Report on Methods for Studying Communities of Practice

Edward Lorenz* and Pierre-Jean Barlatier**

EU Network of Excellence
Dynamics of Institutions and Markets in Europe (DIME)

July 2007

* (Lorenz@gredeg.cnrs.fr) DEMOS, GREDEG UMR 6227 CNRS / University of Nice-Sophia Antipolis
** (Barlatier@gredeg.cnrs.fr) RODIGE, GREDEG UMR 6227 CNRS / University of Nice-Sophia Antipolis.
1. Introduction

The notion of a ‘community of practice’ grew out of the work on the situated nature of learning by John Seely Brown, Paul Duguid, Jean Lave, Lucy Suchman and their associates at the Palo Alto Institute for Research on Learning in the 1980s. In published form, the concept was first introduced by Lave and Wenger (1991) in their book, Situated Learning: Legitimate Peripheral Participation. As Amin and Roberts (2006) observe in their recent review of the literature on CoPs, since this seminal publication there has been a considerable increase in the popularity of the concept in academic discourse, and starting in 1997 one can observe a year-to-year increase in the number of publications focusing on CoPs.

The original brief in the call for contributions to DIME RA1 to which this report responds is, “to compare and contrast qualitative and quantitative approaches to studying CoPs, focussing on what each approach can and cannot capture”. In keeping with the objective of identifying what different methods can and cannot capture, the report is based on a review of empirical studies and investigations of CoPs, or studies which make use of computer simulation methods in order to develop and analyse empirical-like findings on CoPs. More specifically, the following types of studies or investigations are considered: ethnographic studies, case studies based on structured or semi-structured interviews, empirical investigations based on the use of survey data, and studies using computational methods. The report does not address the results of purely theoretical or conceptual papers, or papers whose empirical observations are derived exclusively from references to the original empirical findings of others.

It is worthwhile clarifying the objectives of this report in two further respects. First, its purpose is not to provide a comprehensive overview of the empirical literature, identifying the methodological strengths and weaknesses of individual publications. This would go beyond its scope. Rather, the report sets out to provide an assessment of what different types of approaches to studying CoPs can and cannot capture, using illustrative studies to back-up the conclusions. Second, the report does not aim to be a ‘Manual’ identifying ‘best practice’ techniques for undertaking empirical investigations of CoPs in the spirit, say, of the various OECD manuals for producing science and technology indicators. This, also, would clearly go beyond its scope. Nonetheless, an attempt is made to identify some of the key problems that would have to be solved before such a manual might be drafted.
Section 2 below briefly discusses how the concept of CoP grew out of a larger research programme focussing on the situated nature of knowledge and learning. Section 3 discusses the problem of a lack of cumulativeness in empirical research on CoPs and identifies appropriate empirical methods in relation to a taxonomy of dimensions of CoPs. Sections 4 through 7 discuss, respectively, the strengths and weaknesses of the following methods of study: ethnographic, structured or semi-structured interviews, survey data, and computer simulations. Methods appropriate to the study of virtual communities are discussed in the section on the use of survey data. Section 8 concludes.

2. The origins of the notion of CoP in research on situated practice and learning

Elsewhere, Lorenz (2003) has argued that the notion of CoP, with its emphasis on the situated nature of practice and learning, can be seen as offering a radical alternative to the symbolic or information-processing view of human knowledge and learning, associated notably with the work of Newell and Simon (1976). It is worth calling attention to the fact that the original work on CoPs by Lave, Duguid, Brown and their colleagues at the Palo Alto Institute for Research on Learning grew out of a larger programme of research on the situated nature of knowledge and learning involving researchers at the UC-San Diego Laboratory on Comparative Human Cognition (LCHC) such as Michael Cole, Yrjö Engeström, and Edwin Hutchins. In developing the view that knowledge and learning is situated relative to a local context, this community of researchers for the most part by-passed the question of the internal architecture of human cognition. In particular, they did not seek to challenge the mainstream view by mustering the psychological or neurobiological evidence that might support an alternative description of internal cognitive architecture, in the sense that Newell and Simon used the term. Rather, they challenged the mainstream perspective by providing field evidence on the ‘everyday’ nature of human learning and problem-solving activities, thus providing empirical support for the view that a person’s cognitive capacities depend on his or her interactions with the external environment.

For example, Lave, Murtaugh, and de la Rocha (1984) find that arithmetic procedures used on the job or in one’s daily activities, such as in shopping, are structured and aided by features

---

1 In discussion with Edward Lorenz at the 27-28 October DIME workshop on CoPs at the University of Durham, Jean Lave confirmed that one of her interests in developing the notion of CoPs was its scope for providing an alternative to the physical symbol system approach to human cognition.

2 For an overview of the LCHC research programme in the 1980s, see Cole, Engeström and Vasquez (1998).

3 Also see, Lave (1981) and (1988).
of the physical setting in which they occur. These skills transfer poorly or not at all onto different contextual settings, such as the classroom. Suchman’s (1987) ethnographic account of the difficulties novice users faced in their interactions with a computer based help system attached to a complex photocopier shows how language and words acquire meaning or are “indexicalised” by the shared context in which they are used. In this sense one can say that the shared environment contributes to the intelligibility of conversation.  

These same premises about how a person’s cognitive capacities are linked to his or her interactions with a local environment and about how language acquires meaning through practice and interaction can be seen in the defining characteristics of CoPs as they were set out in the seminal contributions. Thus Lave and Wenger (1991, p. 98) state, “A community of practice is a set of relations among, activity and world, over time and in relation with other tangential and overlapping communities of practice.” And, “learning to become a legitimate participant in a community involves learning how to talk (and be silent) in the manner of full participants” (Lave and Wenger, 1991, p. 105). Brown and Duguid (1991), who draw inspiration in part from Orr’s (1996) ethnographic study of the community formed by Xerox photocopy machine repairers, state, “Not only is learning in this case inseparable from working, but also individual learning is inseparable from collective learning. The insight accumulated is not a private substance, but socially constructed and distributed” (p. 45). And, “From this (social construction) perspective, learners can in one way or another be seen to construct their understanding out of a wide range of materials that include ambient social and physical circumstances and the histories and social relations of the people involved” (Brown and Duguid, 1991, p. 47).

3. Increasing definitional diversity and lack of cumulativeness in subsequent empirical research on CoPs

While the seminal publications are unified through their focus on the way people learn through their shared and situated practice, they nonetheless are silent regarding limits on what kinds of entities might usefully be analysed as CoPs. Now it is true that Duguid and Brown (1991) appear to be mainly concerned with organisational behaviour and with using the CoP framework as a means of analysing the way largely unrecognised non-canonical practice is structured in relation to formal or canonical practice. When we turn to Lave and Wenger

---

4 For an overview of this literature published just prior to Brown and Duguid’s (1991) seminal piece on CoPs, see Brown, Collins and Duguid (1989).
(1991), however, the notion of a CoP is used for gaining understanding into processes of apprenticeship and learning in a diverse array of groups including midwives, tailors, naval quartermasters and non-drinking alcoholics. Although their focus is on the locally situated nature of learning, the relations of the members of some of these communities (e.g. naval quartermasters) are clearly structured by larger sets of formal institutions which play a role in the cultural transmission of practice and knowledge.\(^5\)

Subsequently, an increasing diversity can be observed in the CoP literature. The kinds of things that have been analysed as CoPs include business groups or networks set up by management (Gongla and Rizzuto, 2001), web design project teams (Thompson (2005), primary care clinicians (Gabbay and le May (2004) networks of user-innovators held together by a common enthusiasm for particular products and activities (Franke and Shah, 2003), and virtual communities or networks such as open source software communities (Dahlander and Wallin, 2006). As Amin and Roberts (2006, p. 2) have observed, the diversity is such that, “the blanket use of the term risks weakening the original conceptualisation of CoPs as learning and knowledge generating entities situated in certain kinds of practice.”

Our view is that while this diversity is not bad in itself, it has contributed to a lack of cumulativeness in empirical research on CoPs, since it is often difficult to know whether different researchers are indeed studying the same thing. This state of affairs in applied research on CoPs makes it very difficult for a researcher to build upon a solid foundation of empirical findings and results. Contradictory findings may simply reflect the fact that different kinds of communities are being investigated.

However, we do not believe that a desirable response to this lack of cumulativeness would be for the research community to seek consensus around a rigid set of criteria for identifying CoPs, with anything not conforming to it being excluded from the researcher’s gambit. For one thing, CoPs themselves are evolving, at least in part due to the impact of new information technology, and insisting on a rigid set of criteria might well shut off fruitful exploratory research into these changes. Rather, what we think is required at this phase of empirical research on CoPs is greater clarity concerning their underlying dimensions and in particular greater clarity regarding the cognitive foundations or knowledge characteristics of CoPs.

---

\(^5\) This is a key theme in the research of Hutchins (1995), upon which the naval quartermasters example is based.
Clarity around the underlying dimensions of CoPs would not only respond to the problem of lack of cumulativeness in research, but also would provide useful methodological guidance for the researcher. This is because the different methods available for studying CoPs are more or less suitable for capturing different dimensions. In what follows we draw upon the taxonomy of CoPs proposed by Amin and Roberts (2006) in order to identify the dimensions of CoPs that are relevant for crafting appropriate empirical methodologies. Amin and Roberts identify four basic types of CoPs: craft/task based, professional, expert/creative, and virtual. These different types of CoPs differ according to the following dimensions:

1) Cognitive foundations/knowledge types. CoPs can be distinguished according to the relation of tacit to codified knowledge as well as the aesthetic and kinaesthetic dimensions of knowledge. For example, kinaesthetic and embodied knowledge play a privileged role in craft communities, while codified professional standards play a central role in many professional communities.

2) Physical and cognitive artefacts. This is closely linked to the first dimension. Knowledge may be embedded within machines and tools in different ways and cognitive artefacts, including written text, may play different roles across different types of CoPs.

3) Formality vs. informality. For example, craft/task based CoPs may be more or less formal in terms of the entry criteria and training content of apprenticeship.

4) Proximity relations: CoPs differ in terms of the importance of face to face interaction and communication versus distanciated exchange and communication. This is connected to the question of the boundaries of CoPs. Distanciated communication is of course a central feature of virtual communities while face-to-face communication is central in the transmission of knowledge with in craft communities.

5) Social ties. As we have observed above, CoPs may be embedded within larger institutional structures which bear on the transmission of knowledge and practice across time and space.

The following table briefly indicates the strengths and weaknesses of different methodologies for capturing these different dimensions of CoPs.
Table 1
Strengths and weaknesses of different methods for studying CoPs

<table>
<thead>
<tr>
<th>Method</th>
<th>Dimensions of CoPs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method</strong></td>
<td><strong>Cognitive/knowledge dimension</strong></td>
</tr>
<tr>
<td>Ethnographic studies</td>
<td>Tacit vs. codified knowledge. Aesthetic/kinaesthetic Embodied knowledge.</td>
</tr>
<tr>
<td>Case studies based on interviews</td>
<td>Less suitable for characterising tacit/embodied knowledge. Suitable for capturing codified/articulated knowledge.</td>
</tr>
<tr>
<td></td>
<td>Artefactual dimension: Role of physical and cognitive artefacts in mediating knowledge exchange and information flows in distributed tasks.</td>
</tr>
<tr>
<td></td>
<td>Proximity dimension: Importance of face to face in relation to distanciated communication. Boundaries of CoPs.</td>
</tr>
<tr>
<td></td>
<td>Social/institutional dimension: Interpersonal ties. Shared practice. Institutionalised transmission of knowledge and practice over time and space.</td>
</tr>
<tr>
<td></td>
<td>Organisational dimension: Relation of informal to formal structure. Relation of community to organisational hierarchy.</td>
</tr>
</tbody>
</table>

Ethnographic studies: Especially suitable for rich characterisations of tacit and embodied knowledge.

Case studies based on interviews: Less suitable for characterising tacit/embodied knowledge. Suitable for capturing codified/articulated knowledge.

Suitable for providing rich characterisations of face-to-face communication. Less suitable for characterising distanciated exchanged and communication.

Suitable for providing rich characterisations of interpersonal ties. Less suitable (or at least a relatively costly method) for acquiring first-hand information about larger institutional arrangements.

Suitable for capturing the role of informal practice in relation to formal practice. Less suitable (or at least a relatively costly method) for characterisation the evolution of community structure.

Suitable for providing a rich characterisation of informal practice. Suitable for charactering the formal organisational structure and its evolution over time.
### Table 1 (cont’d)

**Strengths and weaknesses of different methods for studying CoPs**

<table>
<thead>
<tr>
<th>Method</th>
<th>Dimensions of CoPs</th>
</tr>
</thead>
</table>
| **Method**                                     | **Cognitive/knowledge dimension**  
Tacit vs. codified knowledge. Aesthetic/kinaesthetic Embodied knowledge. | **Artefactual dimension**  
Role of physical and cognitive artefacts in mediating knowledge exchange and information flows in distributed tasks. | **Proximity dimension**  
Importance of face to face in relation to distanciated communication. Boundaries of CoPs. | **Social/institutional dimension**  
Interpersonal ties. Shared practice. Institutionalised transmission of knowledge and practice over time and space. | **Organisational dimension**  
Relation of informal to formal structure  
Relation of community to organisational hierarchy. |
| **Exploratory and confirmatory statistical methods using survey data or internet-based data sources** | Unsuitable for capturing the nature and role of tacit/embodied knowledge. Suitable for capturing simplified and representative descriptions of codified/articulated knowledge. | Unsuitable for capturing the role of mediating physical artefacts. Suitable for capturing simplified and representative descriptions of cognitive artefacts. | Suitable for providing simplified and representative characterisations of both face-to-face and distanciated communication and exchange. | Suitable for providing simplified and representative descriptions of the role of interpersonal ties and institutional arrangements. | Suitable for providing simplified and representative information formal and informal practice. |
| **Computational modelling**                    | Suitable for characterising the evolution of codified representations of knowledge. | Suitable for characterising the evolution of internal cognitive artefacts (e.g. mental models). | Suitable for capturing the emergence of communities within larger networks of actors. Suitable for characterising the evolution of the boundaries of communities. | Suitable for modelling the emergence and evolution of interpersonal ties. | Suitable for modelling the emergence and evolution of structure within communities. |
4. Ethnographic approaches: a privileged methodology for studying COPs?

In the seminal research on CoPs the privileged use of ethnographic methods was not a theoretically neutral methodological choice. Rather, as the commentary in Table 1 above suggests, this choice was based on the premise that such methods were suited to capturing the locally situated nature of learning including the way social arrangements and artefacts serve to mediate processes of knowledge exchange and cooperation in the performance of distributed tasks. Indeed, it is difficult to see how many of the key insights of the seminal contributions could have been arrived at or supported in the absence of the use of ethnographic methods.

Consider, for example, Orr’s (1996) classic study of field technicians responsible for repairing photocopiers. One of the key insights coming out of Orr’s study concerns the role of narration in daily practices of knowledge transmission and problem-solving. As he explains, the circulation of stories among technicians is a principal means through which they stay abreast of the knowledge about the subtleties of their machines. Thus, the process of recounting stories can serve to mobilise the collective knowledge of the community and provide essential insights for problem-solving activities. Orr arrived at these insights about the role of narration though accompanying technicians on field visits, attending their meetings and eating lunch with them. This allowed him to gain something of an insider’s understanding of work within this community. It is difficult to see how Orr could have provided a useful characterisation of the role of narration in the absence of direct observation and participation in the activities of the repair technicians.

In the same vein, consider the set of indicators that Wenger (1998) lists for identifying CoPs in his influential study of the claim processing employees of a health insurance company (see Table 2 below).
Table 2
Key indicators of CoPs (Wenger 1998, pp. 125-26)

- Sustained mutual relationships — harmonious or conflictual
- Shared ways of engaging in doing things together
- The rapid flow of information and propagation of innovation
- Absence of introductory preambles, as if conversations and interactions were merely the continuation of an ongoing process
- Very quick setup of a problem to be discussed
- Substantial overlap in participants’ descriptions of who belongs
- Knowing what others know, what they can do, and how they can contribute to an enterprise
- Mutually defining identities
- The ability to assess the appropriateness of actions and products
- Specific tools, representations, and other artefacts
- Local lore, shared stories, inside jokes, knowing laughter
- Jargon and shortcuts to communication as well as the ease of producing new ones
- Certain styles recognised as displaying membership
- A shared discourse reflecting a certain perspective on the world

Some of the indicators listed in Table 2 refer to spontaneous behaviour based on shared understandings that remain tacit or unarticulated by the members of the CoPs. This includes the absence of preliminary preambles, the use of shared stories, the recourse to inside jokes and knowing laughter, and certain styles recognised as displaying membership. As Wenger observes, these are relevant indicators because a community of practice need not be reified or explicitly articulated in the consciousness of its members. As in the case of Orr’s ethnographic study, it is difficult to see how Wenger might have provided useful characterisations of the nature and role of these behaviours in the absence of direct observation of the daily work of the claim processors.⁶

While the assertion that ethnographic methods are more suited than others to capturing the locally contextualised nature of practice based on tacit understandings is probably uncontroversial, these methods nonetheless display certain limitations. First, there is the problem of generalising from the results of specific case studies. The results may be

⁶ One of the few ethnographic studies that explicitly draws on Wenger’s indicators for identifying CoPs is Thompson’s (2005) study of web design teams.
interesting but do they have a wider relevance beyond the specific case or instance? Do communities of the same basic type display the same types of behaviour and learning dynamics?

Second, as laid-out in Table 1 above, ethnographic methods are less suitable, or at any rate less cost efficient methods, for capturing certain relevant dimensions of CoPs. These include:

1) Characterising the wider institutional context in which localised practice is situated.
2) Characterising geographically dispersed communities which rely on distanciated communication. This would apply notably to virtual communities whose geographic scope might well extend across national boundaries or even continents.
3) Characterising the emergence of CoPs and the evolution of their organisation over time. This would apply, for example, to capturing the reproduction cycles of communities referred to by Lave and Wenger (1991, p. 98-99). Direct observation and involvement in the activities of CoPs is impractical as a method for capturing such longer-term dynamics. Even in the case of the naval quartermaster example, which Lave and Wenger describe as having a cycle of relatively short duration, the period is five to six years.

5. The use of structured interview methods
Structured or semi-structured interview methods can be used as a compliment to ethnographic approach in order to provide additional pertinent information and to allow the researcher to check and validate conclusions arrived at on the basis of direct observation. Such methods cannot substitute for ethnographic ones for the reason that they necessarily are limited to capturing what the interviewee is able to articulate. Of course it is possible to pose questions designed to identify the use of behaviour that is not reified in the daily discourse of community members, but clearly providing a contextually rich account of such behaviour would require the use of ethnographic methods.

Our view, however, is that structured interview methods are not primarily valuable for what they can tell us about the largely tacit or unarticulated shared understandings that underlie much meaningful behaviour within CoPs. Rather their value especially lies in their suitability for gaining information from a large number of community members, or possibly acquiring
information about multiple communities, while limiting the expenditure in terms of the researcher’s time and other resources.

Consider, for example, O’Mahony’s (2003) ethnographic study of how community managed software projects protect their work. Over 100 hours of direct observation and involvement in community events such as project meetings and conferences were complimented by 75 semi-structured interviews conducted with contributors to open-source projects. Transcribed interview data in combination with codified information available from on-line archives were analysed with the support of qualitative coding software to identify the practices community managed projects used to manage and protect their work. On this basis O’Mahony identified seven different tactics employed to prevent proprietary appropriation and to assess the extent of their use across different community managed software projects.

O’Mahony’s study also illustrates what in our view is a key condition for the effective use of interview methods: careful attention to the criteria for selecting the cases to be analysed. In order to capture diversity across communities he selected six large, technically mature and well-known community managed software projects with diverse relations with firms, such as Linux kernel, Linux standards base and GNOME GUI desktop. As he observes, “the variance inherent in different types of actors and their different roles facilitated the process of constant comparisons and the development of codes and constructs.” (O’Mahony, 2003, p. 1183).

The empirical work on CoPs that makes use of structured or semi-structured interviews is vast and providing a complete overview of it would go beyond the scope of this report. Other illustrative examples of the use of such methods include Dahlander and Magnusson (2005), Creplet et al. (2001), and Bogenrieder and Nooteboom’s (2004). In each case the authors see the approach as being useful in part because of its suitability for exploring the diversity that may exist within or across communities, as well for validating the statements of individual community members by checking them against the statements of other members.

For example, Dahlander and Magnusson (2005, p. 484) in a study concerned with the relations between firms and open source software (OSS) communities note, “the rationale for studying multiple cases is the need for data regarding different approaches to inter-relating with OSS communities, as well as the perceived difficulties and opportunities involved in using them.” Four firms were chosen for a detailed analysis on the basis of semi-structured
interviews with firm members as well as through the analysis of the mailing lists and forums maintained by the virtual communities. On this basis the authors propose a typology of types of relations that firms may foster with OSS communities: symbiotic, commensalistic and parasitic.⁷

Similarly, Bogenrieder and Nooteboom (2004), in their study of intra-organisational learning groups within a consultancy firm, start from the premise that different groups can have a variety of knowledge and learning characteristics. A series of interview questions are designed so as to provide information regarding the diversity that may exist in terms of such features as the extent to which knowledge is tacit or contextualised and whether its content takes the form of professional expertise, skills, the operation of projects, etc. The notion of CoP is then used as a sort of benchmark in order to compare and contrast the knowledge, relational and governance attributes of different types of groups. Although the paper concludes that none of the intra-organisational groups examined correspond exactly to the authors’ ideal-type understanding of a CoP, it finds that the project teams studied are similar in terms of many fundamentals: a high degree of tacit knowledge, intensive interaction, the use of realistic stories for knowledge exchange, and specific investments for mutual understanding.

Interview methods, of course, cannot claim to generate representative data that can be used to test for the significance of hypotheses in the statistical sense. This is something which most researchers are well aware of, and many papers using structured interview methods offer disclaimers or recognise that larger samples would be needed as a basis for statistically rigorous hypotheses testing. For example, Bogenrieder and Nooteboom (2004, p. 300) observe, “Our empirical study of learning groups does not constitute a rigorous test of hypotheses, with quantified measures and correlations of variables on the basis of a large sample”.

At the same time, it is evident that some researchers using interview methods see them as appropriate to the fact that CoP research is in an exploratory phase of construct development. For example, Bogenrieder and Nooteboom (2004, p. 300) argue that, “at this stage the analysis needs to be more exploratory and process oriented”, and they see the goal of their

⁷ Also see Lee and Cole (2003) and von Krogh et al. (2003) who use mailing lists to follow the activities of OSS communities.
empirical work as being, “more a test of usefulness than of validity”. In a similar vein, Creplet et al (2001, p. 1523) in their study of experts and consultants in relation to epistemic communities and communities of practice, see their work as falling between the objective of grounding, “a previously built theoretical framework empirically,” and trying, “to infer a conceptual formalisation from a data collection.” They observe that their interview material, “provided new insights that allowed us to develop new ideas or link ideas in an original (at least for us) way.”

6. Exploratory and confirmatory statistical methods for studying CoPs

6.1 The use of survey data

Empirical research on CoPs based on survey-data from large samples is in its infancy and to our knowledge no proposals have been made for developing harmonised measures of CoPs that might ultimately find their way into the national-level surveys developed by the statistical services making up the European Statistical System. As the discussion in the above paragraph implies, further theoretical and definitional work clearly would be required as a prerequisite for developing indicators of CoPs that might respond to the interests of a wide range of stakeholders. However, rather than focusing on these theoretical and definitional issues which go beyond the scope of this report, we focus here on certain sampling frame-related problem that would be encountered in developing national-level surveys of CoPs. Some of these problems, in our view, may prove to be intractable.

All large scale national surveys start by identifying the target population that the survey will investigate and the frame population forming the units from which the survey sample are drawn. In many surveys concerned with organisational behaviour the target population is enterprises or establishments above a certain size in the business enterprise sector and the frame population might be an up-to-date business register such as many national statistical services establish specifically for statistical purposes. In other cases, for surveys focusing on working conditions or employee training, the target population might be all occupied persons working in establishments above a certain size and the frame population might be a census of households developed for statistical purposes.

In both of these cases, researchers developing a survey instrument will adapt the choice of statistical unit to the established conventions of the national statistical services. There are, of course, other connected issues that have to addressed in setting appropriate statistical
procedures, including the choice of sampling technique (simple random, stratified sample, cluster sample, etc.), the choice of survey methods (postal, telephone, on-site interviews, etc.), the identification of appropriate respondents, and deciding on the length of the questionnaire.

A basic, and in our view probably intractable, problem in developing national survey instruments at the CoP-level is the absence of relevant registers that could constitute the frame population. Moreover, it is not at all clear how such a register might be constructed since CoPs, unlike enterprises, are not legally established entities that can be identified in an exhaustive manner because they are required to register for census purposes.

This suggests adopting the alternative strategy of using a census of households as the frame population and then defining the target population as some subset of the frame, for example all occupied persons. A questionnaire could then be designed with the intent of collecting information on occupied persons’ involvement in (possibly multiple) CoPs. A key problem linked to this sampling approach, however, pertains to the fact, as Wenger (1998) has stressed, that CoPs are not always reified in the discourse of their members. Thus to be effective, the questionnaire would have to provide a clear definition of a CoP that could be understood in an unambiguous manner even by respondents who are unaware that they are actively involved in CoPs. This might prove impossible to achieve with postal survey methods and suggests that direct interview methods would be required in order to verify the respondents understanding of what CoPs are. Such survey methods are very costly and typically outstrip the budgets of academic research teams. Where they are used, they are generally financed by national statistical services.

Given these problems and limitations, what contribution can survey methods make to the study of CoPs for research teams with limited budgets? At this stage in the development of empirical research on CoPs we believe that survey methods will be useful tools for improving our understanding of CoPs in so far as they are used for studying well-defined CoPs whose membership is easily identified. In this case the target population and frame population could be identical consisting of the entire membership of the CoP. Sampling fractions could then be set in accordance with the size of the CoP membership or particular strata in the event that there was stratification, say according to region.
One recent study that applies this approach is Franke and Shah (2003) statistical study of four sports-related user-innovator communities. There is, of course, a burgeoning literature on the role of users in innovation processes and in common with research on open-source software communities one of the issues that is addressed in this literature is the motivation of persons who contribute their knowledge to the collective pool without apparently anticipating any material benefit. Another central preoccupation concerns the diffusion of innovations within the community. Both of these issues are tackled in Franke and Shah’s study.

For the target and frame population they chose four relatively small and well defined sports enthusiast communities that differed in their make-up, or constituency, and structure. These included: a sailplanning community consisting of students of technical universities in Germany; a canyoning community established in 1995 as a forum in which to organise joint activities; a boardercross community consisting of semi-professional athletes from all over the world who meet in up to ten competitions a year; and a handicapped cycling community whose members are well known from national and international competitions. The total memberships of these communities are, respectively, 170, 123, 170 and 58 persons, giving a total sample size of 521 persons.

The data were collected either by sending paper questionnaires to the members or through e-mail correspondence describing the study and containing a link to an on-line questionnaire. The overall response rate was slightly over 37%. The results were analysed with different statistical methods including exploratory techniques (factor analysis, simple descriptive statistics) and confirmatory methods (parametric statistics to test for difference between different sub-populations, regression analysis). On this basis the study both managed to provide useful characterisations of user-innovator communities and to test for the statistical significance of a number of hypotheses. Amongst the interesting results from the study include evidence that community members are an important source of information for users who innovate and that those users who do innovate are more likely to give assistance to other members of the community, including those that they do not know personally, than are members who do not innovate.
Statistical methods applied to the study of virtual communities

Ethnographic methods as well as on-site interview methods have obvious limitations for capturing the characteristics of virtual communities whose boundaries may stretch across nations and even continents. Structured interviews conducted with key contributors in such communities may, in combination with publicly available documents found on the websites of community projects, constitute important sources of information. But they are likely to provide at best a partial representation of the organisation of such communities and of the learning processes and knowledge exchange that occurs within them.

In our view, the most promising methods being applied for the investigation of virtual communities are those that compliment interview methods and the use of publicly available contextual material with the traces of virtual contacts and exchanges within such communities to be found on the internet (e.g. forums, mailing lists and records of e-mail communications). Such virtual traces can be used to construct representative data sets that are suitable for analysis with the use of exploratory and confirmatory statistical methods. The target population in such an approach is the entire membership of the community and the frame population is constructed from the set of the persons participating in forums, on a particular mailing list, or engaging in e-mail communication.

An example of such an approach to the study of a virtual community is Krogh, Spaeth and Lakhani’s (2003) study of innovation processes within Freenet, a project aimed at developing a decentralised and anonymous peer-to-peer electronic file sharing network. The study is particularly concerned to explore the processes whereby new people join the existing community of software developers and how they initially contribute code.

The statistical data analysis which is largely descriptive, makes use of two sources: the public e-mail conversations stored on the ‘development’ mailing list of the project and the history of changes to the software code that were available via the project’s software repository. While the project had three other mailing lists, interview and comparisons showed that the ‘development list’ was the only list where significant technical discussions took place.

\[8\] For an overview of the burgeoning literature on virtual communities, see Johnson (2001) and Amin and Roberts (2006). Amin and Roberts argue that on-line communities can be sources of knowledge creation and they tentatively conclude that these communities are more likely to achieve this when there is a clear problem-orientation of the members who have domain-specific professional competence and when the community is coordinated by a manager who is sensitive to the importance of fostering a culture of engagement and mutual respect.
set consisting of 11,210 single e-mail messages was constructed and by coding the messages it was possible to identify 356 unique participants. Analysis of the history of changes to the code over the period 1 January to 27 December 2000 identified 1244 source code commits from 30 different developers. This group of thirty developers were then categorised into three different types: ‘joiners’ (those on the mailing list but without access to the source code management tool), ‘newcomers’ (someone who has been given access and just begun to makes changes to the code) and ‘developers’ (someone who has gone beyond newcomer status and is contribution to the project).

By analysing the timing at which persons joined and left the project, the intensity of their communication and their additions to code the authors draw conclusion concerning the transition from joiner to newcomer status as well defining characteristics of newcomers in relation to developers. They use their findings to draw implications for such issues as why open source software community developers contribute for free to the production of a public good.

An alternative method to identifying structure within an OSS community is to differentiate participants according to standard network theory-based measures. This is the approach used by Dahlander and Wallin (2006) in their study of the way firms assign individuals to work within the GNOME free and open source software (FOSS) community with the aim of gaining access to knowledge development activities within this community and possibly influencing the direction of its development. GNOME is a user community consisting of thousands of individuals who interact and, to varying degrees, participate and help each other in the development of a Windows-like desktop environment. Based on preliminary interviews with community members, the authors adopted the GNOME development mailing list as a basic measure of the overall community network. A data set was constructed consisting of 14,644 mails on various topics sent over the period from the project’s initial development in 1999 to December 2004, and on this basis 1659 unique individuals were identified.

By using information on ‘directed relationships’, where a distinction is made between members who initiate contacts from those that receive contacts, the authors develop various measures of interaction patterns within the network. These include an individual’s prominence within the network, or the degree to which others seek to interact with that individual (INDEG), the extent to which an individual helps others and so contributes to the
overall project development (OUTDEG), the extent to which individuals have more persons contacting them than they contact (PRESTIGE), and a measure of the extent to which individuals are connected to central individuals (EIGENVEC). Using these indicators, a number of hypotheses concerning the behaviour of persons who are affiliated to firms that participate in the GNOME community are statistically tested. The results indicate, for example, that firm-sponsored individuals, because of their resource advantage, can interact with many different individuals for the good of the project and on this basis enhance their reputation and legitimate their commercial use of the project. The results also point to differences in the role and influence on the community of individuals sponsored by firms that have incorporated open source in their business models as compared with individuals sponsored by incumbent firms in the software industry.

7. Capturing dynamic processes of emergence or change in CoPs

One of the interesting characteristics of data on virtual communities coming from mailing lists is that it is historical. Because of this, mailing list data also provides unique opportunities for studying dynamic processes of emergence and change in CoPs. For example, Krogh et. al. (2003) are able to use the history of discussions on the development list of the Freenet project in order to analyse processes of joining and becoming a member as well as changes in members’ status over time. In the case of non-virtual CoPs, however, the only data sets we are aware of are cross sectional in nature and poorly adapted to analysing processes of change.

One way in which these constraints can be overcome is through the use of computer simulation methods which can be used to generate statistical measures of the characteristics of members within a structured community. By observing the evolution of these measures over time (simulation periods), it is possible to characterise the evolution of behaviours under different parameter specifications.

A recent example of the use of computer simulation methods to describe CoP dynamics is Muller’s (2006) study of leadership building in CoPs. The simulation starts with a random graph of 250 individuals linked on average to 10 individuals. The links are bidirectional. The network is a relatively ‘sparse’ graph in the sense that that only 4 percent of the possible connections within the network are active. The model assumes that individual members can be of two types: either they are highly committed in the sense that they contribute extensively to the community’s work; or they are slightly committed and chose to profit from the
contributions of others. One of the parameters that is varied in the different simulations performed is the differential in the level of commitment between these two types of members. Another parameter that is varied is the percentage of the population that is committed.

The statistics that are analysed in order to capture the structuring effects of the social network forming the community are measures of ‘degree centrality’ and ‘betweenness centrality’. Degree centrality corresponds to the number of acquaintances and constitutes a standard measure in network theory of an actor’s centrality in the network. Muller assumes that more central individuals in the sense of degree, collect more information and knowledge through interpersonal connections than others. Knowing this, the members of the community tend to copy such members’ behaviour. Muller argues that this is one of the defining characteristics of leaders in CoPs.

The other relevant statistic, ‘between centrality’ refers to an actor’s role in bridging two distant parts of a graph in the sense that contacts between two actors pass via the bridging actor. Muller argues that persons with a high level or betweenness may derive power from their position since they can favour the diffusion of certain categories of information. This allows them to influence the behaviour of other members and constitutes another defining trait of leaders within communities. By analysing these statistics on the basis of a number of simulations under different parameter specifications, Muller arrives at a number of interesting conclusions about the dynamics of leadership building within communities. One of these is that small difference in individual behaviour may induce significant differences in social position with the most committed becoming leaders. Another conclusion is that the nature of leadership is closely tied to the share of highly committed individuals, with communities characterised by many highly committed individuals having a coalitional style of governance and those with a low share of committed members having a more authoritarian style of governance.

The use of simulations to generate relevant statistical measures based on network theory can be combined with other computational methods in order to embed the network within a model that also captures the evolution of knowledge representations. This is the case with the computational models developed by Dupouët et al. (2003) and Dupouët and Yildizoglu (2003) who use a learning classifier system in order to characterise learning process within communities. The learning classifier system developed by J. Holland et al. (1986) is a
machine learning system in which the agents’ knowledge about the environment takes the form of a set of condition-action rules that link a representation of the environment to a particular action. Greater or lesser degrees of ignorance or uncertainty about the environment can be modelled, and when an agent receives a signal from the environment multiple conditions or representations may be matched. A series of parameters are attached to each rule based on its past usefulness, the expected return associated with its activation, and the error associated with this prediction. The matched rules are evaluated on the basis of these parameters and the rule with the highest predicted return is activated. The parameters are then modified on the basis of the realised performance. Changes in the knowledge base (set of condition-action rules) of the agents may be generated by means of genetic algorithms, which, for example, can evolve conditions in the direction of greater generality (the set of environmental conditions that match a rule widens) or greater specificity (the set of environmental conditions that match a rule narrows).

One of the interesting characteristics of learning classifier systems is that agents’ knowledge representations evolve over time as a result of the combined effect of the reinforcement mechanism and the modifications in the individual knowledge bases due to the use of genetic algorithms. It is thus possible, for example, to start out with agents that are completely ignorant about the relevant features of their environment and who over time through the learning mechanism develop a structured representation of their environment (see, for example, Marengo, 1992). For this reason, the classifier system is a legitimate candidate for providing a formal representation of a contextualised learning dynamic. Moreover, by modifying the assumptions regarding the forms of communication that can take place between agents, it is possible to simulate the learning dynamics and assess the performance of diverse organisational structures including network structures.

In Dupouët et al. (2003), the authors assume that each agent within an organisation has a limited knowledge and expertise. When a problem is faced that the agent lacks the necessary knowledge to solve, or when the predicted performance from adopting the solution is too low, the agent communicates and exchanges knowledge (condition/action rules) with a fraction of the other members of the organisation. If a received solution offers a superior performance it is accepted. In this process, the agent develops not only an enriched knowledge base but also ‘know-who’ in the form of an ‘address book’ of contacts. This address book of contacts can be exchanged with partners that are considered trustworthy.
The simulation thus gives rise not only to an evolving knowledge base of the agents but also a network structure which the authors characterise, as in the case of Muller (2006), on the basis of network theory-based statistical measures: the degree of cliquishness, which measures the presence of communities within the overall network of agents; the degree of betweenness, which measures the importance of an agent’s position as intermediary between the members of a community; and the in-degree and out-degree, which measure the importance of the communications received and communications made on the part of an agent. Agents who receive more communications than they make are considered to play a more central role in the community.

On this basis, the authors are able to explore how varying degrees of specialisation of the knowledge base of agents affects the emergence of communities, and how increasing agents’ capacity to communicate in general affects the relative importance of communication directed within and outside communities. In a subsequent paper Dupouët and Yildizoglu (2003) use similar methods to contrast the performance of hierarchical, community and hybrid structures. They conclude that communities can be efficient mechanism for competence building and that community and hierarchy should be seen as complimentary coordination mechanisms rather than as substitutes.

The underlying premise of learning classifiers that human cognitive architecture consists in symbolic representations can, of course, be contested. On the other hand, classifiers are not vulnerable to another criticism often made of artificial intelligence methods that knowledge is conceived as something that exists within the mind independently of the agent’s relation to the external environment. The use of the learning classifier system allows Dupouët et al. (2003) to model a process whereby knowledge representations evolve in part due to what agents learn by interacting with the other agents making up the network. Although their simulation exercises obviously cannot capture the rich detail of qualitative studies of locally contextualised interactions within communities, they have the merit of being explicit about the nature of knowledge and the types of exchanges that are made. The method thus provides a quite powerful tool for exploring the impact of varying assumptions concerning the intensity of communication, or the narrowness of processes of exploration, on community organisation. In our view, such methods can be used to good effect when they are inspired by and complimented by more qualitatively rich ethnographic accounts.
8. Conclusion: in favour of methodological pluralism

The main objective of this report has been to compare different methods for studying CoPs with an emphasis on contrasting what qualitative and quantitative methods can and cannot capture. Drawing clear conclusions in this regard is complicated by the fact that the burgeoning literature on CoPs is characterised by considerable diversity in terms of the kinds of things that are analysed as CoPs. Further, different types of CoPs differ along a number of distinct dimensions, including the type of knowledge, the role of physical and cognitive artefacts, the nature and relative importance face-to-face vs. distanced communication, the importance of interpersonal ties in relation to wider institutional arrangements, and the relative importance of and relation between informal and formal practice. One of the main arguments of this report is that different methods are more or less suitable for capturing different dimensions of CoPs, and this provides a first justification for supporting methodological pluralism in empirical research on CoPs.

An implication of this argument is that researchers will have an interest in crafting their methods to the defining features of the type of CoP being studied, placing greater emphasis on one method or another in accordance with its suitability for capturing specific characteristics. For example, the key role of kinaesthetic and embodied knowledge in the case of craft/task based communities implies a privileged role for ethnographic methods, which are perhaps the only ones that are suitable for generating rich contextualised characterisations of how such knowledge is developed and mobilised in processes of task accomplishment. In the case of professional communities, on the other hand, ethnographic methods are likely to prove less suitable, or at least less cost efficient, as compared to structured interview methods for gaining information about the way wider institutional arrangements structure processes of knowledge generation and transmission. The importance of distanced communication in the case of virtual communities, as well as the fact that such communication is typically codified in the mailing lists of the projects around which such communities are formed, suggests a privileged role for statistical methods for the study of data sets constructed from the traces of virtual communication and interaction.

This conclusion needs to be qualified by noting that regardless of the type of CoP being studied there is scope for making effective use of multiple methods of study. One reason for this is that the dimensional differences we have referred to above are differences of degree
rather than absolute differences. Knowledge, for example, inevitably has tacit and embodied elements and ethnographic methods can be used to good effect for capturing these elements regardless of the type of CoP under consideration. Another reason for this is that the information provided by one method can contribute to the effectiveness of another method. For example, survey data-based studies may draw inspiration from ethnographic accounts or interview methods in generating hypotheses to be tested. They may also make use of on-site interviews in order test the formulation of survey questions for possible ambiguities of interpretation. Thus, a second justification for methodological pluralism is that research methods often prove to be complimentary with respect to one another.

Being open to methodological pluralism can also be justified in relation to the relative advantages of different methods for addressing certain problems that are transversal to different types of CoPs. One of these is generating data that is representative of CoPs. The limitations of qualitative ethnographic or interview-based methods for generating representative data are not, of course, in any way specific to the study of CoPs. We have focused in the report on problems linked to the identification of frame populations and have argued that these problems can best be addressed when the frame and target populations are the entire membership of established and recognised CoPs. Virtual communities will satisfy this criterion in those instances where the community membership can be identified with the contributors to specific mailing lists, and this is a factor in the successful use of statistical methods applied to the study of virtual communities.

Another transversal problem is studying dynamic process of emergence and change in CoPs. In the case of virtual communities, the historical nature of mailing list data offers unique opportunities for studying processes of change, but the more general use of survey methods for analysing dynamics would require the creation of panel data. We have argued that an alternative and complimentary way for studying dynamics is through computer simulation methods. These have the advantage of allowing the researcher at minimal cost to explore a variety of scenarios under well-specified assumptions as to the structure of the interactions between members and under different parameter specifications.

We opened this report by referring to the increasing diversity that can be observed in research on CoPs and to the problem of a lack of cumulativeness in empirical results that this has generated. In conclusion we would like to emphasize that we believe this diversity should be
encouraged in so far as researchers are disciplined about paying attention to differences in the underlying dimensions of different types of CoPs. By being attentive to differences across CoPs in terms of these dimensions, and by using appropriate combinations of empirical methods to characterise and analyse them, the existing diversity can be a positive factor for realising the potential of the CoP framework of analysis for improving our understanding of processes of learning and knowledge generation situated in certain kinds of practice.

References


