An Analysis of Open Content Systems

Ofer Arazy and Raymond Patterson, University of Alberta, School of Business

Abstract

Traditionally, content in organizational Knowledge Bases is created in a highly centralized manner to ensure quality. In Open Content Systems (OCS), on the other hand, content is generated in a distributed and decentralized manner. The most notable examples of OCS are Slashdot, the technology news portal, and Wikipedia, an online encyclopedia. The advantage of such systems is the speed in which content is accumulated, while the risk of open content systems is the lack of traditional quality control mechanisms. The purpose of this paper is to examine the processes that enable an open content system (OCS) to function effectively. We conduct a survey of existing open content systems, and analyze the interplay between the technology underlying OCS, the user community who is responsible for content generation, and the types of content managed by the OCS systems. Our analysis identifies specific settings where open content systems are likely to thrive.

1. Introduction

Open content systems (OCS) have emerged in recent years in the public domain, with little presence in corporate settings. OCS build on the social mechanisms of open source projects [Ljungberg 2000], the technology of many-to-many communication tools, and on the democratic philosophy of participatory journalism to create high quality knowledge bases. Although the technology underlying OCS is not very different from some of the technologies used collaborative applications, there is one fundamental difference between these two types of applications – content is collaborative software is transient and is only an enabler to collaboration, while in OCS generation of high-quality content is the objective.

Two types of architectures have been proposed for OCS: *append-based* (using discussion forums technology) and *overwrite-based* (using wiki technology) architectures. Append-based OCS enhance discussion-forums with feedback and control mechanisms, as demonstrated by Slashdot (http://Slashdot.org), arguably one of the most popular and influential OCS [Stadler and Hirsh 2002][1]. Slashdot is as an online news service, mainly for the Unix-based technological community. The alternative architecture – overwrite-based OCS – is based on wiki technology[2] [Cunningham & Leuf 2001; Wagner 2004], which was introduced in 1995. Overwrite-based systems, and specifically wikis, differ from append-based systems in one fundamental aspect: while in a discussion forum comments are added to one another to form a discussion thread, in wikis new content overwrites old content. A reader is shown only the last version in an overwrite-based system. Wikipedia [http://wikipedia.com; Remy 2002] is an online encyclopedia, which is founded on the wiki technology, and is probably the most popular overwrite-based OCS. Wikipedia came to life in 2001, and soon grew to become a popular online encyclopedia[3]. Examples of other popular OCS in the public domain include ePinions (www.epinions.com), a product recommendations forum which incorporates a feedback and reputation mechanism as well as social networks, and a series of Wiki-based projects [e.g. Wikinfo (http://www.getwiki.net), Wikicities (www.wikicities.com), Memory Alpha (http://memory-alpha.org), Uncyclopedia (http://uncyclopedia.org), and Tolkien Wiki (www.thetolkienwiki.org)].

While OCS emerge as an alternative approach to traditional content management systems in the public domain, the OCS approach is only now making its way into the corporate world.

---

[1] The impact of Slashdot is illustrated through "the Slashdot effect" where links from Slashdot to other sites often lead to high traffic to these referenced sites [Stadler and Hirsh 2002].

[2] Wiki technology is only but one specific operationalization of overwrite-based OCS. However, the term overwrite-based OCS is not restricted to wiki technology, and it is possible that alternative technologies that support overwrite-based OCS will emerge in the future.

[3] A comprehensive analysis of Wikipedia's features is found in [Lih 2004].

Like many other areas of computing, in the study of open content systems practice preceded theory, and since OCS is an emergent phenomenon research on the topic is almost nonexistent. Ljungberg [2000], builds on the earlier works of Kollock [1998] and Raymond [1999], and studies the social mechanisms in open source projects. Although these works focus on open source projects, they are very relevant to OCS. Cunningham and Leuf [2001] provide a comprehensive analysis of wiki philosophy and technology, and list 12 "wiki design principles". Stadler and Hirsh [2002] provide an introduction to the topic[4], by identifying some important characteristics of OCS: feedback and ranking (i.e. peer-review) mechanisms, decentralization of authority to reputable users (which implies the usage of reputation mechanism), free sharing of products, and flexible levels of involvement and responsibility. Wagner [2004] discusses the concept of "conversational knowledge management" and its implementation in wikis. Lampe and Resnick [2004] investigate one key OCS feature – the moderation (i.e. feedback) mechanism, through an empirical study of Slashdot. They find that a decentralized moderation mechanisms "can quickly and consistently separate high and low quality comments" [P. 1]. Barzilai-Nahon and Neumann [2005] conducted an empricial study regulation mechanisms in discussion forums, and describe the types of messages that were deleted. Emigh and Herring [2005] study OCS from yet another perspective – the communication genres of users – and shed light on the ways in which OCS users shape features of content.

2. The Technology of Open Content Systems

The main technological feature of OCS is that it enables decentralization of control and empowerment of users. Such distributed control requires that specialized mechanisms be put in place to enable collaboration and communication between users, to allow users to act as quality assurance personnel, to monitor the contribution of individual users (since user rights are often associated with contribution), and so on. Generally speaking, discussion forums which incorporate feedback and reputation mechanisms (such as Slashdot and ePinios) practice more quality control than wiki-based OCS, although there is some variety in control mechanisms between the various wikis. Content management system operate by three main processes (1) content generation, (2) quality control, and (3) knowledge dissemination or reading rights. Each of these processes can be decentralized, mostly if not entirely. Of these three processes, quality control has the greatest impact on overall content quality, and thus requires the highest levels of administrative control. Knowledge dissemination requires the least control. Content generation, the central process in OCS, is commonly decentralized, and in some applications involves a low level of administrative control (e.g. editing of stories in Slashdot).

3. User Community in Open Content Systems

The success of OCS relies on the user community, just as much as it relies on the technology. Since much of the responsibility in OCS projects lies in the hands of the users, the impact of the user community on project success is much greater in these projects than in traditional KB projects. OCS projects are characterized by a small and devoted community, which is linked by shared goals and strong community values. Work is mostly delegated and is coordinated by one or few "benevolent dictators" [Ljungberg 2000]. Users take pride in their collective efforts, and often the reward for one's contribution is social recognition. Similarly to "gift culture" [Kollock 1998; Raymond 1999], reputation which is gained through contribution is what brings social status and influence [Ljungberg 2000]. **Contribution** levels depend on three user traits: (a) benevolence, (b) competence, and (c) level of activity.

Although **benevolence** is assumed in OCS, as systems grow larger in scale, commitment to community lessens, and new types of users join the system, which are driven by other motivations. Opportunistic users try to "free-ride" and rip the project's fruits without contributing [Ljungberg 2000]; they are motivated by social status and often participate only in actions which contribute directly to their

[4] Stadler & Hirsh [2002] use the term "Open Source Intelligence" to describe OCS. We chose not to employ this term, as it used in the secret services world to refer to information which is gleaned from public sources.

reputation. A third type of users, malicious, aim to deliberately harm the system, by posting low quality or false content, deleting quality content, or tampering with the quality control process (e.g. providing unjustified rankings). Although the driving force behind OCS is benevolent users, successful OCS should be designed to handle users with alternative motivations, and reduce their impact on the system. An example of a successful OCS which serves mostly opportunistic users is ePinions, the product recommendation forum. **Competence** in a specific project relies on domain expertise. The few administrators of OCS are usually domain experts (e.g. meta-moderators in Slashdot and Advisors in ePinions), but this high-level of expertise cannot be assumed in the general user population. Users of a specific OCS share interest in the content topics, and thus have at least some basic domain expertise. Often few users are domain experts, and the rest could be considered professionals. Competence and benevolence of user authors are evident through the contribution of high quality content, and are often rewarded in social reputation. **Levels of activity** differ greatly in OCS communities. Although most users share the project's values, not all users can or want to participate at the same levels. Participation could be easily monitored in online systems, and is often rewarded in privileges (for reading, creating, or controlling the quality of content) and in social reputation.

Another characteristic of the user population is its **size**. OCS communities traditionally start small and homogenous, but become heterogeneous with growth, and original community values erode. Size is, thus, correlated with the other OCS characteristics reviewed above – benevolence (in larger and more mature communities opportunistic and malicious users are more likely to be found), activity levels (in larger communities activity levels will be more diverse, with a substantial group of free riders), and competence (a larger community is more diverse in terms of expertise).

3.1 The Interplay between Technology and People in OCS

Since the user community plays such a significant role in content creation and quality control, OCS should be designed to fit a specific user community. Particularly, there should be correlation between the level of centralization (or openness) the system maintains and the dispersal of contribution levels in the user population (where contribution level is determined by benevolence, activity levels, and competence). Figure 1 illustrates this relationship.

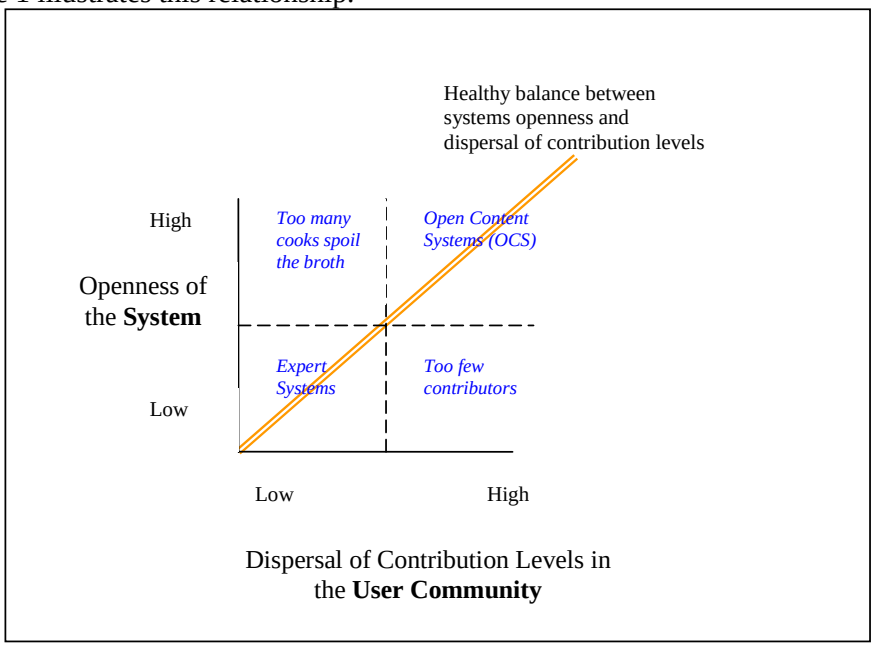

Figure 1: Dispersal of contributions in the user community versus system openness

In communities where only few users posses domain expertise and are willing to contribute, the system has to be highly centralized and allow only these few experts to create content and control for

quality, e.g. expert systems (bottom left cell). In similar communities with only few contributing users, an open system which allows the larger user population to take active part in content creation, will result in low quality content (top left cell), as unqualified users will contribute low-quality content. In an active community, where most users posses some level of expertise and are willing to contribute, systems with tight administrative control will under-utilize the community's potential (bottom right cell), as these many qualified and willing users will not be able to contribute. On the other hand, in active and collaborative communities, an open system will encourage participation and will utilize the potential of the community, as is the case with open content systems (top right cell). The diagonal line represents a healthy balance between the characteristics of the user population and the system's design; we can expect to see successful projects on that line. Expert systems, which have been used successfully in organizations for over 30 years now, represent one end of the line, while the newly emerging OCS represent the other end.

3.2. Control in Open Content systems

While it is likely that completely open and decentralized systems (such as Wikipedia) will succeed with the ideal cooperative user community, these pure OCS may fail with a more heterogeneous user population. To handle such populations, control processes are often used. These control processes could be performed by centralized administration, be decentralized to empower highly-contributing users, or be automated. **Editing of content** before it is published is often performed by administrators. Administrators may also **monitor some of the decentralized control processes** to remove low-quality content (Barzilai-Nahon & Neuman 2004) or punish members who misuse their privileges. For example, in Slashdot, few extremely highly-contributing users may become "meta-moderators" and participate in monitoring the decentralized ranking process. These "meta-moderators" monitor the feedback process, and punish offenders by revoking their privileges for a given time period. Similarly, in wiki-based projects administrators reserve the right to remove offensive users.

Decentralization of control empowers users to act as quality controllers, and is very useful in large-scale OCS. **Feedback mechanisms** allow users to rank the quality of posted content and often to reply with comments (or – in wikis – to overwrite existing content). Feedback serves two roles: (a) it is part of the content generation process, and (b) it used to determine the quality of postings, allowing readers to filter content based on its quality. The construction of online **social networks** within the user community and the specification of trusted sources could be used to personalize and filter the content a user views. For example, Slashdot allows a user to define his friends and foes and to bring to front postings by friends while filtering out postings by foes, and similarly ePinions allows users to define their trusted peers and uses the trust network to filter content.

Whenever possible, control processes should be automated. **Monitoring users' activity history** allows determining activity levels, and this information could be utilized in granting reading, writing, or quality control privileges (e.g. in Slashdot user receive moderation rights based on their use of previous moderation rights and their seniority in the system). Automatic control could also be used to detect malicious users and low-quality content, e.g. the anti-spam software robot in Kayak Wiki. **Reputation mechanisms** calculate users' contribution, often based on users' activity levels and the quality of their postings (as evident through the rankings these postings receive), or based on user's centrality in the network (i.e. the number of "friend members" they have). Reputation mechanisms have three roles. First, publishing and quality control privileges could be based on user's reputation. Second, they serve to motivate users to actively contribute to the projects, as reputation scores are commonly made available to the entire community and one's reputation reflects its social status. Moreover, monetary compensation could be associated with reputation, as in ePinions. Lastly, readers could filter content based on author's reputation. Reputation mechanisms, which became very popular with the success of eBay, prove to be powerful tools and are an essential component of large-scale OCS projects. As discussed above, one of the central roles of control mechanisms on OCS is to determine user privileges. Table 1 describes the role of control mechanisms in the various OCS processes.

Control Mechanisms	Dissemination (Reading Rights)	Knowledge Creation (Publishing Rights)	Quality Control (Moderation Rights)
Centralized			
Editing content for publication			Administrators edit content before it is published.
Control ranking process			Administrators control the ranking process (e.g. meta-moderation in Slashdot)
Decentralized			
Feedback Mechanisms	Reading of content could be filtered by the ranking of content		
Social Networks	Reading of content could be personalized and filtered based on user's social network		
Automatic			
Monitoring User Activity	Reading rights depend on activity (e.g. only registered users can view content, or malicious users' rights could be revoked)	Publishing rights depend on activity history (e.g., only users who were active readers in the past 3 months can publish)	Moderation (i.e. ranking) rights depend on activity history (e.g. only users who used the system for 2 years can moderate). Meta-moderation rights depend on activity history.
Reputation Mechanism	Reading of content could be filtered by the reputation of content's author	Publishing rights depend on user reputation	Moderation rights depend on reputation. Meta-moderation rights depend on reputation.

Table 1 - control mechanisms affect user privileges in OCS processes

OCS based on wiki open technology require additional control mechanisms. One of wikis' features is the continuous archival of content pages, and the ease with which pages could be reversed to their archived content, making it more difficult to damage wiki pages than to restore the original content. This imbalance tips the symmetry between benevolent and malicious users, and is one of the key contributors to wikis' success [Stadler & Hirsh 2002]. An additional control mechanism used in some wiki projects, for instance Wikipedia, allows users to receive alerts when pre-defined content pages have been modified, enabling immediate user reaction.

In the private domain, due to the lack of experience, it is hard to determine what types of controls would be appropriate. However, since adherence to norms could be enforced through organizational apparatus, we believe that the need for control mechanisms is not as high.

4. Content Types for OCS

Two characteristics of content should be considered. First, the extent to which knowledge of the topic is common is correlated with the number of potential contributors. OCS dealing with esoteric knowledge may attract small number of readers, and even fewer experts. Second, the extent to which knowledge is factual (as opposed to knowledge that is ambiguous and depends on opinions) is important, as opinion-based knowledge is open to interpretation and requires special conflict resolution mechanisms.

Esoteric knowledge types imply a small user community. As long as there are some experts in the community, and the community remains small, it may thrive. However, if the project attracts more users, and these users contribute to content generation despite their lack of expertise, the system's quality is

bound to deteriorate. Thus common knowledge types are more appropriate for large-scale OCS. Some OCS projects encourage multiple world-views (e.g. Wikinfo, which promotes "sympathetic" postings, and Uncyclopedia, a humor wiki), while other projects aim at consensus (e.g. Wikipedia). OCS which tolerate contradicting opinions can deal with either factual or opinion-based (e.g. politics) knowledge. On the other hand, OCS which seek consensus, deal better with factual knowledge, and require special conflict-resolution mechanisms when handling opinion-based knowledge (this is especially true when the community grows, and users no longer share the same philosophy and goals). Opinion-based knowledge, when used in OCS projects with a diversified user-base, require special conflict resolution mechanisms. For example, discussion of opinions outside the main content area, polling or voting on controversial issues, or the ability of administrators to dictate the "official" world-view in conflict situations. Wikipedia employs all these mechanisms [Lih 2004], and this is probably one of the main reasons the project continues to thrive despite the inherent open nature of wiki technology and the growing diversification in its user base [Lih 2004].

5. Conclusions

This paper studied open content systems in an effort to understand if and how decentralization of control could yield high-quality content. We analyzed the interplay between technology, user community, and content types in OCS. Our analysis suggests that (1) the most critical process in OCS is quality assurance (2) a collaborative user-community is key for OCS projects success, although (3) effective control measures – either automatic, decentralized, or centralized - can help deal with non-benevolent users, and that (4) the openness of OCS should be correlated with the dispersal of contributions in the user community. An economic analysis of corporate versus internet settings demonstrates that the impact of quality assurance mechanisms, such as reputation, differs between these settings. Our analysis indicates that corporate settings may need additional OCS features, compared to internet settings, to thwart non-benevolent behaviour. We hope that this paper will draw attention to the emerging OCS approach in the research community, and believe that our findings could be used to direct the design of new OCS.

6. References

Barzilai-Nahon K. and Neumann S., 2005, Bounded in Cyberspace: An empirical model of self-regulation in virtual communities, HICSS-38 (Hawaii International Conference on system Sciences).

Cunningham W. and Leuf B., 2001, The Wiki Way, Quick Collaboration on the Web. Addison-Wesley.

Emigh W. & Herring S., 2005, Collaborative Authoring on the Web: A Genre Analysys of Online Encyclopedia, Proc. Of the 38th Hawaii International Conference on System Sciences.

Kollock P., 1998, The Economies of Online cooperation, In Communities in Cyberspace, Routeledge, London.

Lampe C. & Resnick P., 2004, Slash(dot) and Burn: Distributed Moderation in Large Online Conversation Space, Proc. Of ACM Computer Human Interaction, Vienna.

Lih A., 2004, The Foundations of Participatory Journalism, Conference paper (Association for Education in Journalism and Mass communication; Toronto, August 7, 2004).

Ljungberg, 2000, Open Source Movement as a model for Organizing, European journal of Information Systems, 9, pp. 208-216.

Raymond E.S., 1999, The Cathedral and the Bazaar, O'Reilly, Sebastopol.

Stadler & Hirsh, 2002, Open source Intelligence, First Monday, Vol. 7, No. 6.

Wagner C., 2004, "Wiki: A technology for conversational knowledge management and group collaboration". Communications of the association for Information systems. (Volume 13, 2004) 265-289.