The Problem of Private Information in Large Software Organizations

Jonathan L. Krein[†], Patrick Wagstrom[‡], Stanley M. Sutton Jr.[‡], Clay Williams[‡], Charles D. Knutson[†]

Brigham Young University[†]
2204 TMCB, Provo, UT 84602
jonathankrein@byu.net, knutson@cs.byu.edu

IBM T. J. Watson Research Center[‡]
P.O. Box 704, Yorktown Heights, NY 10598
{pwagstro, suttons, clayw}@us.ibm.com

ABSTRACT

Coordination of project stakeholders is critical to timely and consistent software delivery. In this short paper we present *the problem of private information* as a guiding framework or lens through which to interpret coordination dynamics within software organizations. We provide evidence of this problem in the form of specific challenges, collected via interviews from a diverse set of *extended* (i.e., non-development) stakeholders in a globally distributed software development organization.

Categories and Subject Descriptors

D.2.9 [**Software Engineering**]: Management; K.6.1 [**Management of Computing and Information Systems**]: Project and People Management—*life cycle, management techniques, strategic information systems planning*

General Terms

Economics, Human Factors, Management, Theory

Keywords

Coordination, Extended Stakeholders, Information Management, Private Information, Software Process

1. INTRODUCTION

The ultimate challenge facing any software development organization is the efficient delivery of quality software that meets intended user needs. To achieve this goal, software engineering researchers have traditionally focused on improving processes for software developers. For large projects, researchers have studied developer coordination, related to software artifacts [3, 12]. Some of these studies address coordination within large-scale development [1, 13]; others focus on challenges arising within distributed development [7, 10].

Though critical, product development is only a small part of product delivery. *Extended* (i.e., non-development) project stakeholders play vital roles in the delivery of software products, and yet have historically been under-studied within software engineering. Few papers in the literature discuss the needs of the broader stakeholder community. Williams et al. [14] discuss the criticality of extended stakeholders to software delivery—including marketing, sales, services, strategy, support, legal, executives, and customers—and propose a framework for enterprise software development coordination. Berndt, Jones, and Finch [2] also outline the need for coordination support across the extended stakeholder community and discuss information markets as a possible solution.

Coordination of extended stakeholder communities is critical to solving fundamental problems of timely and consistent software delivery. For large corporations in particular, delivering software products requires the coordination of multiple layers of non-development stakeholders. Consequently, we set out to understand the mechanisms and challenges of coordination in software organizations from the perspective of extended software project stakeholders.

We conducted a series of interviews with stakeholders across IBM, utilizing a grounded theory approach. Our goal was to understand the coordination processes and pain points of the organization, with a focus on more fully adapting corporate process to the global delivery of complex systems. In this paper we present one of the theories that emerged from our study, *the problem of private information*, as well as an initial formulation of that problem. As a lens through which to view coordination dynamics, we believe this theory can reveal novel solutions to long-standing coordination problems.

In Section 2 we introduce the problem of private information, leveraging prior work in economics. We present stakeholder coordination challenges in Section 3 as evidence of the problem from a large software organization. Section 4 relates the interview data to the central theory. In Section 5 we present conclusions.

2. THE PROBLEM OF PRIVATE INFORMATION

Borrowing from the literature in economics [9], we refer to the challenge of utilizing distributed knowledge in an organization as *the problem of private information*, where *private* refers to information possessed by a relatively small segment of the population—as opposed to information that is widely held.[1] In addition to substantive, objective knowledge, *private information* includes subjective, biased knowledge such as heuristics, intuition, and preference. The privacy (or

[1]Although the term *local information* is arguably more precise given this definition, *private information* is already an established term in economics [8].

ICSE '11, May 21–28, 2011, Waikiki, Honolulu, HI, USA

locality) of information becomes a problem when information is not available where and when it may be needed or useful.

Although not specifically treated in software engineering, the problem of private information is explored in economics literature. In *The Use of Knowledge in Society* [9], Friedrich Hayek describes the purpose of economics as the production of a rational economic order—the optimal allocation of resources as circumstances evolve. Hayek posits that *if* a mastermind could exist, one possessing all relevant information, then by means of economic calculus a central authority could guarantee a rational economic order. He then declares:

> This, however, is emphatically *not* the economic problem which society faces... The peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form, but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess. The economic problem of society is thus not merely a problem of how to allocate "given" resources—if "given" is taken to mean given to a single mind which deliberately solves the problem set by these "data." It is rather a problem of how to secure the best use of resources known to any of the members of society, for ends whose relative importance only these individuals know. Or, to put it briefly, it is a problem of the utilization of knowledge not given to anyone in its totality. [9, pp. 519–20]

In other words, a mastermind not only *does not* exist, it *cannot* exist. Hayek's point is that the distributed nature of information requires us to think differently about available solutions. This distinction is important because the clear identification of a problem, relative to a particular context, heavily influences the visible solution set for that context.

The fundamental "problem of the utilization of knowledge not given to anyone in its totality" is inherent to all large-scale, socially-dependent production tasks—in part because information is discovered (or created) locally. Causes of private information in software production include logistical constraints that can be alleviated but not eliminated. Further, many challenges that projects encounter, as discussed in the next two sections, seem to occur because of complications related to the management of private information.

In the next section we present evidence of the problem of private information, specific to the context of extended stakeholder communities in large-scale software development.

3. EVIDENCE OF THE PROBLEM

We conducted nine one-hour interviews with extended software project stakeholders. Twelve subjects represented projects from across IBM. Multiple scribes assisted with each interview, and five of the interviews were recorded. We used a semi-structured format, beginning with pre-selected open-ended questions designed to elicit coordination challenges and crisis examples, after which we explored the context of each through unstructured discussion.

Coordination challenges represent the subjects' own evaluative generalizations of coordination problems, whereas *crisis examples* reflect the underlying experiences. A "crisis" is any scenario in which a project's key success indicators are threatened; crises are not necessarily extreme events and may occur multiple times throughout a project's life cycle. We did not encourage subjects to link coordination challenges to crisis examples in order to allow for cross-validation of categories against multiple data sources.

Interview responses reflect broad experience. Most subjects have worked in multiple company divisions, and some currently serve in multiple capacities within their division. Of the twelve subjects, four serve as corporate process owners (managing process globally), six as process administrators (managing process within a division), two as upper managers, two as middle managers, and one as a lower-level manager. Although the data represent stakeholder challenges from the perspective of management, all of the subjects work with, coach, and train people from various capacities, so they are familiar with other perspectives as well. Most of the subjects have long histories with corporate process and its practical application to software development/delivery.

All stakeholders utilize a common corporate process for delivering products. This process, referred to as the Integrated Product Development (IPD) process, has been used company-wide for all regular products since 1998 [6]. IPD is a Stage-Gate^{TM} process that explicitly supports coordination among cross-functional stakeholders. IPD is maintained at the corporate level and deployed in each division with minor customizations. The process has been used to deliver thousands of products and is quite mature.

Analysis of interview data followed the structured approach of Corbin and Strauss [4]. Three authors identified coordination challenges/crises (open coding), after which they cross-analyzed these data sources to produce a set of eight generalized categories (axial coding). We selected private information as a lens through which to interpret coordination mechanisms because of its prominence in both challenge descriptions and crisis examples. Based on this choice of central theory, we eliminated two top-level categories[2] and merged three others (selective coding), thereby arriving at the four top-level classes presented in this paper.

This study utilizes three validation strategies: 1) cross-validation, which involves comparing final categories against both crisis examples and challenge descriptions to confirm their grounding in the data; 2) external review—provided by advisors from both Computer Science and Sociology; and 3) member checking, which confirmed the salience of our categories. Regarding member checking, IBM's process team has previously addressed many of the challenges we discuss. However, in this paper we focus on articulating the problem rather than examining existing solutions.

In the following subsections we present the four coordination challenges that featured prominently in stakeholder interviews.[3] In Section 4 we discuss how each of these challenges relates to the problem of private information.

3.1 Structural/Operational Diversity

All large organizations deal with structural and operational diversity as a result of functional specialization.[4] Cross-functional teams may be an effective tool to mitigate such

[2]We pruned *process auditability* and *stakeholder overload*; these categories were not well-saturated in the data.

[3]Stakeholders spoke in terms of specific processes and products; details (including specific quotes) are omitted in agreement with requests for confidentiality.

[4]Note that *structural/operational diversity* does not refer to the heterogeneity of individuals and personal experiences.

obstacles. However, complications still emerge from the natural diversity of large organizations, including: cross-functional breadth of teams, team size, number of products managed, geographic distribution of team members, tooling, and operational rules. Also, meeting schedules for upper and middle management were described as varying significantly in both frequency and length. Teams that meet less often or for shorter periods of time likely coordinate in informal ways, outside of standard process, which tends to produce additional operational diversity. All of these factors impact coordination and complicate internal software tooling.

Several subjects commented specifically on solutions delivery, which requires integration of the primary business units: software, systems, and services. Although business units utilize a common corporate process structure, they each maintain customized process frameworks, based on domain-specific needs. Process customization is an indispensable consequence of diversity in large organizations. However, it also opposes the unifying purpose of standardized process, thereby introducing additional coordination complexities. Ultimately, in large organizations the need for global consistency must be balanced against the need for local flexibility.

The interview subjects also identified several complicating elements, including acquisitions, short-staffing, and a reliance on ad hoc dynamic communication channels between teams (informal coordination). In general, interview subjects seemed to agree that structural and operational diversity, although not entirely avoidable, create process inefficiencies and can inhibit cross-functional coordination.

3.2 Tooling Diversity

Although internal software tools function well locally, most tools cannot be universally shared or fully integrated due to the size of the organization, the diversity of tooling needs, and the sheer number of tools utilized. According to stakeholders, the massive diversity and consequent interoperability limitations of tools presents unique obstacles for coordination both within and across stakeholder teams.

Why are general cross-domain software tools essentially non-existent industry-wide? A software organization may provide its developers a standard environment (e.g. Eclipse), but that same interface will not be acceptable (or even useful) to business executives (or to marketing, sales, strategy, etc.). Because each domain communicates in terms of its own disparate parlance and goals, concepts cannot simply be shuttled from one domain to another. Global sharing of information requires filtering knowledge and translating semantics across domains—a hard problem, even for humans.

3.3 Process Experience and Awareness

Multiple subjects noted that process inexperience and unawareness are normal challenges that often arise from common business practices. For example, employee rotations involve members joining teams for a limited period of time, after which they move to another area of the company. Turnover introduces new members with little or no company experience. Reorganizations, which can occur in response to market forces, customer needs, or acquisitions, cause general staffing changes. All of these changes lead to temporary process inexperience as members adjust to new roles.

Process inexperience is particularly problematic when it affects core positions at higher organizational levels. A core stakeholder team actively manages one or more projects, but each member of that project team also belongs to a functional team (thus project teams are cross-functional). Other members of a stakeholder's functional team may substitute when the stakeholder is unable to fulfill a particular responsibility (e.g., due to sickness). Alternates should be as knowledgeable about the process as the principal stakeholder, but for similar reasons may not be.

Interview subjects also described some stakeholders (conceptually) as "corporate consultants"—employees who provide expert advice on specific topics to the rest of the company. Corporate consultants generally interact with corporate process on a more specialized level than do core stakeholders or functional teams. Thus process awareness tends to be more localized and specific for corporate consultants.

The same appears to be true of employees at lower organizational levels. Employees on lower-level teams may be involved with corporate process only in the context of a specific organizational unit or function, and so may lack a cross-organizational or cross-functional perspective. They may be very experienced within their own roles, but may lack awareness of or information about the concerns of the larger process. Thus for both corporate consultants and low-level teams, organizational specialization promotes process specialization, which can result in process unawareness.

3.4 Information Flow

The fourth challenge describes information flow between stakeholders. We discuss this challenge from three perspectives, as reported by interview subjects: *loss of specialized information* (due to staffing changes), *information pull*, and *information push*, where *pull* and *push* refer to information sharing between core project stakeholders and subject matter experts[5] (SMEs). In general, the management of organizational knowledge, which is unavoidably distributed, lies at the heart of the problem of private information.

3.4.1 Loss of Specialized Information

Loss of specialized information arises from both reorganizations and loss of employees. Reorganizations invariably require some realignment and training for team members and can create confusion and disagreement regarding role assignments, both of which create and prolong information gaps in teams. In addition to causing problems with process unawareness (Section 3.3), employee loss also creates information gaps, particularly for teams that depend on cross-functional representation. In general, cross-functional coordination is threatened by staffing changes, which can create coordination gaps that prevent specialized information from being passed between project stakeholders.

3.4.2 Information Pull

Stakeholders often struggle to identify necessary SME collaborations within the organization, and when collaborations are known, they often struggle to utilize the "available" information. We refer to these struggles as the challenge of *information pull*.

According to one subject, stakeholder relationships are similar to the layers of an onion, with core stakeholders at the

[5]Interview subjects used the term *subject matter expert* to refer to any individual that manages a specialized body of knowledge—of course, it is helpful to recognize that most stakeholders possess specialized information on some topic, even if not formally tasked with the care of that knowledge.

center, relying on peripheral individuals at varying degrees of connectedness that act as SMEs (see the discussion of core stakeholders, functional teams, and corporate consultants in Section 3.3). SMEs are not necessarily connected into the core, and existing connections may be weak. Generally, the responsibility lies with core stakeholders to establish and maintain connections with SMEs, pulling information as needed. Because core stakeholders bear the primary responsibility for obtaining critical project information, they must maintain a broad awareness of the shifting information-landscape within the organization, and once new information becomes available, they must become aware of the change and locate the relevant SMEs.

Lack of awareness by core stakeholders of a need for information often contributes to coordination breakdowns. Multiple subjects indicated that the amount of changing information is just too great for corporate process (or any central authority, including a project team) to manage all of the relevant connections. The problem is magnified by the large, diverse information needs of core stakeholder teams, the global distribution of stakeholders and SMEs[6], and the relative disconnectedness of day-to-day core stakeholder communication networks from those of SMEs.

This challenge surfaced in several crisis situations collected during the interviews. In one example, a team had to deal with the release of their product in a country where the import/export regulations had recently been changed. In that case, the project team did not become aware of a key SME until after the crisis began.

In general, information pull appears to be a costly process that results in decreased efficiency and loss of knowledge because stakeholders often lack a complete picture of available information in the organization.

3.4.3 Information Push

Dissemination of information to teams is a frequent challenge for SMEs. In the previous example, in which a country altered its import/export laws, a SME in the organization knew about the change before the crisis. In fact, in most of the crisis examples collected during the interviews, someone in the organization possessed information that could have prevented the crisis had it been disseminated to the appropriate stakeholders. Sometimes critical information is delayed in processing, but often SMEs are unaware of or cannot locate other stakeholders with a need to know. Organizational diversity and staffing changes compound the problem.

To grapple with these complexities, SMEs utilize information repositories (e.g., wikis) to publish information to stakeholders. Under this model, each stakeholder is responsible for tracking changes by referencing the relevant repositories. However, due to structural/operational diversity, these databases are presented in specialized, diverse, and evolving formats. Interview subjects indicated that the sheer number of SMEs, coupled with the bulk of "available" information, creates significant logistical complexities for establishing and maintaining information channels. We found that even in cases where a SME was attempting to share private information thought to be relevant, the information often was not fully ingested. In such cases, the SME could not understand precisely which information the stakeholders actually needed or how best to present/distribute that information.

[6]For example, one project was spread across Germany, Australia, China, India, and two sites in the United States.

The need for information exchange is always known to the organization as a whole. When knowledge from two or more individuals is juxtaposed, it becomes clear whether they need to share information. However, due to the distributed nature of large organizations, both information push and pull seem fraught with complications. Because of their limited perspectives, core stakeholders and SMEs both seem inadequately positioned to orchestrate an optimal flow of information.

4. THE PROBLEM IN CONTEXT

The complementary struggles of information pull and push are direct expressions of the problem of private information. No central authority is capable of coordinating an optimal flow of information—core stakeholders encounter limitations in pulling information from SMEs, and SMEs encounter limitations in pushing information to core stakeholders.

Like market economies, the solution must necessarily be a distributed one. Hierarchical organizations are an attempt to create such distributed solutions. However, the diversity that results from specialization amplifies the problem of private information, making it more difficult to overcome. For example, tooling diversity results in too many tools to effectively interface, and in connection with structural/operational diversity, it makes standardization of communication channels between sub-organizations difficult. Structural/operational diversity also inhibits stakeholder awareness of cross-functional information, and since information is distributed among individuals, the organization is constantly under threat of losing specialized information as the organization evolves.

Corporate process attempts to mitigate many of these challenges by systematizing coordination, but natural organizational change and functional specialization create additional complications for ensuring universal process experience and awareness among stakeholders. Further, no process can anticipate emerging changes quickly enough to prevent all crises. For improved crisis management, the solution must lie in some type of distributed sensing mechanism that can expose emergent behaviors for quick short-term resolutions, while leading to long-term process changes and strategies.

From this analysis we formulate a set of premises which we believe describe the problem of private information as it occurs generally within large software development environments (bracketed terms refer to the challenge categories):

The organization of interest is large, with thousands of participants, most of whom serve critical roles as stewards of some set of specialized information. Whether under assignment or simply as a matter of circumstance, SMEs are critical to the success of the organization [*Loss of Specialized Information*, *Information Pull* and *Information Push*].

Further, the sum of information in the organization is too great for any individual to fully possess. Thus the organization functions best when information flows optimally between participants [*Information Pull* and *Information Push*]. Hierarchical structure and corporate process are designed to facilitate the flow of information [*Structural/Operational Diversity* and *Process Experience and Awareness*].

As one moves outward from any particular set of core project stakeholders, the number of relevant actors grows quickly, as does their relational complexity and the volume of specialized information. Sorting out who needs information from whom at any given moment, let alone over time as circumstances change, is a complex problem [*Information Pull* and *Information Push*].

Diversity across organizational subunits [*Structural/Operational Diversity*], including the diversity and non-interoperability of coordination tools and information repositories [*Tooling Diversity*], as well as turnover and short-staffing [*Loss of Specialized Information*] complicate the problem.

In this environment, no central authority is capable of fully knowing or guaranteeing the optimal flow of information.

5. CONCLUSIONS

In our efforts to understand the coordination needs of extended stakeholders, we discovered that effective management of private information among stakeholders is critical to timely and consistent product delivery. The *problem of private information* lays the foundation of a conceptual framework for interpreting coordination processes in software organizations. We believe this framework can guide the articulation of new concepts from which we may generate novel solutions to long-standing coordination problems.

As an example of new approaches enabled by the notion of private information, consider *information markets*. Hayek's theory of market economics demonstrates that markets aggregate and disseminate information of all types, and resolve informational conflicts [8]. Economists have leveraged these concepts to design asset markets for predicting future events [11]. In 2005, Google reported success in forecasting events using internal information markets [5]. The company has also studied the trading behavior of its employees within the markets [15]. Google's studies suggest two research objectives for software development: 1) test whether market aggregation of stakeholder estimates can produce useful software project indicators [2]; and 2) test whether the dynamics of stakeholder participation in an aggregate estimate can reveal additional project indicators. These objectives represent methods for eliciting, collecting, and interpreting private information.

However, information markets only scratch the surface of potential solutions to the problem of private information in software organizations. For example, they do not address many of the complexities that core stakeholders face in pulling information from SMEs. Rather, they demonstrate that within large groups of people, private information is often under-utilized, offering at least one technique for taking advantage of some of that information. Information markets are also well-explored in economics literature and have been demonstrated to work as aggregators of distributed information in numerous contexts. Thus they represent a validating example of the type of innovative thinking that can emerge from formulating coordination mechanisms in terms of the problem of private information.

Since the problem of private information is a new theory, we utilized qualitative (complexity-preserving) methods to discover its features and implications. However, this study is only a beginning. Without additional investigations into other contexts, findings are limited and may not generalize well to other organizations. For example, further work needs to be conducted to determine the impact of private information on small software organizations—not to mention its impact on other large organizations. We must refine/extend this theoretical construction in both breadth and depth, as well as explore its implications.

6. REFERENCES

[1] A. Begel, N. Nagappan, C. Poile, and L. Layman. Coordination in large-scale software teams. In *International Workshop on Cooperative and Human Aspects of Software Engineering*, pages 1–7, Washington, DC, USA, 2009. IEEE Computer Society.

[2] D. J. Berndt, J. L. Jones, and D. Finch. Milestone markets: Software cost estimation through market trading. In *Hawaii International Conference on System Sciences*, volume 9, page 239a, Washington, DC, USA, 2006. IEEE Computer Society.

[3] M. Cataldo, P. A. Wagstrom, J. D. Herbsleb, and K. M. Carley. Identification of coordination requirements: Implications for the design of collaboration and awareness tools. In *Conference on Computer Supported Cooperative Work*, pages 353–362, New York, NY, USA, 2006. ACM.

[4] J. Corbin and A. Strauss. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications, London, 3rd edition, 2007.

[5] B. Cowgill. Putting crowd wisdom to work. `http://googleblog.blogspot.com/2005/09/putting-crowd-wisdom-to-work.html`, 2005. Last accessed, February 2011.

[6] J. Dickerson. Improve product development using IPD. *Quality Progress*, 39(8), 2006.

[7] J. A. Espinosa, R. E. Kraut, S. A. Slaughter, J. F. Lerch, J. D. Herbsleb, and A. Mockus. Shared mental models, familiarity, and coordination: A multi-method study of distributed software teams. In *International Conference on Information Systems*, pages 425–433, 2002.

[8] R. Hanson, R. Oprea, and D. Porter. Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior & Organization*, 60(4):449–459, 2006.

[9] F. A. Hayek. The use of knowledge in society. *The American Economic Review*, 35(4):519–530, 1945.

[10] P. Hinds and C. McGrath. Structures that work: Social structure, work structure and coordination ease in geographically distributed teams. In *Conference on Computer Supported Cooperative Work*, pages 343–352, New York, NY, USA, 2006. ACM.

[11] J. L. Jones, R. W. Collins, and D. J. Berndt. Information markets: A research landscape. *Communications of the AIS*, 25(1):289–305, 2009.

[12] R. E. Kraut and L. A. Streeter. Coordination in software development. *Communications of the ACM*, 38(3):69–81, 1995.

[13] C. Poile, A. Begel, N. Nagappan, and L. Layman. Coordination in large-scale software development: Helpful and unhelpful behaviors. Technical report, Microsoft Research, 2009.

[14] C. Williams, P. Wagstrom, K. Ehrlich, D. Gabriel, T. Klinger, J. Martino, and P. Tarr. Supporting enterprise stakeholders in software projects. In *International Workshop on Cooperative and Human Aspects of Software Engineering*, pages 109–112, New York, NY, USA, 2010. ACM Press.

[15] J. Wolfers. Prediction markets at Google: A guest post. `http://freakonomics.blogs.nytimes.com/2008/01/14/prediction-markets-at-google-a-guest-post/`, 2008. Last accessed, February 2010.