

Forming and Sustaining Online Communities: The Role of Distributed Collective Information Practices

Robert J. Sandusky

Assistant Professor, School of Information Sciences, University of Tennessee, Knoxville, TN 37996

sandusky (at) utk (dot) edu

Abstract. This paper describes an approach to understanding how online communities form and sustain themselves by focusing attention on distributed collective information practices. Distributed collective information practices are the fundamental human information-related activities, mediated by specific socio-technical constraints and contexts, which frame participation in online communities. Research results from an examination of distributed software problem management in a thriving open source software development community are presented to illustrate distributed collective information practices.

Introduction

We are surrounded by communities that combine face to face and distributed modalities. My research includes the examination of the problem solving activities of a Free / Libre / Open Source Software (FLOSS) development community. Instant messaging, e-mail, wikis and blogs are among the media I use to try to stay in touch with people with whom I once shared physical proximity. In my professional life, I teach graduate students in information sciences using both synchronous and asynchronous distance education technologies. Members of local communities are full partners with academics in action research projects that aim to empower, transform and create new community capabilities, often through

the application of information and communications technologies to local problems.

In this paper, I propose that researcher sensitivity to the existence and use of *distributed collective information practices* (DCIP) may be useful in determining how, why, and under which conditions online communities provide incentives for participation and ultimately sustain themselves. I present examples drawn from research into the distributed software problem management practices used by a large, thriving FLOSS development community.

Distributed Collective Information Practices in FLOSS

I am part of a research group that is examining the software problem management (i.e., bug fixing) activities of the Mozilla project, one of the largest and most successful FLOSS development projects. Software problem management (SWPM) is one part of the larger FLOSS software development process. The Mozilla project has been responsible for producing more than twenty software products and sub-products, two of which, the Thunderbird e-mail client and the Firefox Web browser, have been adopted by millions of users around the world. The Mozilla community's mission statement lists the following responsibilities, which are focused on centralized code integration:

- Provide technical and architectural direction for the project
- Collect and synchronize code changes
- Create source code releases
- Operate discussion forums
- Coordinate bug management (that is, perform software problem management)
- Provide documentation (code "road maps")
- Help people achieve consensus

Like many FLOSS projects, large and small, the Mozilla project depends upon a large, distributed group of people to perform most of the work: design, coding, and software problem management. A smaller, core group manages the project web site, runs the bug report repository and code repositories, etc. This model is sometimes termed "distributed development with centralized integration."

We view distributed FLOSS development generally and distributed SWPM in particular as exemplars of *distributed collective practice* (DCP). A DCP is comprised of several elements, including *activities*, *information objects*, *socio-technical infrastructures*, *common problems* and *common methods and processes* (Gasser & Ripoché, 2003). We consider SWPM as a sub-DCP of the larger software development process and examined SWPM across all of the twenty-odd Mozilla products. The Mozilla community can be viewed as comprising multiple

sub-DCPs, although it may be difficult to draw precise lines between them. For example, the provision of technical / architectural direction or the creation of source code releases could be considered separate sub-DCPs within the larger software development process.

One approach to identifying and understanding the mechanisms by which online communities form, self-sustain and provide incentives for participation is through the identification of *distributed collective information practices* (DCIP) (Sandusky, Gasser & Ripoché, 2004b). The concept of DCIP emerged as our analysis progressed, and we identified several ways in which activity, information objects, methods and processes, and socio-technical infrastructure interact in distributed SWPM management (Sandusky, 2005).

The notion of DCIP is grounded in well-established research traditions in information science and can be conceptualized as a series of nested boxes (see Figure 1, below). The innermost box represents the continuing tradition of experimental, quantitative and qualitative research into information searching by individuals. The enclosing box represents research into information seeking by individuals, a more general class of behaviors that includes other activities such as browsing (Rice, McReadie & Chang, 2001). The next box represents research into individual human information behaviors more generally, which includes activities other than seeking and searching, such as information categorization, use and creation. The fourth box represents information practices, which acknowledges the social nature of human information behaviors in co-located group or community settings (McKenzie, 2003; Fisher, Durrance & Hinton, 2004). The fifth, outermost box represents DCIP, which translates these existing research traditions to the world of distributed collective practices, the ICT-mediated social information behaviors of the members of distributed communities.

In the FLOSS setting we examined, community members identify and mark individual bug reports as being related to other bug reports, creating new virtual information objects, *bug report networks* (BRNs). Linking bug reports together simultaneously re-orders information (e.g., marking dependency or duplicate relationships), creates new information objects (the bug report network), and transforms the community's social order (by bringing the people associated with separate bug reports into a new relationship around the bug report network). BRNs are common in this community: 65% of the bug reports sampled are members of one or more bug report networks and 33% are marked resolved because they are duplicates of another existing bug report. Informal relationships (that is, various *see also* relationships that are not formally supported by the bug report repository) occur in 33% of the sampled bug reports (Sandusky, 2005).

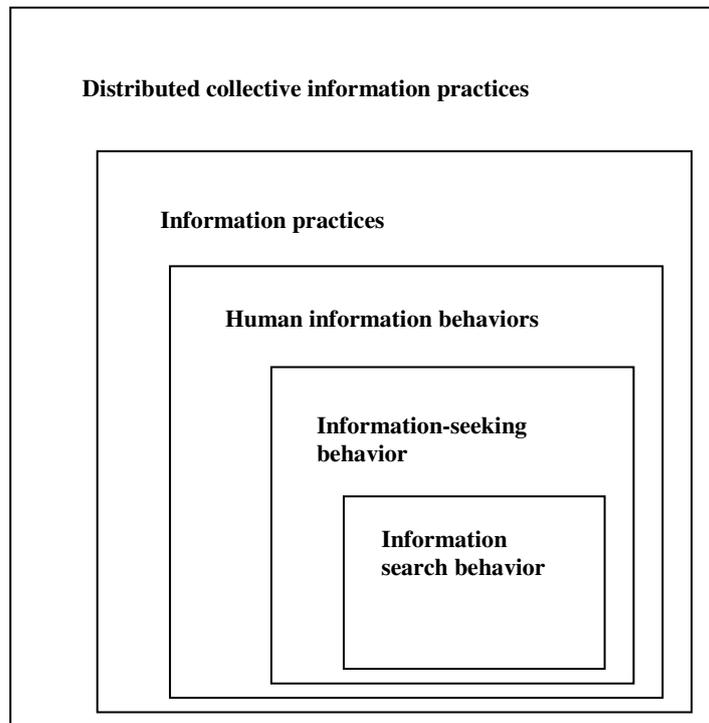


Figure 1 – Relationship of DCIP to Information Research Traditions

Data from our FLOSS research and results reported by other investigators suggests that DCIPs occur in many different settings. Some identified so far include *information compounding* (Paepcke, 1996): information compounds are new forms of information spontaneously created by the members of a community from multiple, disparate information sources. Paepcke (1996) identified this practice in several departments within a computer manufacturing company. Information compounding was also found to be a common practice in the real-time management of distributed infrastructure (Sandusky, 2003). Second, the creation of *networks of information objects* (Sandusky, 2005) as community members assert specific formal and various informal relationships between information objects contained in a communal repository was found to be a common activity in the FLOSS development community we studied, as the description of bug report networks (above) reveals. Third, *negotiation* (Strauss, 1978) is a basic social process frequently employed by the Mozilla community during SWPM to both manage information and coordinate activity to support problem solving (Sandusky & Gasser, 2005). The parties to the negotiation present evidence (code; system behaviour; system-generated artifacts such as stack traces) to support their planned or actual courses of action. Finally, the Mozilla community also employs a communal, distributed search strategy as part of its SWPM management process (Sandusky, 2005). The communal search strategy is used to prevent the creation of duplicate bug reports, identify duplicate bug reports when they exist, and identify and formalize relationships between bug reports in support of solving software problems.

Conclusion

The incentive structures within FLOSS development communities have been examined from several points of view in recent years. Factors such as the enhancement of a programmer's professional reputation, personal interest in particular problems ("scratching an itch"), or work-for-pay within FLOSS projects that have hybrid volunteer / commercial organizational structures (such as the Mozilla project) have been identified, but further empirical verification is warranted.

Additional research and analysis of DCIPs should be performed in a variety of online and mixed community settings to verify the assertion that information practices, when revealed, can contribute to our understanding of how and why some online communities thrive and others do not. Improved understanding of DCIP can then be used to enrich social and technical mechanisms that support the formation and sustenance of these communities. Settings that have been examined through the lens of distributed collective practice include scientific collaboratories, distributed engineering, software engineering, medical informatics and geographic information systems (Turner, et. al., 2004). All of these settings are DCPs: they involve *activities, information objects, socio-technical infrastructures, common problems and common methods and processes.*

Acknowledgments

This material is based in part upon work supported by the National Science Foundation ITR (Digital Society and Technologies) program under Grant No. 0205346. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. I thank the workshop organizers for their helpful comments on this paper.

References

- Fisher, K.E., Durrance, J.C. & Hinton, M.B. (2004). Information grounds and the use of need-based services by immigrants in Queens, New York: A context-based, outcome evaluation approach. *Journal of the American Society for Information Science and Technology*, 55(8), 754-766.
- Gasser, L., & Ripoche, G. (2003). Distributed collective practices and free/open-source software problem management: perspectives and methods. *2003 Conference on Cooperation, Innovation & Technologie (CITE'03) (Université de Technologie de Troyes, France, December 3-4, 2003).*

- McKenzie, P. J. (2003). A model of information practices in accounts of everyday-life information seeking. *Journal of Documentation*, 59(1), 19-40.
- Paepcke, A. 1996. Information needs in technical work settings and their implications for the design of computer tools. *Computer Supported Cooperative Work: The Journal of Cooperative Computing*, 5(1):63-92.
- Rice, R. E., McCreadie, M., & Chang, S.-J. L. (2001). *Accessing and browsing information and communication*. Cambridge, MA: MIT Press.
- Sandusky, R.J. (2003). Infrastructure management as cooperative work: implications for systems design. *Computer Supported Cooperative Work: The Journal of Collaborative Computing*, 12(1), 97-122.
- Sandusky, R. J. (2005). Software problem management as information management in a F/OSS development community. *OSS 2005: The First International Conference on Open Source Systems*. Genova, Italy, July 11 - 15, 2005.
- Sandusky, R. J., & L. Gasser. (2005). Negotiation and the coordination of information and activity in distributed software problem management. *GROUP '05: ACM 2005 International Conference on Supporting Group Work*. Sanibel Island, Florida, November 6 - 9, 2005.
- Sandusky, R.J., Gasser, L., & Ripoche, G. (2004a). Bug report networks: varieties, strategies, and impacts in a F/OSS development community. *MSR 2004: International Workshop on Mining Software Repositories, ICSE 2004, IEEE International Conference on Software Engineering (Edinburgh, Scotland, UK, May 25, 2004)*.
- Sandusky, R. J., Gasser, L., & Ripoche, G. (2004b). Information practices as an object of DCP research. *Distributed Collective Practice: Building New Directions for Infrastructural Studies. Workshop of the CSCW 2004 Conference*. (Chicago, IL, US, November 6, 2004).
- Strauss, A. (1978). *Negotiations: varieties, contexts, processes, and social order*. San Francisco: Jossey-Bass.
- Turner, W., Zacklad, M., Bowker, G., Gasser, L., Karasti, H., Schmidt, K. (2004). Distributed Collective Practice: Building new Directions for Infrastructural Studies, Workshop of the CSCW 2004 Conference. Available at: <http://tech-web-n2.utt.fr/cscw04/#Accepted%20Position%20Papers>. Accessed: [2005-10-14].