Analysis and Representation of Field Research: Activity patterns in intellectual collaboration

Peter H. Jones, Redesign Research, Dayton, Ohio, USA peter@redesignresearch.com

Introduction

The following position paper briefly discusses analysis and representation approaches selected for a field study of scientific collaboration in a molecular biology research center. A hybrid framework was adopted for this study, incorporating field research approaches based on Rapid Ethnography (Millen, 2000), an analytical framework based on activity theory analysis (Engeström, 1999, Nardi, 1996), and developments of representations adapted from Engeström, Contextual Design, and prior field practice.

In CSCW and HCI research, several commonly-cited field data analysis methods recur in the literature for supporting software design and complex system requirements. Among the methods used in software and requirements we find scenarios (Carroll, 1995), stories (Erickson, 1995), use case modeling and UML (Jacobson, Booch, and Rumbaugh, 1999), and the well-documented models of Contextual Design (Holtzblatt and Beyer, 1993, 1997). However, for representing field data specifically identifying key aspects of collaboration, few models are available that lend themselves to visual representation. As in Contextual Design, I propose a position that multiple representations must be used to triangulate the issues of collaboration, supported by ethnographic details to describe the context of the real practices under analysis.

Significant philosophical issues remain among researchers regarding representation approaches and the validity of abstracting from ethnography. Researchers have held mixed views of the conceptual and field validity of representations. Throughout the CSCW literature, a dialogue has surfaced the question of representation and validity, ranging from the shift from ethnography to design (Hughes, Randall, and Shapiro, 1992) to appropriate social science methodology (Shapiro, 1994), to the politics of representation covered by a special issue of the CACM (Suchman, 1995, Bannon, 1995).

Background

The methods used by the field researcher determine the type of data collected – and the analysis methods applied to that data generate the products of the research. Whether approaching field research for the purposes of understanding (research) or design (innovation), our awareness of methods and their selection influences the final analysis.

Ethnographic methodologies have grown in popularity for studying collaborative activity. Representations of field data for the purpose of software design have been developed and critiqued in the literature, even as many social systems designers have proposed specific or hybrid methodologies intended to address the weaknesses of traditional representations. While traditional approaches have been criticized for being too cognitive (Shapiro, 1994), failing to capture the social organization of work (Suchman, 1995), and for inherently biasing the work practice by glossing over the situational details (Sachs, 1995).

These issues may affect us as *designers* working in a research discipline; given the practitioner's context within a design team, we may run the risk of treating ethnographic methodology as a field technique, when our field situations call for a professional sensitivity to these "academic" issues. For example, when conducting research in a business organization, our sponsor's issues may drive the interview protocols and the representation approaches. As Bannon (1995) notes:

> "Our critique of modeling argues for the essential limitations of any and every form of representation. Models are thus seen, in our view, as interpretations, as constructions, which for some purposes, under certain conditions, used by certain people, in certain situations may be found useful, not true or false."

He and Suchman hold that the underlying perspective taken by the researcher will be reflected in the representations of work activity – and these perspectives become hidden in the neatly summarized views or diagrams we use to document our fieldwork.

Bannon (1995) discusses a philosophical distinction about how work processes are represented in design. Significant controversy exists among practitioners about the methods used to describe work practices. Suchman (1995) summarizes this issue as, "the relation between normative accounts of how work gets done and *actual practice*." Bannon asks us to consider not whether one method is better than another, but to question whether *any techniques* in the current models could ever capture all that is required, and how the remainder left outside the representation is managed. He argues that work processes should "be conceived as heuristic devices that provide resources on particular occasions for particular groups of people."

The politics addressed in this perspective involve the ways in which system designers can choose certain representations to be used as formal work descriptions, and then leave others out, either through choice, avoidance, mis-analysis, or from just "not seeing."

In 1996, John Bowers urged CSCW researchers dealing with these issues to do the following:

1. (Encourage a) problem-driven, practical field with commitment to development. Allows involvement of practical social scientists.
2. Should continue to focus on social dynamics of computer work - it should remain radically empirical: "Go and see!"
3. Theories and conceptual frameworks from other disciplines should be regarded as resources, good ideas, useful constructs ("Go and see!")
4. "Host people" should make methodical and methodological contributions - Nike theory of CSCW, action-oriented approach.

I propose taking both Bowers and Bannon in stride. We should look at our current approaches to analysis and representation with an eye toward improving the state-of-the-art, while exploring the opportunity for transparency and self-reflection in fieldwork and design. I present my perspective on these issues. Then following with approaches and reflections on selecting appropriate methodologies for representation in a current field research effort.

Issues in Representation

Whenever we choose to abstract, aggregate, or even selectively illustrate from field findings to construct representations for communication, we incur certain risks:

- Failing to elicit findings of possible significance, because they don't lend themselves to our representation approach. (Like "Type II error" – a meaningful pattern exists, but our method was insensitive to it).
- Focusing on findings that may have been due to situational relationships, but by representing them in analysis, we charge them with a "false generality." (The error of representing what we find, but we may not have solid grounding for the selected representation).
- Letting research goals and hypotheses intrude, instead of letting the data tell the story.
- Pushing an abstracted view of the collaborative situation instead of a truer, but perhaps less communicable view.

In the practice of analysis and even more so in design, these issues of abstraction commonly elude our grasp. In qualitative analysis of field data (notes, video, interviews, protocols), one iterates through the verbal information to produce summaries, coding schemes, and responses to key inquiries. In building comprehensive scenarios of work activity in a specific domain, each level of further representation paints a clearer picture, but also removes us further from the work experience being modeled. With each crank of the analysis dial, we may introduce useful, but perhaps misleading abstractions based on theory. In the context of design goals we look for the data to support requirements for our developing design proposals and prototypes.

Based on this background, I raise four issues related to practice for exploration in the workshop:

1. What is the **purpose of the representation** in the first place?
   For illuminating issues from the field study for better understanding? For communicating the story of the work practice to designers or stakeholders? For developing meaningful models of practice?

2. How do we first **define the scope of collaborative activity** so that our analysis and representation stays true to its subject of discussion?

3. What **level of abstraction** do we select for "rolling up" field data to representations? To what extent do we support representations by selected field details?

4. Once we open the representation up to the social realm, where do we stop? **How do we focus the scope of analysis** when the *social* requires understanding (and perhaps representing) the team, workgroup, Community of Practice, extended community, organization?

Developing Methodology: From a Current Study

Purpose

The current project is the initial field study in a multi-phase field research program focusing on the research question: "how do scientists conduct research and seek and use information in their research context?" We further develop the following research purposes from this inquiry:

- Understand individual and social behavior of scientists information-seeking & use
- Create a framework for analyzing new findings within this context
- To uncover discoveries useful for design and extending the current body of literature

A key focus of the study is to develop and initial understanding of the drivers of *collaboration* and the work of scientists on defined research *projects*.

A week-long field visit was sponsored by the RNA Center for Molecular Biology at Case-Western University. A rapid ethnography field approach was used, consisting of semi-structured interviews and with five principal investigators and interviews and observations with 11 researchers. Field data was collected using videotaped interviews, notes, artifact notation and collection, and interactive walkthroughs of information activities. Collaboration data was gathered through videotaped interviews, group discussions, observations within the lab environment, and information task diaries used by the researchers.

Defining the Scope of Collaborative Activity

A literature review (Jones, 2002) of information-seeking in the scientific research environment shows most research focusing on individual and cognitive models and studies. Across a ten-year period, a core group of 25 frequently-cited studies were found defining this research focus. Only about four of these were based on ethnographic field methods; most were framed as individual cognitive studies, even though different disciplines and roles in information-seeking were characterized. Of these studies, three focused on *interdisciplinary* research activity, considered a significant trend in the sociology of science.

Focusing on the context of activity, categories were defined for the *unit of analysis* of the studies. I derived four categories based on context:

- Scientific Discipline
- Institution or Department
- Research Program or Project
- Individual

I further organized the research by Cognitive/Analytical vs. Social/Collaborative to show how many focused on individual behavior or the social processes of research. This model is represented in Table 1. Most of the core studies behind this model focused on the *Individual* as information-seeker, and most presented theoretical models of individual factors derived from research or empirical study. Of the 25 studies, 13 fit this category. Three of these also fit the second largest category, that of Discipline, analyzing factors unique to a scientific discipline. A fourth fit both Discipline and *Institution*, generalizing its findings to disciplinary distinctions and not individual. *Research Program or Project* was the context with least representation of studies, with only one, a review of *methods*, with no empirical data. (Articles and review available on request.)

Context	***Cognitive / Analytical***	***Social / Collaborative***
1. Individual	Models of individual behavior for information seeking, use, and cognitive processing.	Individual use of information in interaction and collaboration with other individuals.
2. Research Program or Project	Individuals seeking and using information as part of specific research project goals.	Individuals in small groups working together on goal-driven information seeking and use tasks (research).
3. Institution or Department	Faculty, students research contributions to broader programs in department or university.	Faculty and students working in collaborations toward internal programs or department projects.
4. Scientific Discipline	Research within the context of publishing or theoretical contributions to discipline.	Collaborative projects, workshops, panels, etc. specifically aimed at discipline development.

Table 1. Hierarchical model of research activity.

Four gaps in the current research were discovered, which we are attempting to address with the molecular biology research center study:

- Describing the information behavior of *research teams*, from the level of research project.
- *Motivators* of scientists' information behavior, especially understanding the social motivations in research collaboration that lead to information use.
- *Cycles* of information -intensive work, describing the temporal events and activities characterizing research projects.
- *Predictors* of information use, attempting to predict types of information activities based on field research and analysis of research projects, motivators, and temporal cycles.

Level of Abstraction

Focusing on the appropriate unit of analysis is a critical issue for fieldwork, analysis, and representation. The traditional design of information seeking studies assumes an individual level of focus, and representations derive from this focus. However, studies cite trends in scientific research toward collaboration, interdisciplinarity, and large research projects. These trends indicate that research projects and collaboration should motivate and extend information seeking and use beyond

the individual. Therefore, the scope for the study was defined to focus on the social use of information in scientific research collaboration. The unit of analysis was the research project, a small group of scientists consisting of typically (in this research lab) 4-7 researchers.

To analyze three of the four dimensions of information behavior, an activity theory approach was adopted. Activity theory (Cole and Engeström, 1991, Nardi, 1996) offers an analytical approach to distributed cognition, based on understanding actual work activity as the unit of analysis. Engeström's framework for modeling activity systems enables interpretation of individuals or multiple participants in social activity, from macro-level research projects to micro-level individual research tasks, while preserving the social context across analysis and representation.

Activity theory develops from the position that human behavior can be represented as activity, including cognition and collective activity. The essential concepts are summarized by considering individual and social activity as "oriented towards the world, they include both internal and external components, they are hierarchically organized, mediated, and developing." Fundamental to the theory is the hierarchical organization of activity, expressed in the levels of activity, action, and operation. **Activity** can be defined as the overall object or motive of organized behavior, essentially composed of a behaving subject, an object of activity, actions, and operations. **Actions** are consciously performed, goal-directed processes undertaken to fulfill the activity. **Operations** are non-conscious, automatically performed behaviors comprising an action. These three levels interrelate within an environment, as shown in Table 2:

Context	***Activity Level***	***Type of Analysis***
Scientific Discipline	**Environment**, largest boundary of analysis	Researcher activity or info seeking that contributes to discipline
Institution or Department	**Context**, immediate boundary of analysis	Activity and info seeking involved in university or department leadership
Research Program or Project	**Activity**, unit of analysis, focus on work practices and joint actions	Activity and info seeking involved in research projects with specific goals
Individual	**Action**, specific tasks as part of ongoing activity	**Individual actions** to accomplish tasks relating to any level (Dunbar, Covi) Analysis of micro-level **operations** required of actions (Xie, Wilson)

Table 2. Hierarchical model of research activity.

When analyzing at the activity level, for example, the activity of final preparation of a research paper for publication, we can focus on the lifecycle of tasks making up that activity, and analyze the different individual behaviors involved. It's not enough to use a general model that suggests the scientist is "identifying" information they require for filling in a paragraph or "locating" references for an argument. We must evaluate the whole picture of activity as a meaningful set of related tasks from the perspective of the researcher. Then we can understand the rationale of individual behaviors within the context of a goal-directed process.

Focusing the Scope of Analysis

Very little research has focused on the information behavior of the scientist and their role within a specific research program. Although all scientist studies imply the conduct of research, they often describe ongoing research activity, and do not describe the activities within specific research projects. No ethnographic studies are found focusing on scientist interaction with information in pursuit of specific goals in the context of a bounded project, with specific objectives such as an

experiment or article. These goals and artifacts of the research world are important to understanding the context of information seeking, and are insufficiently represented. In the sociology of science, Latour's (1986) *Laboratory Life*, and Lynch and Woolgar's (1990) *Representation in Scientific Practice* offer comprehensive ethnographic studies of research activity. However, these rich, real world accounts show little focus on information seeking and use, and do not afford analysis of information seeking within the current context of available online information resources. For the most part this literature in the sociology of science deconstructs the practices of research to show how scientific work is conducted like any other job or work practice.

Engeström's Activity Theory model offers a means of specifying the goals and influences on individual activity within a meaningful context of use.

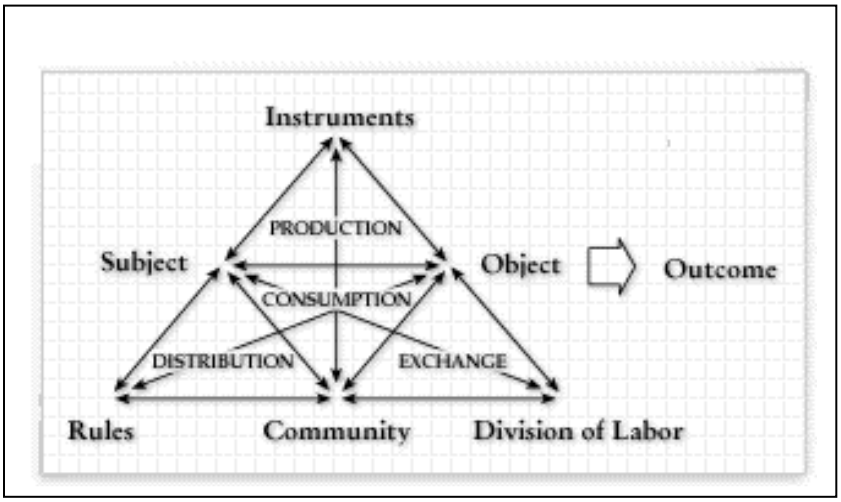

Figure 1. Engeström's model of activity.

The three basic elements of activity require a *subject* (researcher) interacting with an *instrument* (tools used in practice) to mediate an *outcome* or goal. Engeström's model extends the concepts of activity theory by integrating three dimensions that specifically serve organizational analysis. These include community, rules, and division of labor.

Engeström's model describes the six dimensions portrayed in Figure 1. The model supports description of coordinated social activity based on relationships among the dimensions. This descriptive approach was taken with the data from the organizational analysis in the research findings and analyses, in which claims were analyzed for interactions using the model. The six dimensions are defined as follows:

Subject – an individual or group whose point of view is represented in the activity.

Object-outcome – The objective or problem space which defines the scope of directed activity, which transforms through instrumental action into a designed or mediated outcome.

Instrument – Any tool, artifact, process, device, or thought system enabling action toward the space of the outcome.

Rules – Shared understandings and conventions, values, explicit and implicit regulations, and norms that enable interaction and constrain behavior.

Community – A specific, distinct collection of individuals or groups oriented toward the general object of activity.

Division of Labor (**Work Structure**) - The division of tasks horizontally among community members and vertically in distinctions of power and status.

A proposed representation comprises two visual components for each level of activity. The Engeström model represents the continuous and specific mediated tasks performed within a given context. An activity model can be developed for the context of Scientific Discipline, Institution or Department, and Research Program or Project. The Individual level can represent a focus on activity or *actions*.

Engeström (1996) points out how any complex activity requires a dynamic approach to analysis using the model. As an activity system is analyzed over time, constituents of the model often change positions based on the perspective of the subject. "What initially appears as object may soon be transformed into an outcome, then turned into an instrument, and perhaps later into a rule." Therefore, any static representation of activity using the model must be seen as a snapshot in time, as a collection of qualities in relationship to one another, as opposed to occupants of categorical positions.

> "An activity system does not exist in a vacuum. It interacts with a network of other activity systems. For example, it receives rules and instruments from certain activity systems (e.g., management), and produces outcomes for certain other activity systems (e.g., clients). Thus, influences from outside 'intrude' into the activity systems."

Holland and Reeves (1996) clarify the influence of collective perceptions in project activity, describing how jointly composed interpretations develop through cooperative action. These perceptions become powerful motivators for collective activity, and establish the basis for explanations of team behavior. In this way, collective perceptions influence the outcomes of projects and products.

> "For one, teams symbolize themselves and the project. To a degree they discursively construct, with the type of semiotic mediating devices emphasized by Vygotsky, representations of their stance toward their projects. Use of these mediating devices solidifies their perspective and orients their subsequent work.." (1996, p. 270)

These concepts comprise a foundation for representation that supports an integrated analysis of field findings for coordinated and unstructured social activity in context of the scientists' research program.

Proposed Representations: Activity Model

An example activity model for Research Project/Program identifies factors for a range of complex information situations. (The collected data has not been analyzed as of yet). Although this example was necessarily abstracted, a specific example from field data will be available for presentation in the workshop. The activity model shows Subject as being one of several types of researchers that may participate in a program: the Principal Investigator (PI), a professor/researcher, a lead researcher (assistant professor), and doctoral/graduate student researchers. We could trace activity for any one of these subjects, each with a different information behavior "trajectory."

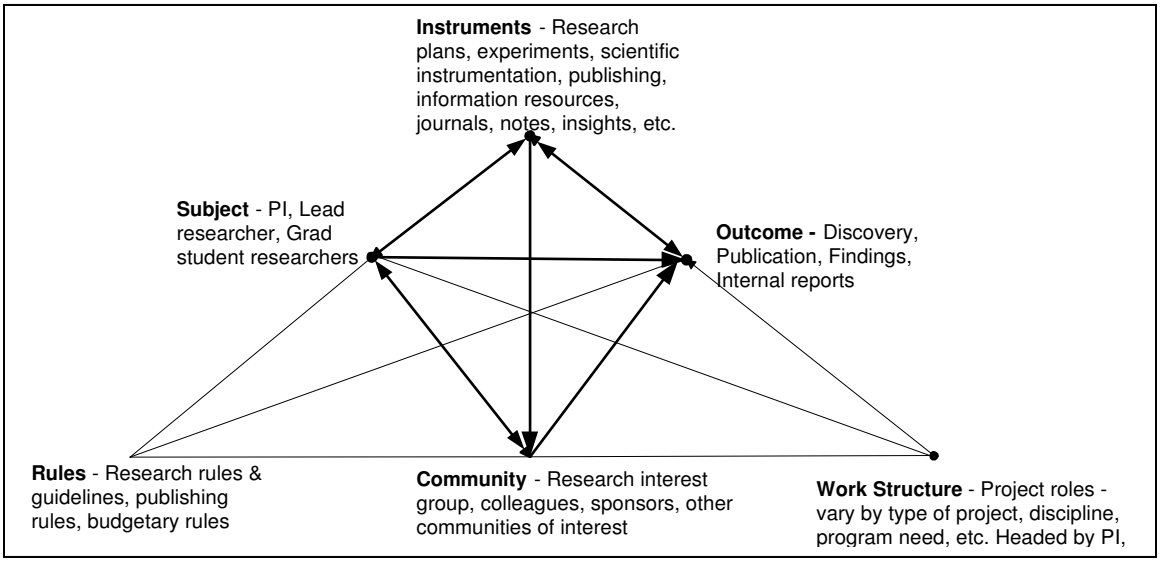

Figure 3. Activity Model at Research Project Level.

This initial model shows how work is oriented toward a discovery or publishing – project outcomes shared by all team members, and even the larger research community. The model indicates frequent and influential collaboration with the research community (bi-directional vectors), and the influence of this community on the project outcome. Within the project, both rules and work structure play a lesser role – activities on a project may be bounded by some role restrictions. In research, assignments and information tasks are dynamic, determined by necessity in the course of research, and not parceled out in advance by Principal Investigators.

The instruments mediating the project objectives include all the information sources, facilities, labs, and research tools involved in the research. In fact, in a project setting any information source can become a mediating tool, such as insights, notes, or discussions. And these sources of information and insight are often shared with the community as from the subject, indicated by the strong vector from Instrument to Community (which should be bi-directional also).

In the literature, most studies describe the tools of research as mediating *individual* actions; when scientists browse journals or conduct searches, it is considered an individual behavior, which it can be. But when analyzing activity, the individual actions of information seeking derive their motivation from motivations and requirements of the group, of the research project itself. This group level of analysis allows us to inquire about the coordination of team resources, and to better locate differing roles on projects. The likelihood of a PI conducting their own literature searches within a project context is low – this role is found fulfilled by graduate students, who also check references, read new papers, and summarize findings for quick discussion and analysis by the project team.

Research project activity also engages the use of tools from the Community, not just between key "subjects" of the research team. The community represents the network of support for the project, including librarians, lab directors and assistants, professors from other departments, etc.

Information Lifecycle

The information lifecycle of the project also cannot be generalized too far without case data, so a caveat is appropriate for interpretation. This initial model indicates a "typical" lifecycle of activities for an investigatory research program, involving a planned intervention such as an experiment or field study, and resulting in the publication of findings. Because the lifecycle covers so many aspects of the research project, it is worth presenting as a framework for validation and further research.

The milestones indicated for the lifecycle show a series of typical points that define a mid to large project plan for the research effort (very large programs would show a much different lifecycle). We show this program starting with an initial research paper, as opposed to the more proposal. All proposals stem from promising research ideas, and for larger projects, a conference paper or working paper may be reviewed to generate interest and further develop the issues. Depending on the discipline, this period may range from a few weeks to a year (physics shorter, social sciences, longer). Emerging interdisciplinary fields may take even longer, if the project crosses normal departmental boundaries.

As the detail indicates, an initial paper identifies issues that future research might explore, and presents a possible focus for future work. This is also a typical bridge between dissertation research and post-doctoral or larger academic research. The information use spikes as one or multiple individuals contribute to developing the issues, and at some point an exhaustive review of related research is conducted prior to or as part of the research project proposal. We might position the proposal itself at the peak of this process, at which an administrative phase takes the project into consideration, and information use drops off. Of course, from a designer's point of view, we might envision this identification, ideation, review, and proposal process as a unique set of information activities, differing from the activities involved in paper publishing.

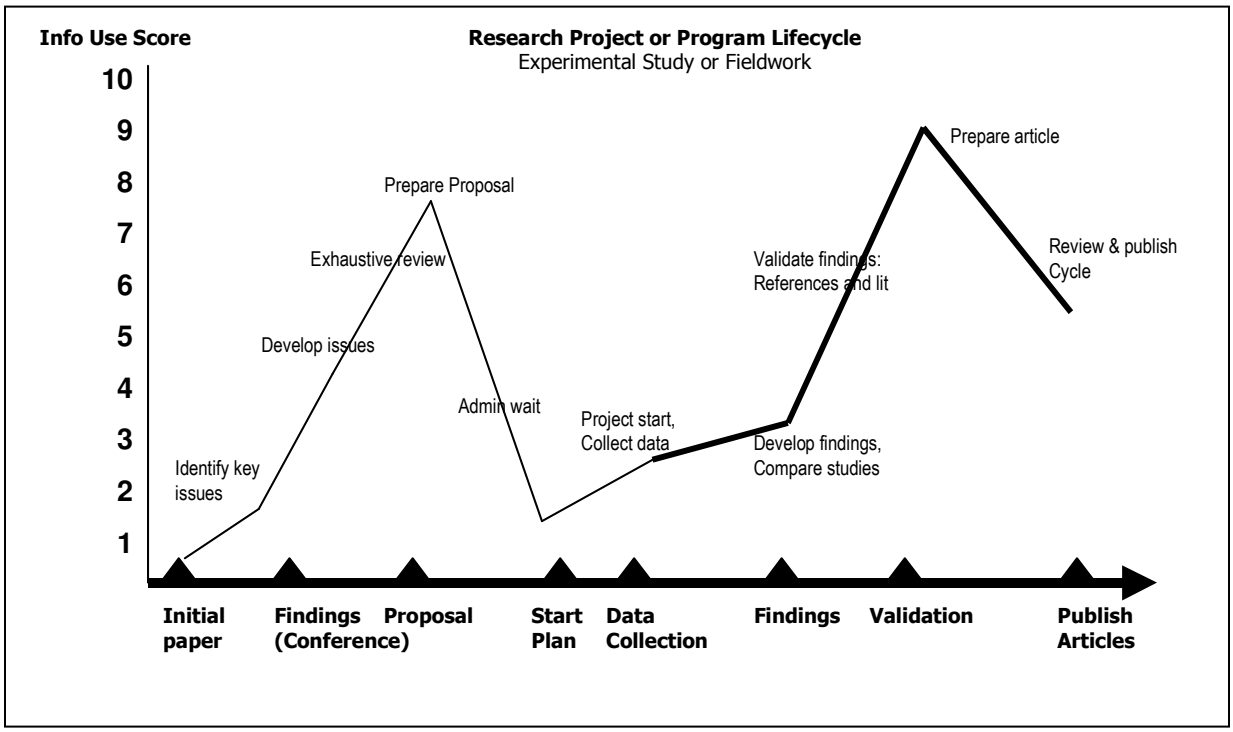

Figure 4. Information Cycle of Research Project.

Final Points

The goal of presenting this initial model in such detail is to present some of the known limitations in field research and analysis methods in research in scientific research collaboration. I also propose an adaptation of Engeström's activity model used as a representation in concert with analysis of field data. I further propose the information cycle as a model of temporal behavior to show both continuity of analyzed behaviors and the specified points of change along a real cycle of work.

I present these representation approaches in the spirit of both Bowers and Bannon. Bowers charges us to "go and see" how we might use effective representations in a pragmatic and conscious application to our field research. Bannon asks us to refrain from abstracting beyond the path visible by the light of our data. While I've taken more than my share of space in defining this research program, I should also state we have integrated representations from Contextual Design into the final research instruments and analyses. We have selected the Context Model, Information Flow Model, Artifact Model, and Physical Model as maps to guide assemblage of the field research data in the analysis task. We wish to report both our findings and our experience of the effectiveness of this hybrid methodology in future reports.

References

1. Bannon, L.J. (1995). The politics of design: representing work. *Communications of the ACM*, 38 (9), pp. 66-68.
2. Carroll, J. M. (1995). *Scenario-based design: Envisioning work and technology in system development*. New York: John Wiley & Sons.
3. Cole, M. and Engeström, Y. (1991). A cultural historical approach to distributed cognition. In G. Salamon (Ed.), *Distributed cognition* (pp. 1-47). Cambridge: Cambridge University Press.
4. Engeström, Y. (1999). Activity theory and transformation. In Y. Engeström, R. Miettinen, and R-L. Punamaki (Eds.), Perspectives on activity theory, (pp. 19-38). Cambridge, U.K.: Cambridge University Press.

5. Erickson, T. (1995). Notes on design practice: Stories and prototypes as catalysts. In *Scenario-Based Design*, John M. Carroll (ed.). New York: John Wiley & Sons.
6. Holland, D. and Reeves, J.R. (1996). Activity theory and the view from somewhere. In B. Nardi (Ed.), Context and consciousness: Activity theory and human-computer interaction (pp. 257-281). Cambridge, MA: MIT Press.
7. Holtzblatt, K. and Beyer, H. (1993). Making customer-centered design work for teams. *Communications of the ACM*. *3*6(10): 92-103.
8. Holtzblatt, K. and Beyer, H. (1997). *Contextual design: A customer-centered approach to systems design*. New York: Morgan Kaufmann Publishers.
9. Hughes, J.A., Randall, D., and Shapiro, D. (1992). Faltering from ethnography to design. *In Proceedings of CSCW 92*. Toronto, November, pp. 115–122.
10. Jacobson, I., Booch, G., Rumbaugh, J. (1999). The unified software development process. Boston: Addison-Wesley.
11. Kyng, Morten. (1995). Making representations work. *Communications of the ACM*. 38(9): 46-55.
12. Latour, B. and Woolgar, S. (1986). *Laboratory life: The social construction of scientific facts*. London: Sage.
13. Latour, B. (1987). *Science in action*. Cambridge, MA: MIT Press.
14. Lynch, M. and Woolgar, S. (1990). *Representation in scientific practice*. Cambridge, MA: MIT Press.
15. Millen, D. (2000). Rapid Ethnography: Time deepening strategies for HCI field research. *Proceedings of DIS 2000*, 280-286. New York: ACM.
16. Nardi, B.A. (1996). Studying context: A comparison of activity theory, situated action models, and distributed cognition. In B. Nardi, (Ed.), Context and consciousness: Activity theory and human-computer interaction, (pp. 69-102). Cambridge, MA: MIT Press.
17. Nardi, B.A. and O'Day, V.L. 1999. *Information ecologies: Using technology with heart*. Cambridge, MA: MIT Press.
18. Sachs, P. (1995). Transforming work: Collaboration, learning, and design. *Communications of the ACM, 38* (9), 36-45.
19. Shapiro, D. (1994). The limits of ethnography: Combining social science for CSCW. *In Proceedings of CSCW 94*. Chapel Hill, NC, Oct. 22–26, pp. 417-428.
20. Spinuzzi, C. (2000). The ecologies of technologically mediated work: three perspectives: Investigating the technology-work relationship: a critical comparison of three qualitative field methods. *Proceedings of IEEE professional communication society international professional communication conference*.
21. Star, S.L. (1995). The politics of formal representations: Wizards, gurus, and organizational complexity. In S.L. Star (ed.) *Ecologies of knowledge: Work and politics in science and technology*, pp. 88-118. Albany, NY: State University of New York.
22. Star, S.L. and Ruhleder, K. 1994. Steps toward an ecology of infrastructure. In *Proceedings of CSCW 94*. Chapel Hill, NC, Oct. 22–26, pp. 253–264.
23. Suchman, L. (1995). Representations of work. *Communications of the ACM*, September, 38 (9).
24. Woolgar, S. 1995. Representation, cognition, and self: What hope for integration of psychology and sociology?. In S.L. Star (ed.) *Ecologies of knowledge: Work and politics in science and technology*, pp. 154-179. Albany, NY: State University of New York.