedia *"The Free Encyclopedia"*

Wikipedia is a Web-based encyclopedia developed in a collaborative fashion by volunteers from all over the world[28]. The goal of the Wikipedia project is a very ambitious one and it is likely to take many years to be accomplished:

> Wikipedia's goal is to create a free, reliable encyclopedia - indeed, the largest encyclopedia in history, in terms of both breadth and depth.

Wikipedia consists of 195 language editions and currently contains over 1.5 million articles (most of them in the English language).

The main interaction in the community is in the form of creating and editing encyclopedic entries. For each entry in the encyclopedia a dedicated discussion forum

provides editors with a way to communicate, debate, and reach an agreement about the contents of the entry. Entries are organized in categories and readers have access to both alphabetical and hierarchical indices.

An open contribution model ensures that any reader can easily edit the contents of any entry without the need to have special access rights. As a result all entries are under constant revision by any of its members who collectively try to enforce a "neutral point of view" policy. The open nature of Wikipedia leads to rapid editing and growth but also to inaccuracy and vandalism. The quality of the contributions is typically managed by the sheer volume of edits and as a result any inaccuracies will be quickly edited by the readers. However, readers with administrative privileges have the power to protect and delete pages. Getting the administrator status requires a history of contributions to the Wikipedia community and requires a nomination and a healthy percent (80% or more) of supporting votes from the community. Diffusion of information is open and all content is made available under the GNU Free Documentation License (GFDL). This license permits the redistribution, creation of derivative works, and commercial use of the content provided that 1) its authors are attributed and 2) the content remains available under the GFDL license. Currently, all Wikipedia text is available under this license. However, a significant percent of images and sounds are non-free.

Bibliography

[1] Kafensky Bob. Mars clickworkers faq. Online: http:// clickworkers.arc.nasa.gov/ faq.html. Accessed: June 20, 2005.

[2] Kafensky Bob. Mars clickworkers results. Online: http:// clickworkers.arc.nasa.gov/ results. Accessed: June 20, 2005.

[3] Brian Butler. Community effort in online groups: Who does the work and why. Accessed: June 20, 2005, 2002.

[4] Tzu-Ying Chan. A comparative study of online user communities involvement in product innovation and development. Accessed: June 20, 2005.

[5] Henry Chesbrough. sometimes success begins at failure. *Harvard Business Review, Working Knowledge*, 2003.

[6] Henry Chesbrough. Managing open innovation: in uncertain markets, new metrics can help companies to play poker as well as chess. 2004.

[7] Ely Dahan. The predictive power of internet-based product concept testing using visual depiction and animation. *Journal of Product Innovation Management*, mar 2000.

[8] Moltzen Edward. Ibm unveils alphaworks web site. *Computer Reseller News*, 701:37, 1996.

[9] Andrea Hemetsberger. Fostering cooperation on the internet: social exchange processes in innovative virtual consumer communities. 2001.

[10] Hewlett-Packard. hp support forums - overview and faqs. Online: http://www-128.ibm.com/ developerworks/ aboutdw/ index.html. Accessed: June 20, 2005.

[11] IBM. About alphaworks. Online: http://www.alphaworks.ibm.com/about. Accessed: June 20, 2005.

[12] IBM. About developerworks. Online: http://www-128.ibm.com/ developerworks/ aboutdw/index.html. Accessed: June 20, 2005.

[13] IBM. Editorsćontent wish list. Online: http://www-106.ibm.com/ developerworks/ wishlist/. Accessed: June 20, 2005.

[14] IBM. Power architecture author faq. Online: http://www-128.ibm.com/ developerworks/ library/ pa-authorfaq.html. Accessed: June 20, 2005.

[15] Joseph M. Reagle Jr. Socialization in open technical communities. Accessed: June 20, 2005.

[16] Robert V. Kozinets. The field behind the screen: Using netnography for marketing research in online communities. feb 2002.

[17] Satish Nambisan. Designing virtual customer environments for new product development: Toward a theory. *Academy of Management Review*, jul 2002.

[18] Andreas Neus. Managing information quality in virtual communities of practice. *IQ 2001: The 6th International Conference on Information Quality at MIT.* Accessed: June 20, 2005.

[19] Slashdot. About this site. Slashdot, Online: http://slashdot.org/ about.shtml. Accessed: June 20, 2005.

[20] ThinkCylce. About thincycle. ThinkCycle, Online: http://www.thinkcycle.org /about. Accessed: June 20, 2005.

[21] ThinkCylce. Post new contribution. ThinkCycle, Online: http://www.thinkcycle.org /tc-notes /newt. Accessed: June 20, 2005.

[22] I. Tuomi. *Networks of Innovation: Change and Meaning in the Age of the Internet.* Oxford University Press, 2002.

[23] Menelaos Tzouris. Software freedom, open software and the participant's motivation - a multidisciplinary study. Master's thesis, London School of Economics and Political Science, 2002.

[24] Fernanda Viegas. Studying cooperation and conflict between authors with history flow visualizations. *Proceedings of the SIGCHI conference on Human factors in computing systems*, apr 2004.

[25] Eric von Hippel. Innovation by user communities: Learning from open-source software. *MIT Sloan Management Review*, 2001.

[26] Etienne C. Wenger. Communities of practice: The organizational frontier. *Harvard Business Review*, jan 2000.

[27] Wikipedia. Slashdot effect. Wikipedia, Online: http://en.wikipedia.org/wiki/Slashdot_effect. Accessed: June 20, 2005.

[28] Wikipedia. Wikipedia. Wikipedia, Online: http://en.wikipedia.org/wiki/Wikipedia. Accessed: April 10, 2005.

[29] Wei Zhang. Peripheral members in online communities. 2001.