HAL
open science

A Situated Approach of Roles and Participation in Open Source Software Communities

Flore Barcellini, Françoise Détienne, Jean-Marie Burkhardt

To cite this version:

Flore Barcellini, Françoise Détienne, Jean-Marie Burkhardt. A Situated Approach of Roles and Participation in Open Source Software Communities. Human-Computer Interaction, 2013, vol. 29 (n° 3), pp. 205-255. 10.1080/07370024.2013.812409 . halshs-00860554

HAL Id: halshs-00860554

https://shs.hal.science/halshs-00860554v1

Submitted on 10 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Barcellini, F., Détienne, F., Burkhardt, J.M. (2013). A situated Approach of Roles and Participation in Open Source Software Communities. *Human Computer Interaction,* DOI: 10.1080/07370024.2013.812409

A situated approach of roles and participation in Open Source Software Communities

Flore Barcellini* (1), Françoise Détienne (2), Jean-Marie Burkhardt (3)[1]

* *Corresponding author*

(1) Le Cnam, CRTD, Ergonomics team, Paris, France

(2) LTCI- UMR 5141 - CNRS - Telecom Paris Tech, Paris, France

(3) IFSTTAR LPC, Versailles-Satory France

RUNNING HEAD A situated approach of roles and participation in Open Source Software Communities

Corresponding Author's Email address: flore.barcellini@cnam.fr

Abstract. Our research aims at understanding the various forms of participation in Open Source Software (OSS) design, seen as distributed design in online spaces of actions - discussion, implementation and boundary between these spaces. We propose a methodology - based on situated analyses of a formal design process used in the Python project- to identify the distribution of actual roles (implementation, interactive, group and design oriented) performed by participants into and between the spaces (defining boundary spaces). This notion of roles is grounded in collaborative design activities performed online by participants. This way, our findings complete the core-periphery model of participation in OSS. Concerning the distribution of roles between spaces, we reveal a map of participation in OSS: the majority of participants are pure discussants but all participants in the implementation spaces do also act in the discussion space and only few participants act at boundary spaces. Concerning the distribution of roles between participants in the discussion space, we reveal that interactions are structured by a central hub (occupied by key-participants) and that, whereas design-oriented roles are spread among all participants, group-oriented roles are performed by one or two participants in the respective spaces and at their boundary. Finally, combination of roles reveals five individual profiles performed by participants. Our approach could be extended to other design situations to explore relationships between forms of participation- in particular those revealing use-oriented contributions- performance, and quality of the design product. Finally, it could be a basis for specifying tools to monitor and manage community activity for both research issues and support of online community.

---

[1] **Flore Barcellini** is an ergonomist with an interest in cooperative and mediated activities and online communities; she is a associate professor in the Research Center on Work and Development (CRTD) of Le Cnam. **Françoise Détienne** is a cognitive psychologist and ergonomist with an interest in collaborative design, technology-mediated collaboration and online communities; she is a research professor at CNRS in the Department of Social, Economics and Human Sciences of LTCI (CNRS - Telecom ParisTech). **Jean-Marie Burkhardt** is a psychologist and ergonomist with an interest in collaborative design, mixed realities and learning technologies; he is a research professor in the Laboratory of Driver Psychology of IFSTTAR - Institut Francais des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux.

Keywords. Open Source Software, roles, participation, activity, boundary spaces

Contents

1. INTRODUCTION 3

2. PREVIOUS WORKS ON DISTRIBUTION OF DESIGN AND FORMS OF PARTICIPATION IN OSSD 6

2.1. OSSD: A DISTRIBUTED DESIGN PROCESS ON THE INTERNET 6

2.2. FORMS OF PARTICIPATION IN OSS COMMUNITIES : AN "ONION-RING" CORE-PERIPHERY VIEW OF PARTICIPATION 8

2.3. ADDRESSING PARTICIPATION IN OSSD 10

2.4. TOWARD AN HYBRID AND A SITUATED APPROACH TO ANALYZE PARTICIPATION IN OSSD 11

3. PREVIOUS WORKS ON ROLES AND PROFILES TO EMBRACE PARTICIPATION IN A SITUATED DESIGN CONTEXT 12

3.1. DEFINITION OF ROLES AND THEIR DIMENSIONS 13

Interactive roles: a communication network-based analysis 13

Task-oriented roles: Goal-oriented activities contributing to the performance of the design 14

Group oriented roles: Socially-oriented activities to support and reinforce the team 15

3.2. PROFILES OF PARTICIPATION AS COMBINATIONS OF ROLES IN SITUATED DESIGN CONTEXTS 15

3.3. GENERAL TRENDS OF PARTICIPATION IN SOFTWARE DESIGN SITUATIONS 16

4. RESEARCH QUESTIONS AND STRATEGY 17

4.1. RESEARCH QUESTIONS 17

4.2. RESEARCH STRATEGY 19

Four types of roles characterizing participation in OSS context 19

Investigating participation in boundary spaces 20

5. METHOD 21

5.1. CONSTRUCTION OF THE CORPUS 21

Identification of situated data of design: selection of PEP process 21

Collection of actions in the implementation and discussion spaces 22

5.2. DESCRIPTION OF THE SELECTED CORPUS 24

Description and chronological organization of the PEP 327 corpus 24

Actions and discussions considered to analyse roles 24

Position of PEP 327 in the whole PEPs ' scenery *25*
Key-participants initially identified in the selected corpus *26*
5.3. CODING SCHEME OF ROLES AND CONSTRUCTION OF PROFILES 27
Coding scheme *27*
Construction of profiles *28*
5.4. CONFRONTATION OF ANALYSIS 30
6. RESULTS : FROM PARTICIPATION AND ROLES TO PARTICIPANTS PROFILES 31
6.1. PARTICIPATION AT THE BOUNDARIES DURING THE PROCESS 31
6.2. DISTRIBUTION OF ROLES AND PARTICIPATION BETWEEN AND ACROSS SPACES: A FOCUS ON THE USE DISCUSSION SPACE AND THE DESIGN DISCUSSION SPACE 33
Distribution of interactions: "y-forks" structured around a central hub *34*
Distribution of design-oriented roles and concentration of group-oriented roles *36*
6.3. SYNTHESIS : INDIVIDUAL PROFILES OF PARTICIPANTS 37
7. DISCUSSIONS AND FURTHER WORKS 40
7.1. A NEW APPROACH OF PARTICIPATION IN OSSD 40
Insights *40*
7.2. IMPLICATIONS 42
Understanding more deeply a successful design organisation *42*
Generality of the approach and its potential applications to other design contexts *43*
Automation of the approach *45*
8. ACKNOWLEDGMENTS ..... ERREUR ! SIGNET NON DÉFINI.
9. REFERENCES 46
10. APPENDIX 52

1. Introduction

Through the impulse of Internet tools and emergence of online communities[2] the question of online participation is currently attracting increasing interest. Among all the various types of community (social support, epistemic…) we are

[2] a group of people connecting via the Internet with a common goal (Preece, 2000)

focusing more precisely on online epistemic communities (e.g. Preece, 2000; Cohendet, Creplet & Dupoüet, 2003). In these communities, collective activities occur and are finalized by a common objective of knowledge construction (Barcellini, Détienne & Burkhardt, 2010a), such as in Open Source Software (OSS) projects (e.g. Crowston, Wei, Howison & Wiggins, 2012; Barcellini, Détienne & Burkhardt., 2009; Crowston & Howison, 2005; Scacchi, 2001) or in Wikipedia (e.g. Welser et al., 2011; Gleave, Wesler, Lento & Smith, 2009; Bryant, Forte & Bruckman, 2005, Fréard et al., 2010). The research presented in this paper is a part of research program (Barcellini, Détienne, Burkhardt & Sack, 2005; Sack et al., 2006; Barcellini, Détienne & Burkhardt, 2008, 2009; Fréard et al., 2010) aiming at establishing theoretical and methodological approaches for achieving a deep understanding of processes underlying these situations of knowledge co-elaboration (Détienne et al., 2012).

In this paper, we illustrate our approach by investigating Open Source Software Design performed by volunteer participants, connecting via the internet with a common goal of producing and constructing knowledge about the software they develop or about using this software. Our position is to view knowledge co-elaboration in this case as a *collaborative design activity distributed* across several *'spaces'* on the Internet (mainly discussions spaces and implementation spaces). Indeed, As Curtis (1986) asked : *'By the way did anyone study any real programmers*?' we would also like to pose the question, ' *By the way, did anyone study any actual collaborative design activities in Open Source Software communities?*'.

In a cognitive point of view, *design activity* was classically considered as ill-structured problem solving process (Détienne, 2006; Visser, 2006) characterized by many degrees of freedom in the problem's initial state; unstable design requirements in terms of constraints and sub-goals; co-evolution between problem-solutions(s); various possible solutions with no definite way for choosing the best one. Nowadays, design activity is considered as a *collaborative* and *social process*. Design solutions are not only based on purely technical problem-solving criteria but are the results of a negotiating process among disciplines and perspectives (Bucciarelli, 1998) and has to be apprehended in a socio-cognitive perspective (Détienne, 2006). In OSS, the collaborative process mainly consists in the co-elaboration of knowledge objects (piece of software), which are the results of design choices and compromises among participants from various backgrounds and application domains. Assessing negotiations underlying design process requires to focus on actual interactions between participants pursuing a common goal. Indeed, we consider activities in online epistemic communities as finalized by common goal at a global level, e.g. designing an open-source software and by specific common goal linked to tasks or subtasks at more local levels, e.g. designing a specific OSS function. In the sense, our research aims at focusing on tasks oriented by co-elaboration of design solutions and not coordination among

participants taking in charge tasks (e.g. Herbsleb, 2007; Crowston, Qing, Kangning, Eseryel & Howison, 2007; Zhou & Mockus, 2010). In this view of design, collaborative design activities may occur in various phases of design process : from phases in which solutions are abstractly defined (e.g. elicitation of needs), coming through the specification phase to the implementation phase (e.g. coding) in which design choices are reified – and sometimes re-elaborated -into a software artifact[3]. In the case of OSS design process supposed new functionalities can always be proposed and discussed whatever the step in the project (Gasser, Scacchi, Ripoche & Penne, 2003).

Moreover, we assume that these new design situations are of particular interest for HCI research because online volunteer participation leads to more *distributed* design processes (e.g. Crowston et al., 2012) compared to traditional forms of design such as in small collocated design teams (e.g. Olson, Olson, Carter & Storrosten, 1992; d'Astous, Détienne, Robillard & Visser, 2001; d'Astous, Détienne, Visser & Robillard, 2004; Détienne, 2006). Here, the term "distributed" encompasses several types of distribution : (1) design processes can be distributed in time, in space, and among organizations, (2) they are potentially performed by a large number of participants whatever their statuses (Détienne et al., 2012; Barcellini et al., 2008a), and (3) they are mainly distributed among internet artifacts (such as e-mails, web cooperative platforms, blogs, wikis …), that are defining two *spaces of actions* (see section 2.1) (e.g. Sack et al., 2006; Barcellini et al., 2005; Détienne et al., 2012)[4] : *discussion space* (mailing-lists, forums, chats…) and *implementation space* (versioning of online documentation, revisions of code[5]). In these situations, participation becomes hybrid (Ducheneaut, 2005; Sack et al., 2006) as one person may potentially participate to many actions, i.e. many spaces. An issue is thus to investigate how actions of participants are spread in and between these space, i.e. at their boundaries.

In this paper, we propose to embrace participation in and between spaces through the notion of actual roles and profiles performed by participants. Our position is that distinction of participants based on their actual activities, i.e. on their roles, would enable to clarify and to characterize participation in OSS context. By actual, we mean that roles are strongly linked to activities effectively performed by participants, and emerging from interactions between participants pursuing a common goal (Baker, Détienne, Lund & Séjourné, 2009; 2003). As we will show later, this is a complementary perspective to studies that have mostly view participation in OSS as based on status, static roles or positions, performed

---

[3] Coding activity may be mostly individual whereas it may be collective in some specific methodology such as pair programming for instance.

[4] We previously made a distinction between three spaces named discussion, documentation and implementation. However, we believe that it is ambiguous to refer to a single space embracing the proper implementation composed of documentation and coding.

[5] Supported by the Concurrent Version System (CVS) or by Subversion

mainly in a single space (e.g. the implementation space in OSS) (e.g. Crowston et al., 2012 for a review).

The research reported here is focused on a major OSS project, Python, which is an object-oriented programming language (www.python.org). We analyze online actions and collaborative design activities performed by Python participants engaged in a specific design process used to define and to specify revisions and improvements to the Python language- *Python Enhancement Proposal*- or PEP.

After a review of previous works dealing with distribution and forms of participations in OSSD and the various perspectives concerning the notion of roles we will present our research approach and questions. In order to understand participation, i.e. distribution of roles in the various spaces of actions OSS design process, we choose to perform situated analyses of activities in a selected design process. Finally, we will present and discuss our results.

2. Previous works on distribution of design and forms of participation in OSSD

In the first section, we review organizational characteristics of OSSD design stressing and illustrating its distribution in both spaces of actions (implementation and discussion). In the following section, the forms of participation in OSSD are highlighted stressing that they are mainly based on development tasks performed by participants and not on analysis of actual collaborative design activities they performed.

2.1.OSSD: A distributed design process on the Internet

OSSD is a highly mediated design situation as participants rarely meet face to face and actions of participants occur through series of online artifacts (mailing-list or fora, versioning systems, platforms).

We propose to view OSSD has been distributed between two *spaces* of actions - implementation and discussion – on the Internet (Barcellini et al., 2005; Sack et al., 2006; Détienne et al., 2012) (Figure 1). These spaces are defined by actions made by participants on different artifacts and are named after the functions of these artifacts:

-*discussion spaces* are named after artifacts dedicated to discussion between participants (mailing lists, forums, chat). Usually, OSS projects have at least two types of mailing-lists:

- a development mailing-list defining a *design discussion space*. For instance, in Python *python-dev* is the list where the future of the language and its implementation (Python design issues, release mechanics, and maintenance of existing releases) is discussed.

- a general mailing-list dedicated to general issues when using the software defining a use discussion space, such as *python-list* in the Python project.

-*implementation spaces* is named after artifacts dedicated to production and documentation of code. It is composed of:

- *documentation space* related to producing and maintaining online documentation. Documentation relative to code can be spread on several web artifacts (blogs, wiki, projects documentation on project website), even if some projects tend to centralize their documentation (e.g. www.python.org/doc) and their versions (svn.python.org).
- and *coding space* related to actions of producing new pieces of code (module for instance), maintaining code (producing patches), bug reporting.

[FIGURE 1 ABOUT HERE]

Figure 1 Distribution of OSS design process in discussion and implementation spaces

Indeed, the history of a project's implementation and its different versions is supported by versioning systems like CVS or Subversion (coding space). Bug report or feature request systems can be also provided by default by OSS repository platforms such as Sourceforge or bugs.python.org (coding space). However, important steps in design – requirements analysis, specification and elaboration of solutions- are often spread in discussions, sometimes in feature request system (Crowston et al., 2012).

To try to frame this, large projects may elaborate explicit design coordination mechanisms (Crowston et al., 2007; Mockus, Fielding & Herbsleb, 2002). These mechanisms are also partially supported by online artifacts. For instance, several OSS projects have settled some formal enhancement proposal process framing feature requesting, from which Python was a pioneer with the Python Enhancement Proposal (PEP). PEP process is an explicit means for proposing changes and extensions to the Python language. These processes are quite similar to two design processes used in conventional software projects: Request For Comments (RFCs) and technical review meetings. RFCs have been practiced for decades to define standards for the Internet (especially by the Internet Engineering Task Force, IETF). Technical Review Meetings (D'Astous et al., 2004) have been practiced in many corporate and governmental settings. The PEP process is also similar to other design processes used by distributed design communities, like the XEP process (Xmpp Extension Protocol) of the jabber community (www.jabber.org), and the PLIP process (Plone Improvement Proposal) (www.plone.org) of the Plone project. These formal processes are also close to the consensus-based decision making of Apache (www.apache.org) in which design issues are discussed and chosen online.

Studying Enhancement Processes is a way to embrace collaborative design activities occurring online in both spaces of actions (e.g. Sack et al., 2006;

Barcellini, Détienne, Burkhardt & Sack, 2008b). Indeed, during the PEP process, community input are collected in the discussion spaces and the chosen design decisions are documented (in the documentation space via a PEP document6). Archives of discussion, decisions regarding the PEP, and the different versions of PEP are archived in the different spaces. After a PEP has been accepted, a new piece of code (for instance a module) is written to implement the PEP. This code is integrated into the project's code archive: the coding space. Thus, a PEP design process includes actions of documentation, but also of discussion and implementation: it is distributed in both spaces of actions.

One issue is to understand how activities and participation are distributed between and across implementation and discussions (both design and use) spaces, which embrace different types of activities: interactions about artifact to be designed or usage of Python in the discussion spaces, production and maintenance of documents regarding this artifact in the documentation space and reification of this artifact into code in the implementation space (code commit and revisions of code). Moreover, in an OSS design, studying participation also refers to studying distribution of this participation between certain use space and design space, i.e. spaces in which issues regarding design and/or use (elicitations of requirements for instance) are processed. OSSD projects offer clear separated spaces where design or use (use-oriented mailing-list) can be discussed (through dedicated development or usage mailing-lists). This question is of particular interest in OSS communities, since the divide between use and design is supposed to be less strict than in a "traditional" design organization.

2.2.Forms of participation in OSS communities : An "onion-ring" core-periphery view of participation

Forms of participation in an OSS project are often understood based on a "core-periphery" hierarchy (Crowston, 2011) represented either by an "onion ring" model of participation (Jergensen, Sarma, & Wagstrom, 2011; Bach & Twidale, 2010; Hedberg & Livari, 2009; Jensen & Scacchi, 2007; Crowston & Howison, 2005; Ducheneaut, 2005; Nakakoji, Yamamoto, Nishinaka, Kishida, & Ye, 2002) or by a pyramidal model of participation (Jensen & Scacchi, 2007; Gacek & Arief, 2004).

All hierarchies reviewed are based on a classical distinction in epistemology of design between *developers* and *users* (Bach & Twidale, 2010; Jensen & Scacchi, 2007; Crowston & Howison; 2006; Gacek & Arief, 2004; Mahendran ; 2003; Nakakoji et al., 2002). By default, all participants involved in OSS may be *users* of the software. *Passive users* sometimes called *observers* (Jensen & Scacchi,

[6] The PEP document is written to describe a new language feature: it is intended to provide a concise technical specification of the feature, the rationale for the feature, and a reference implementation

2007) or *readers* (Nakakoji et al., 2002) are those users who "only" use the software, and possibly look at online resources (mailing-lists, fora, documentation). There is no online trace of their implications in OSSD neither in the discussion spaces nor in the implementation spaces. These so called "users" can be more or less skilled in computer science according to the OSS considered (e.g. the web browser Firefox vs. the programming language Python) (Ducheneaut, 2005).

Afterwards in the hierarchy, participants are mostly named after development tasks, i.e. regarding code production or documentation (i.e. actions in the implementation spaces), possibly performed by participants. As put by Jergensen et al. (2011) "*high skills, high reputation roles (are) in the center and low technical skills and reputation (are) at the periphery*". However, the denomination of participants and rules employed to define their positions are quite fuzzy. Indeed, hierarchy of participation is indifferently based on technical skills in computer science likely needed to performed development tasks (Mahendran, 2003; Ducheneaut, 2005) or on power to modify the code and finally on involvement in decision-making process (Hedberg & Livari, 2009; Bach & Twidale, 2010).

Going to the center of the onion, progressing in skills, power and decision-making, we can find several categories of participants :

- Those participants "*who do not contribute code but provide use-cases and bug-reports as well as testing new releases* » (Crowston & Howison, 2006). There are called *bug reporter* (Gacek & Arief, 2004; Ducheneaut, 2005), *bug submitter* (Jensen & Scacchi 2007), *active users* (Crowston & Howison, 2006; Hedberg & Livari, 2009; Bach & Twidale, 2010), *contributors* (Gacek & Arief, 2004), *testers* (Antikainen, Aaltonen, & Väisänen 2007). This position is supposed to require less technical skills, less power to modify the code and less weight in decision (as none modification can be directly done to the code). We can add that those users may participate in mailing list discussions, for instance as informants for newcomers (Mahendran, 2003) or as *feature requesters* (Jensen and Scacchi, 2007) and they may also be *documenters* (Ducheneaut, 2005).
  - Participants who are *providesr of patches* (Ducheneaut, 2005), *bug fixer* (Nakakoji et al., 2002), once again *contributors* (Hedberg & Livari, 2009), or peripheral developers (Nakakoji et al., 2002; Bach & Twidale, 2010)
  - *co-developers* « *who submit patches (e.g. bug fixes), which are reviewed and checked in by core developers* » (Crowston & Howison, 2006);
  - *maintener* (Gacek & Arief, 2004; Ducheneaut, 2005), or *module developers* (Jensen & Scacchi, 2007), who are the owner of a part of a code.
  - *commiter* (Bach & Twidale, 2010) or *formal developers* (Barcellini et al., 2008a) who have the rights to commit new code on versioning systems (commit privileges).

Finally, the top of the hierarchy is occupied by participants named after managerial and coordination tasks, power to modify the code and strong involvement in decision-making process:

- *project leader* (Nakakoji et al., 2002; Mahendran, 2003), *initiator* (Crowston & Howison, 2006), member of the *core team* or *core developer* (Ducheneaut, 2005; Gacek & Arief 2004), *administrator* (Mahendran, 2003) refer to community or project *manager* (Jensen & Scacchi, 2007);
- *release coordinator* (Crowston & Howison, 2006) are those "*who contribute most of the code and oversee the design and evolution of the project*". They can be *veteran testers* (Jensen & Scacchi, 2007) as they add to gain reputation to reach this position. Because of "*their technical capabilities and activity, the core team is being respected and having authority to make the decisions related to what to include in the code base*" (Hedberg & Livari, 2009), and have formal rights to modify the code.

The hierarchies presented are not static in time, it is possible for a participant to "migrate" within the hierarchy by gaining reputation on the basis of technical skills (quality of code provided) and discursive skills (ability to engage and to maintain discussion) that participants may show (Mahendran, 2003; Ducheneaut, 2005; Sack et al., 2006; Jensen & Scacchi, 2007; Jergensen et al., 2011).

2.3.Addressing participation in OSSD

This "core-periphery" hierarchy model has been validated thanks to several types of methodologies. Most of them have focused only on contributions of participants in the implementation spaces (modifications of code, bug reports and bug corrections) sometimes using Social Network Analysis (e.g Madey, Freeh, & Tynan, 2002) to reveal code interactions among participants (e.g. Crowston & Howison, 2006; Ripoche & Sansonnet, 2006; Gonzales-Barahona, Lopes & Robles, 2004; de Souza, Froelich, & Dourish, 2005). For instance, they highlight to what extent participants are central or peripheral in bug reporting or code production (Crowston & Howison, 2006; Ducheneaut, 2005). They also reveal a centralization of *code production* since a small proportion of participants (less than 10%) performed 80 to 90% of code (German, 2003; Capilutti et al., 2003 Ghosh, Glott, Krieger, & Robles, 2002; Mockus et al., 2002).

The few studies analyzing participation in discussions space have been mainly focused on the level of interactions: they analyze, for instance, the level of participation through the number of posted messages or they apply Social Network Analysis to corpus of interactions in OSS (Sowe, Stamelos & Angelis, 2006, 2008; Hendry, 2008). As for implementation, results have highlighted that very few participants are in central position and very few participants are high contributors (Sowe et al., 2006, 2008). The latter study shows for instance that one single participant posted over 40% of messages of the considered mailing-

lists. It also reveals to what extent they span different mailing-lists (Sowe et al., 2006), revealing so-called knowledge brokers that span different mailing-lists of an OSSD project (in this case Debian).

However, some others studies addressing forms of participation reveal gaps between the core-periphery model of participation and participants own views on their participation. These studies are mainly based on interviews (e.g. Mahendran, 2003; Gutwin, Penner, & Schneider, 2004; Barcellini et al., 2009), sometimes complementing SNA analysis (e.g. Crowston & Howison, 2006). Indeed, Crowston and Howison (2006) stress that Apache developers argue that non-hierarchical and decentralized structures are preferred and more efficient than a top-down 'cathedral' model of proprietary software engineering, whereas studies review above reveal that hierarchy does exist in OSS communities[7]. In the same line, we highlight that Python community's participants report consistently that distinction between *users* and *developers* is not relevant as a principle to describe participation (Barcellini, Détienne, & Burkhardt, 2010b). There was no agreement between participants on the meaning of the "developer" term. Some interviewees referred to developers as participants in the design-oriented mailing-lists, others interviewers to technical contributors, others to participants that owned rights to modify the code, others to *gurus* or even to "those of members who contribute "*significantly*" to the implementation of the OSS code". The only agreement between participants concerns the status of project leader: all participants recognize his power and legitimacy, linked to his demonstrated expertise and power, in the project.

2.4. Toward an hybrid and a situated approach to analyze participation in OSSD

The core-periphery model is essential to picture a broad view of OSSD project, e.g. what the number of participants taking part to the project is, what are their trajectory of participation in the project (e.g. Ducheneaut, 2005), who are central or peripheral participants in discussions or code production? However, the gaps highlighted above argue for the development of complementary methodologies to understand forms of participation in OSSD design.

Firstly, embracing participation in OSSD requires to take into account all the spaces of actions of participants. Indeed, the core-periphery hierarchy appear to provide a static view of participation and neglect others types of participation, especially those occurring outside the code production task: for instance a documenter can also be an significant contributor, some core developer or bug reporter may be a key participant in use-oriented discussion, so called active users

[7] The real difference in OSS and proprietary software situations may lie in more open coordination practices, i.e. auto-attribution of tasks (Crowston et al., 2007).

may perform collaborative design activities (e.g. Barcellini, et al., 2008a), as we stress below. Following some OSSD‘ scholars, we assume that hybrid methodologies (Ducheneaut, 2005; Sack et al., 2006; Crowston & Howison, 2006; Crowston et al., 2012) combining analysis in both the implementation space and the discussion space are required to embrace participation in OSSD. Moreover, we assume that embracing participation implies to focus not only on what happens inside a space, but also between implementation spaces and discussion spaces, i.e. at the boundaries, as participation in OSSD – and performance of OSSD process - requires several types of actions supported by various tools that support the coordination of the design process and the development of project across these spaces.

Secondly, the data used in studies reviewed above are often undifferentiated, i.e. all SVN or all discussions along a long period of time are considered, i.e. they aggregate various goals and activities that participants may pursue during this period. In this sense, those studies cannot provide evidence to account for collaborative design activities that are grounded in specific situations. This argues for identifying participation corresponding to an engagement of participants around a same goal. This leads to two methodological implications. (1) It implies that data about specific design process should be isolated before processing the analyses, instead of taking undifferentiated corpora with respect to design problems or goals. More precisely, data that we are targeting must deals with collaborative design activities in which participants are engaged to elaborate design solutions and reach a consensus about them. (2) It also implies that a target on interactions between participants is not sufficient to address the complexity of participation and that addressing the content of these interactions – mostly in discussion- is essential to embrace participation.

In the following, we describe the previous works on roles that we propose to use to embrace participation in OSSD.

3. Previous works on roles and profiles to embrace participation in a situated design context

Our approach is grounded in the notion of *roles*, which is used to understand participation in distributed and collaborative design situations. Definitions of roles are related to what actual activities performed by participants are and in which spaces of actions are they occurring. In a first section we present the definition of the three types of roles that we propose by articulating research on collaborative design activities (e.g. d’Astous et al. 2004, Darses, Détienne, Falzon & Visser, 2001; Olson et al., 1992) and interactionist approach of role (e.g. Bales, 1950). These three type of roles are : interactive role, task-oriented role and group-oriented role. In our case, the task in which participants are involved in

collaborative design, research on collaborative designed are thus used to define this role.
This leads us to present and discuss the notion of *profiles* as an aggregation of roles, used to capture the variety of roles performed by participants in different situations (for instance different in time, different in problems addressed by participants…) (Baker et al., 2009; 2003). We outline that few studies are highlighting general forms of participation in design situations (Sonnenwald, 1996), small software design teams (e.g. d'Astous et al., 2001), or group training situations (e.g. Benne & Sheats, 1948), but none is using the notion of roles – as we defined it- to understand collaborative activity in design situations.

3.1.Definition of roles and their dimensions

We define roles as actual and emerging participants' behaviors that appear and differentiate in local contexts of interaction (Baker et al., 2009; 2003; Barcellini et al., 2008a; Détienne et al., 2012). This way, roles are strongly linked to activities actually performed by participants and this notion is distinct from the notion of status, as seen in social psychology, which refers to ranks or formal social positions. In others words, a role is specified by dimensions that characterize "behaviors of participants".
The interactionist approach specifically fits to our objective because it develops a notion of role focusing on small groups of participants engaged in task-oriented collective activities, such as problem-solving tasks (e.g. Bales, 1950 in Oberlé, 1995), which is not so far from design problem solving tasks on which we focus. In this approach, structure of interactions (places of participants in a communication network between all participants, levels of contributions of participants) as well as the content of interactions is used to characterize the roles of participants. We propose that positions of participants in the structure of interactions defined *interactive roles*, whereas analyzing contents of interactions oriented toward the task itself, or toward coordination and support to the group of participants defined respectively *task-oriented roles* (in our case *design-oriented roles) or group-oriented roles* (Benne & Sheats, 1948; Bales, 1950 in Oberlé, 1995), as we described in the following.

Interactive roles: a communication network-based analysis

Interactive roles refer to the level of participation and connectedness of participants in interactions. They are identified through the structural analysis of communications, revealing: *(1)* high vs. low participation characterized by numbers of contributions (messages, turns in discussions) and *(2)* centrality vs. periphery of participation in communications network. These give rise to specific communication network patterns (Figure 2), in which nodes represent participants

and links represent the relationships between them: for instance “who answers to whom” in a face to face discussion.

[FIGURE 2 ABOUT HERE]

Figure 2 Structures of communication network (from Oberlé, 1995)

As illustrated in Figure 2, typical patterns of communications are reported and identified: in the *wheel* pattern –also called *star*- there is one central participant - a *hub*- characterized by a maximum number of links with other participants – each node is linked to this hub and only this; in the *circle* pattern no one is central, each node being linked to only two other nodes; a *chain* represents a continuation of node linked unilaterally to an other one; and finally in a *y* pattern (or fork), several external nodes are linked to a chain by one participant - a bridge. Actually, an edge is said to be a *bridge* if deleting it will cause its endpoints to lie in different components of the network (Wellman and Berkowitz, 1988).

In mediated communication, we use the same type of analyze considering interaction as “who replies to whom” or “who quotes whom”.

Task-oriented roles: Goal-oriented activities contributing to the performance of the design

Task-oriented roles are related to participants’ activities oriented toward production regarding the task i.e. the generation of new ideas, making evaluations or suggestions, requesting or providing of information, or sharing knowledge. Benne and Sheats (1948) describe several roles that fall into this category: *opinion* or *information giver* or *seeker*, i.e. people who evaluate (opinion giver) and provide or search for knowledge to clarify points (information giver or seeker); *elaborators*, those who offer rationale or suggestions, or those who clarify suggestions among various ideas; and finally, *initiator-contributors* which refers to participants who provide the group with new ideas or new ways of regarding group problems or goals.

In design situations these roles are related to the process of solving and evaluating various aspects of design problems, referred to as generation-evaluation activities in design situations (Stempfle & Badke-Schaub, 2002) or in software design situations (e.g. d’Astous, et al. 2004; Détienne , Boujut & Hohmann, 2004; Olson & Olson, 2000; Olson et al., 1992). These activities concern the elaboration of design problems, and the proposal and evaluation of alternative solutions. In our case, task-oriented roles are related to design and will be called in the following *design-oriented roles*.

Group oriented roles: Socially-oriented activities to support and reinforce the team

Group oriented roles differentiate *coordination* roles and *socio-relational* roles.

Coordination roles are characterized by participants' activities oriented toward definition and reformulation of objectives, or synthesis activities oriented towards group management. Benne and Sheats (1948) refer to participants as *coordinators*, the ones who try to pull ideas together and to arrange tasks of various members; or *orienters*, the ones who define the position of the group, summarize and finally *energizers* who prod the group into action or decision, or stimulate the group into performing "high-quality" tasks. In design situations, these activities concern the coordination of people and resources: the allocation and planning of tasks, the ordering and postponing of topics or tasks (e.g. d'Astous, et al. 2004).

Socio-relational roles are associated with activities the objective of which is to facilitate interpersonal relationships i.e. reducing conflict, harmonizing, and seeking a consensus. They concern the creation of a free, expressive and participative context that is essential for group cohesion and learning or for performance, as analyzed in training situations (Bennes & Sheats, 1948). Participants performing these activities are called *encouragers* or *harmonizers*, who praise and agree with the contributions of others, or mediate the differences between members (Benne & Sheats, 1948). As far as we know, these types of socio-relational activities have not been analyzed in design situations whereas there is every reason to believe of their particular interest for achieving design effectiveness. Indeed, in the socio-cognitive perspective of design we consider (Détienne, 2006), group management and socio-relational issues are key components of achieving design effectiveness, i.e. for design solutions negotiation and hability for "the group" to reach a consensus.

3.2. Profiles of participation as combinations of roles in situated design contexts

The notion of *profile* (aggregate of roles) proposed by Baker et al. (2009; 2003) seems appropriate for referring to the various roles describing a participant in terms of behaviors compared with other participants in a specific situation. Indeed, the categorization of roles presented previously is not sufficient to characterize participation, principally for two reasons: *(1)* authors neither refer explicitly to the analysis of activities nor to the context in which these activities are performed, and subsequently links between roles and activities describing these role are implicit. This implies a difficulty in deciding whether a participant belongs to one category or another. *(2)* Most researchers agree that these roles are distributed in small teams and that this distribution is not permanent, i.e. that

participants may have different facets (e.g. Bales & Slater, 1955 in Oberlé, 1995; Nygaard & Handlykken, 1981 in Hendry, 2008; Sonnenwald, 1996). This combination of roles performed by participants is therefore time and context dependent: a participant may perform one or several of these roles according to a situation, e.g. the topic discussed, his/her experience in domains related to situation. Participants are not limited to one or other of these roles and may perform them differently depending on time or context. Some participants may perform roles in a systematic manner, i.e. either taking on the same roles in various contexts or at various times, or not (Baker, 2004).

3.3. General trends of participation in software design situations

In design literature, and especially in software design literature, one can find some descriptions of specific forms of participation, sometimes called roles, even if analyses are often not situated in a clearly delimitated task context and are not targeting actual activities: some are partly based on activity analyses, such as *boundary spanners* profiles, others are based more on peer-recognition mechanisms, such as *leader* or *expert* profiles.

Boundary spanner is one form of participation revealed by software design literature (e.g. Curtis, Krasner & Iscoe, 1988; Sonnenwald, 1996; Grinter, 1999). It has been mostly described on the basis of activity analyses, but not really in situated contexts. *Boundary spanners* are literally persons who span the gap between their organization and external ones (Sarant, 2004). They are identified by "*communication or behavior between two or more networks or groups*" (Sonnenwald, 1996). They move across different teams transferring information about the state of a project/task. This suggests specific design role oriented toward knowledge sharing, coordination and group roles oriented toward group support. They are seen as essential in design situation to compensate for communications deficits between different groups of designers (Krasner, Curtis, & Iscoe, 1987; Sonnenwald, 1996). Despite these examples, as far as we know, few studies have been able to give detailed descriptions of profiles of boundary spanners emerging in situated tasks, neither in a given design process nor in more large-scale distributed teams, such as OSS communities.

There is a more abundant literature identifying general trends of participation based on peer recognition (e.g. de Visscher, 1995): the focus is mostly (sometimes indistinctively) put on the power or the influence attributed to a participant, and rarely on their patterns of behaviors. In this latter case, there are no systematic and explicit reference to the context in which the identification of these patterns has been done and thus no explicit references to activities and roles.

Leaders and *experts* are two examples of this peer characterization of participation (de Visscher, 1995).

The *leaders* are the ones who have *authority* based on a legitimate power. They are not necessarily official managers. Some of the following trends emerge among leaders (de Visscher, 1995):

- high participation and central position in interactions, which could be derived from the characterization of interactive roles as we described below;
- an orientation toward the creation of harmonious work settings and coordination behaviors, which could fit with group-oriented roles.

Experts are recognized by their peers on the basis of their competencies. In software design, experts are referred to as *super-experts* or *gurus* (Sonnentag, 1998; Curtis et al., 1988). Some trends among *super-experts* have been revealed, they:

- participate to a high degree in design meetings;
- are able to integrate knowledge from various domains, in particular design and application domains;
- are provider of support activities for other participants, which could indicate specific design and group-oriented roles;
- have a broad experience in various design projects (Sonnentag, 1998; Sonnentag, 2001).

In a design context, leaders as well as super-experts may have some stability in time, i.e. their position has been established for the duration of a project and is not only linked to a specific moment or situation (Sonnentag, 1998).

General characterizations of participants, tagged as leaders, super-experts or boundary spanners, may not be sufficient to capture the diversity of participants' activities. This point emphasizes the need for a systematic analysis of what participants actually do in distributed and situated design contexts, i.e. their actual roles, in order to characterize their participations.

This must be particularly true in an emerging organizational context such as in OSSD. Indeed, whereas small teams are composed of stable members, online design communities -such as OSS communities- are composed of a large and unstable number of participants whose participation and roles are constructed as they interact. One may expect a more important diversity of roles performed by participants according to the various encountered design situations.

4. Research questions and strategy

4.1. Research questions

We stress that the core-periphery model of OSSD participation may be enriched by the characterization of participation (1) in and between various spaces of actions to understand how participation is coordinated in and between those

spaces and (2) on the basis of collaborative design activities actually performed by participants engaged in a situated design process. Consequently, we address the following research questions:

- How does participation work in and between spaces (implementation, discussion) and their sub-spaces? What types of boundary spaces may be identified and who participates in them?
- What is the distribution of roles among these spaces, including boundary spaces?
- What types of profiles of participations may be identified? In particular, what characterizes roles in the design process?

By addressing these research questions, we operationalize for the first time in OSSD the notion of profile - as specific aggregations of roles - proposed by Baker et al. (2003; 2009). This notion may be a good candidate to enrich the static core-periphery model of participation in OSSD. We also aim at discussing the relation between roles and general trends of participation highlighted in literature (such as boundary spanners, leaders and super-experts), as these trends are often described with no explicit references to activities performed by participants and in which context.

As a first attempt to develop a complementary view of participation in OSSD, we have analyzed collaborative design activities in the Python OSS project performed through the discussions associated with specific design problems – PEP related discussion (Barcellini et al., 2009; 2008a,b). These studies proposed a methodology that associates :

- the analysis of *structure of discussions*, i.e. interactions - positions of participants in discussion reconstructed on the basis of a quotation-linked organization of messages (Barcellini et al., 2008b);
- and *content of discussions* – analyses on the basis of collaborative design activities performed by participants and the nature of knowledge exchanges (computer science knowledge vs. application domain knowledge for instance).

The analysis of interactions between participants (who is quoting whom?) (Barcellini et al., 2005; Barcellini et al., 2008b), provided a sound basis to a communication network. For instance, the project leader tends to be in a closing position (no one quotes his message) or in a beginning position (he/she is quoted by several participants)[8].

We reveal some patterns of actions associated with specific participants in design related discussions (Barcellini et al. 2008b). For instance, some participants that may be qualified as *active users* do perform collaborative design

[8] Contrary to other studies analyzing online discussions, we have used the quotation link rather than the reply-to link to reconstruct the discussions. In a previous paper (Barcellini et al 2008b), we have shown that the organization of messages according to the quoting link is relevant to reconstruct the thematic consistency of online discussions and to understand the interactions between participants in terms of verbal turns.

activities: not only in the bug report phase as suggested by the OSS literature (e.g. Raymond, 1999) but also in design related discussions (Barcellini et al., 2009).

Finally, these studies identified the presence of *boundary spanners* who cross-participate in use and design-oriented mailing-lists on same topics discussions and that held a central position in interactions. They are central in the mediation between use and design that operates in such collaborative distributed design situation (Barcellini et al., 2009; 2008a). These specific participants –mainly core developers - provided and shared knowledge about both the use-oriented application domain and the design-oriented programming domain (Barcellini et al. 2008a). Moreover, their presence seems to enhance the success of design process (Barcellini et al., 2009).

In this paper we aim at extending these studies in two directions: *(1)* by considering the distribution of activities in *all* spaces (implementation as well as discussion) of an OSS project, and not only in implementation space (as in most studies investigating the core-periphery model in OSS) or in discussion space as the previous research (Barcellini et al., 2008) *(2)* by considering each participant as individual and not categories of participants with respect to the core-periphery model (e.g. active users vs. core developers vs. developers).

4.2.Research strategy

An originality of our approach is to explore design participants' roles across all spaces of actions combining structural and content analysis of activities. Since various implementation and discussion sub-spaces of design does exist in OSSD, participation have to be analyzed at the boundary between these spaces.

The remaining part of this section reviews our approach and the main notions relevant to such a framework for situated analysis of role and participation. Then, we summarize the research questions.

Four types of roles characterizing participation in OSS context

Our perspective on roles aims to account for emerging activities of participants in situated, i.e. localized interactions. Thus, these roles are semantics (i.e. problem-content), time and context dependent. In order to capture richer patterns, we will use the notion of profile: combination of roles in a specific context. Analyzing these profiles should also help to analyze the profiles of key-participants, i.e. participants recognized as leaders, or super-experts by their peers.

Taking into account dimensions of behaviors defining roles (interaction, design, group) is essential to construct profiles. These dimensions have to be defined regarding the context of the research: goals of participants engaged and spaces of actions. In our case, participants can act in two main spaces (implementation and discussion), which defined implementation role or discussion role. In discussion,

participants are engaged in interactions that support collaborative design activities. From this, we define specific design-oriented and group-oriented roles (in addition to an interactive role). This leads to consider four types of roles:

- *the implementation role*, i.e. as revealed on the basis of actions actually performed by each participant (code revisions or commits, contributions to specifications…) that contribute specifically to the design process in the *implementation space* (i.e. documentation and/or coding spaces) ;
- *the discussion roles*, i.e. as revealed by collaborative design activities performed in discussions related to the design process considered (*discussions space*). There are three types:
  - *interactive role* linked to position in a network of communication and level of participation;
  - *design-oriented role* performed through cognitive and epistemic activities related to the definition of design solutions (generation-evaluation, clarification…);
  - *group-oriented roles* linked to coordination and socio-relational activities involved in collaborative design (coordination, acknowledgment).

Investigating participation in boundary spaces

The spaces of actions embrace potentially multiple intersections and boundaries (illustrated by arrows on the figure 2). Moreover, studying participation in OSS design also refers to studying distribution of this participation between some use and design spaces, i.e. spaces in which issues regarding design and/or use (elicitations of requirements for instance) are processed. This question is of particular interest in OSS communities, since the divide between use and design is supposed to be less strict than in "traditional" design organization.

Intersections in participation between various spaces define what we have called *boundary spaces*: a boundary space is a virtual space created by the fact that participants act in parallel on the different sides of boundaries between two, three, or four spaces for the considered design process.

As illustrated in Figure 3 by the "?" sign, several types of boundary spaces may be defined inside discussion or implementation spaces. These spaces embrace potentially multiple intersections and boundaries (illustrated by arrows on the figure).

[FIGURE 3 ABOUT HERE]

Figure 3 Distribution of the OSSD process in the 2 spaces of actions and their boundaries

Inside the implementation spaces, a "document-coding boundary space" may be defined. In the discussion spaces, sub-spaces dealing with design or use can be bounded by the usual two major mailing-lists in an OSS project: one is the use-

oriented mailing-list (python-list in the case of the Python project) and the other is the design-oriented mailing-list (python-dev for Python). Thus, we consider that participants posting to the use-oriented mailing-list are acting in the use space and those posting to the design-oriented mailing-list are acting in the design space. Consequently, participants posting in both two mailing-lists are working in a "design-use boundary space".

5. Method

Our method is divided into four steps aiming at:

(1) Identifying a corpus containing data of a situated design process. It implies:
   a. selecting a situated design process (a PEP - Python Enhancement Proposal- process);
   b. constructing the corpus by searching and selecting for actions in discussion space (messages posted on some Python mailing-lists on a situated design process) and implementation space (commits or revisions of code and document related to the situated design process)

(2) Selecting some relevant data (especially discussions) in the corpus, organizing these data among spaces and time, and placing them in a global PEPs ' scenery and identifying participants in the design process.

(3) Identifying roles and profiles:
   a. developing a coding scheme to analyze roles based on collaborative design activities and actions of participants and
   b. defining rules used to construct profiles of participation.

(4) Discussing and validating our findings with participants in the selected design process through interviews.

5.1.Construction of the corpus

Identification of situated data of design: selection of PEP process

To approach activities, we have selected data regarding interactions related to particular design issues – *Python Enhancement Proposal* (PEP) - of the Python project, across the various spaces (discussion and implementation).

Focusing on PEP processes makes possible for us to consider units of corpora distributed in spaces of actions, as well as distributed over time in relation with the different formal steps of a design process: from needs' elicitation (identification and proposals of new functionalities), specification, and finally to implementation and evaluation of software. Among these data, our interest is in collaborative design activities that can occur at boundaries between spaces. These

activities may appear at different phases of a design process, but more likely during need's elicitation, specification or after evaluating functionality. This focus will help us to reduce the PEP related data of the project and constitute the corpus of data of our research.

To identify a PEP candidate, we conducted 13 semi-structured interviews with participants in the Python project (11 "users" of the Python language, the project leader and one "developer"). Questions asked can be categorized according to two different objectives: *(1)* the first one refers to a previous study -out of the scope of this paper- dealing with characterization of mechanism underlying construction of social awareness in the Python community, a full report can be found in Barcellini et al. (2010b); *(2)* the second objective was to identify - among all design proposals of the Python project - one that can be of particular interest for our research. As the Python community uses the PEP process to formalize a new design proposal, interviewees were asked which PEPs had been most significant for them, or for some other users in the community according to them.

The interviews were then transcribed and one design proposal transformed in PEP (PEP 327), i.e. a successful design proposal, was outlined among the 161 PEPs of the Python project at this time.

PEP 327 concerns the introduction of a decimal type in Python and its related decimal.py module. Before PEP 327 was accepted, there have been several unsuccessful proposals, i.e. not scored by an accepted PEP and its implementation, to introduce a decimal type in Python. A first decimal.py module was also proposed but was still to be implemented. Then, a successful proposal was initiated by a user of Python, whom we will refer to as the user-champion, who is a developer of a project in a financial application domain of Python (http://sourceforge.net/projects/sigefi). The first need of the user-champion was for a money type, but it appeared that before introducing a money type, work had to be done in the decimal type in Python. The user-champion formalized this proposal through a PEP document that became an accepted PEP (PEP 327). This successful design proposal was selected for several reasons: *(1)* the champion[9] of this PEP was a relatively newcomer in Python (we shall refer to him as the champion) *(2)* the PEP is related to one very clear end-user application domain of Python (financial application). We assume that this process would be relatively "open" for a variety of participants and for design/use discussions at boundaries.

Collection of actions in the implementation and discussion spaces

We collected three types of *actions* during the entire PEP 327 design process in the three spaces : *(1)* messages related to PEP 327 design process in the two main mailing-lists of Python (*actions in the design and use discussion spaces*), *(2)*

[9] The initiator of the discussion

commits or revisions to the code of the module 327 implemented in relation to this PEP (*actions in the coding space of implementation spaces*), and *(3)* revisions of the PEP document (*actions in the documentation space of the implementation spaces*). The collection was performed from the first formal attempt to introduce a decimal module (May 2001) to the beginning of this study (May 2006).
These actions and the way they are collected are described in the following:

- *Actions in the discussions spaces* are messages posted by a participants in all discussions related to a PEP 327 in the two mailing-lists considered (*python-list* and *python-dev*). *Python-list* is a mailing-list dedicated to general issues when using Python and *python-dev* is the list where key Python developers discuss the future of the language and its implementation. Topics include Python design issues, release mechanics, and maintenance of existing releases (see http://mail.python.org/mailman/listinfo/python-dev/ or python-list for description and archives of the mailing-lists). The messages were gathered by searching - by hand – on python-list's and python-dev's archives for the following keywords in the header of discussions : decimal, money, currency, PEP 327 and the name of the user-champion.
- *Actions in the implementation space* are commits or revisions of code or revisions of the document describing the new functionality proposed by the PEP (PEP document). Two types of actions are considered: *(1)* indirect revision is when the author of a revision explicitly mentions that the revision is based on the work of other participants - this may occur when a participant does not have the formal right to modify the code by himself (i.e. he/she is not a formal "developer"); *(2)* other revisions are called direct revisions, i.e. made on purpose of the author himself/herself. These revisions are tracked in the versioning system repository of Python (svn.python.org) [10]:
    - Coding (commits or revisions of the code of the 327 module). The module 327 have been implemented in relation to PEP 327 design process. We search for commits or revisions of this module and their authors in the related repository (python/trunk/Lib/decimal.py).
    - Documentation (revisions of PEP 327 document). The revisions of documents PEPs are also tracked by a SVN repository on svn.python.org. We search for modifications dealing with the PEP 327 documents and their authors in that repository.

[10] Subversion systems that trace all revisions of code, their content and authors.

5.2.Description of the selected corpus

Description and chronological organization of the PEP 327 corpus

The collection method described below leads us to construct the following corpus of data:

- *Discussion spaces* : the complete data obtained correspond to 51 discussions and 746 messages posted by 130 participants. As we described in the following section, the content of messages of some selected discussions have been analyzed to identified roles of participants.
- *Implementation spaces*
    - Coding: 44 commits or revisions of module decimal.py made by 7 authors.
    - Document: 9 revisions of PEP 327 document made by 9 authors.

To have a broad view of how the process is organized, a simplified view of the chronology of actions and discussions in the corpus across the spaces has been constructed (Figure 4). To simplify representation, this figure does not represent each action (each discussion, each revision) but we group together actions (discussions or commits/revisions). These groups are symbolized by dots and are represented in parallel in the three activity spaces according to time. *Actions* in the discussion spaces (discussions in design or use-oriented mailing-lists) are grouped together according to formal design phases they spread (specification of the proposal, refinements, promotion or evaluation), *commits/revisions* of code (implementation coding) are grouped together when they occur in a same period of time relatively to the whole process (i.e. the same month). As there are few revisions of documents, we do not group them together and all are displayed. Extensive description of this process (with display of all actions) is given in Barcellini et al. (2009).

[FIGURE 4 ABOUT HERE]

Figure 4 Simplified overview of PEP 327 process among the spaces of actions (adapted from Barcellini et al., 2009)

Actions and discussions considered to analyse roles

The fine coding of the 51 discussions would have been very costly, so we choose to reduce the corpus considered and focus on several discussions to apply our methodology. Two rules linked to our research questions have been applied to select discussions: *(1)* we need to identify how roles can be distributed between design (python-dev) and use (python-list) spaces; *(2)* we are interested in

discussions in which collaborative design activities occurs (proposing and evaluating solutions, reaching a consensus).

(1) Among the 51 discussions, some are occurring only in *python-dev* (e.g. a group of discussions called *refinements*), others occurred only in python-list (e.g. the group of discussions called *valorisation)*, and very interestingly some discussions (5) with the same topic (same header of messages) are occurring simultaneously in both lists (these group of parallel same-topic discussions are symbolized by a vertical link in Figure 4).

(2) In a previous research we have already studied python-dev discussions (see Barcellini et al., 2005; Sack et al., 2006; Barcellini et al, 2008b) and in this case, python-dev discussions dealt mainly with refinements of design choices made previously (in parallel discussions). Finally, we assume that python-list discussions will be less focused on design (indeed they dealt with the "valorization" and diffusion of the *decimal.py* module),

We assume that those 5 parallel discussions – occurring simultaneously- would be of particular interest to identify broad forms of participation (involving well-knows developers of the project and other participants) and to understand what is occurring at boundary (in this case between design space and use space), compared to discussions occurring in one or another list. Four of the 5 discussions occurred at the beginning of the process (called *specification* phase on Figure 4) in which collaborative design activities may be concentrated and two discussions during a phase called evaluation, that lead to transform *decimal.py* module (Barcellini et al., 2009).

Additionally, all the corresponding implementation actions (revisions in the documentation space and actions in the coding space) were collected.

The size of the corpus on which our research is based is described in [FIGURE

5 ABOUT HERE]

Figure 5.

[FIGURE 5 ABOUT HERE]

Figure 5 PEP 327 corpus analyzed in the discussion and implementation spaces

Position of PEP 327 in the whole PEPs ' scenery

Our purpose was to select a proposal – a case- in which use and design can "confront", for which design-use boundary spaces may exist. We can place the selected corpus in a broader picture of PEPs in Python during the period of our analysis (Barcellini, 2008). At that time, 161 PEPs proposing new functionalities

had been discussed. During 2004 (year when occurs the majority of discussions analyzed), 36 PEPs were discussed on *python-dev* and 25 on *python-list*, and 7 on both lists, one of theme being PEP 327.

Concerning participation in the five selected discussions, discussions related to all PEPs (still in 2004) involved: on *python-list* 168 different participants posting 699 messages; on *python-dev* 159 participants posting 1579 messages[11]. Having said that, the five discussions finely analyzed in our research concern (see Barcellini et al., 2008b):

- 31% (42/168) of all participants involved in PEP-discussions in python-list and 13% (21/159) of all participants involved in PEP-discussions in python-dev in 2004;
- 25% (173/699) of all PEP messages posted in 2004 in python-list and 3% (46/1579) of all PEP posted at that time in python-dev.

In sum, we can say that the specificity of the selected corpus is that: *(1)* it encompasses a significant population and discussions involved in PEPs process at the time of analysis *(2)* it generates more discussions than other ones and *(3)* that it is more discussed in the use space than other ones, which confirms that it is a good candidate to analyze collaborative design activities at the boundary between design and use.

Key-participants initially identified in the selected corpus

Among the participants in the selected discussions and/or collected implementation actions, there are some that can be initially *identified* on the basis of their statuses in the project (Mahendran, 2003) and/or their recognition by their peers, as revealed during interviews (Barcellini et al., 2010b). These socially visible participants present in the PEP under study are:

- The project leader, labeled GvR. He is the founder of the project and he is called the BDFL, the Benevolent Dictator For Life. He is the guarantor of the project follow-up, and is very well respected in the project;
- The champion of the design proposal (labeled FB). FB is the person who proposed the new functionality and wrote its related PEP document. At the beginning of the discussions, he was a "user" coming from a financial application domain. He would later acquire the status of developer, i.e. would gain the rights and login information to directly modify the source code in Python repositories;
- Four participants identified as super-experts of the project thanks to interviews performed in a previous study (Barcellini et al., 2010b). These participants (RH, TP, Aa, AM) are well-known developers. They all have participated since the early beginning of the Python project. These participants are super-

[11] For the record, broad traffic of these lists is as follows: python-list (5370 participants posting 51495 messages); python-dev 451 participants posting 8955 messages)

experts in specifically relevant areas, for instance TP is an expert in scientific calculation, which is an area of expertise dealing with the considered PEP.

5.3.Coding scheme of roles and construction of profiles

Coding scheme

Analysis of roles of participants was made by the way of complementary approaches (structural and content analysis) to investigate participation in the discussion and implementation (**Erreur ! Source du renvoi introuvable.**). A full description of the coding schema used to analyze discussions can be found in Barcellini et al. (2008a).

In the *discussion spaces*, we performed:

1. *a structural analysis* to address interactive roles and to enable the construction of the network of communication between participants. The network constructed is based on a citation link between each participant. To construct the network, we follow the steps described below:
   - Citations between participants are identified "by hand": within the texts of the messages, quotes are usually identifiable according to a simple syntax: greater-than (">") symbols precede each quoted line. But this syntax is not always reliable and so, for our analysis, we have also compared messages "by hand" to identify the message(s) quoted, its authors and authors quoting.
   - An authors*authors table is then generated, in which: lines are authors citing, columns are authors being cited, and cells contains the numbers of citations.
   - A graph is finally constructed (by hand using a presentation software) on the basis of this table. The occurrences of citations are represented by a link between participants (node), wherein the thickness of the link symbolizes the strength of their relationship in terms of number of citations (see for instance [FIGURE 10 ABOUT HERE]
   - Figure 11). For a better readability, our citations graphs will be based only on a number of citation superior or equal to the third quartile (Q3) of the distribution of the number of citations over all participants (Q3=2 citations). This choice excludes 19 participants who quote each other only once, and 71 links between participants. A complete graph including participants who quote or have been quoted by others' only once can be found in appendix 1.
2. *A content analysis* of messages to address design and group roles. The corpus obtained (each message of the 5 selected discussions) is segmented into units, corresponding to comments. All messages contain sequences of

quotes (from one or several previous messages) and comments. On the basis of our previous study (Barcellini et al., 2005), we consider that a comment corresponds to an individual turn in an online message. In relation to the definition of design-oriented roles and group-oriented roles, we code for (Table2) :

- collaborative design activities, reflecting the functions of the turns in the design discussion (e.g. making a proposal, a reformulation, an argumentation).
- activities related to group management (coordination, decision, synthesis) or interpersonal relationship management (acknowledgment of others' work, support).

Each comment or sub-unit of a comment (if there are several activities in one comment) is categorized using this coding scheme : collaborative design activities or group management activities is associated with each comment or sub-unit of comments.

In the implementation spaces, we performed an analysis focused on technical actions (revisions or modifications of documents or code[12]) in order to address implementation roles (Figure 6). We distinguish between direct and indirect actions, i.e. actions performed on behalf of someone, who may not have the rights to act in the implementation spaces.

[FIGURE 6 ABOUT HERE]

Figure 6 Coding scheme of roles

Construction of profiles

Profiles have been constructed by the aggregation of specific roles performed by each participant relatively to activities of other participants for a situated design process. These comparison of participants supposed certain types of metrics to characterize the role performed, i.e. what characterizes a participant within the group. Two kinds of metrics are used to construct participation profiles: intensity of participation and coloration of participation of each participant (tendency to perform one or another role).

[12] These commits and revisions are one part of the implementation space: we do not consider actions into issue tracking systems, as the Python project did not use it at the beginning of the process. Moreover, we do not consider branches in the decimal.py modules.

Intensity of participation in implementation and discussion spaces

Intensity of participation aims at characterizing the implication of a participant in a given space. It is evaluated by comparing the number of actions in the implementation and discussion spaces of one participant with the distribution of actions among all participants.

For *characterizing implementation role*, actions considered are revisions of code or document. A participant is considered to be:

- *a high* contributor in implementation space or in discussion- if his/her number of actions in this space is ≥ Q3, i.e. is in the fourth quartile of the distribution (i.e. the participant is one of those who contributes up to 75% of actions);
- *a low* contributor in implementation space - if his/her number of actions in this space is ≤ Q1, i.e. is in the first quartile of distribution (i.e. the participant is one of those who contributes to less than 25% of actions);
- *a mean* contributor in implementation space or mid-central participant in discussion- if his/her number of actions is between the former high and low thresholds. In other words (the participant is one of those who contributes to more than 25% of actions but less than 75% of actions).

It is important to note that the "high", "low" and "mean" terms refers to the quantity of contributions and not the quality, for instance the potential impact of a given contribution on the design activity and/or solutions.

For characterizing *interactive role in discussion*, the actions considered are the number of interactions of one participant with other participants; i.e. a participant citing and being cited by another one. This metric has been preferred to a number of messages for characterizing the interactive role because it also gives information about the position of the participant in discussions (centrality/periphery). Using the same criterion as for actions in the implementation, we distinguish between :

- *central* interactant (number of interaction ≥ Q3) ;
- *mid-central* interactant (Q1< number of interaction <Q3);
- *peripheral* interactant in discussion ≤ Q1).

Coloration of participation in discussion spaces

Coloration of participation is used to characterize task or group-oriented roles. It aims at characterizing the tendency of a participant to perform one or another of these roles. For example, if one participant exhibits more design-oriented activities than all participants on average, he will be characterized as high for these activities. *Coloration* is determined on the basis of a contingency table in which lines are authors, columns are roles (design or group-oriented) and cells contains the number of activities performed by participants (related to design or group-oriented roles) in the discussion space.

Coloration is evaluated on the basis of relative deviation (RD) of this table. RD is a descriptive statistic used for Contingency Tables. RD can be used when more than the two third of the cells in the contingency table have frequencies superior to zero (Rouanet, Leroux & Bert, 1987)[13]. RD highlights how each modality of a categorical variable in a column is associated with each modality of another categorical variable in a row. Thus, RDs are used here to reveal strength of associations between variables that may be difficult to outline in simply analyzing a contingency table (of activities for instance) (Bernard, 2003). They are calculated for each pair of modalities from the two variables on the basis of a comparison between observed and expected frequencies (i.e. those that would have been obtained, if there was no association between the two variables). There is an attraction when the RD is positive, and a repulsion – when it is negative.

This metric has been preferred to an evaluation of intensity of performance of a given activity, because profiles of participants are aimed at characterizing the specificity of participants – referred to as *coloration*- compared with other participants.

By convention, we have considered that :

- there is a *high positive association* (respect. negative) between two modalities of the variables when absolute RD value is ≥ 0,2 in other words, a participant tends to perform this role noticeably more (respect. Less) than other participants;
- there is no preferential association for -0,2<RD<0,2, in other words the participant is in the *mean* of the group, he/she does not perform the activity more or less than other participants.

Construction of profiles

Profiles are thus characterized by the four roles, their intensity and coloration as described in Figure 7(the same color code will be used in the results sections). To categorize profile, we then applied the following criteria: the spaces and boundary spaces in which participants are acting; the centrality in interactions; the coloration of participation (design-oriented, group-oriented or both).

[FIGURE 7 ABOUT HERE]

Figure 7 Synthesis of rules used to construct profiles of participation

[13] This requirement is filled in our case as we have 42 cells (21 participants vs. design and group-oriented activities) and 4 “0”.

5.4.Confrontation of analysis

To complete the previous analysis, we conduct interviews with participants in the PEP 327 process. All the 58 participants were appealed to and four answered positively to our request (the user-champion, the project leader, one super-expert and one developer).
Three interviews were e-mail-based and consisted in three to four e-mails exchanges (1500 to 2000 lines of text for each interviews), which were limited by participants' availability. The project leader was interviewed face-to-face (interview duration = 40 mins).
All interviews were composed of two parts: the first part dealt with general questions described in section 4.1 (identification of situated data of design); the second dealt with PEP 327 process. Interviewees were asked to retrace the history of PEP 327 process, to describe their own forms of participation and the other ones. The data gathered is used to interpret our findings.

6. Results : From participation and roles to participants profiles

In a first section, we characterize participation within each space and at the boundaries between spaces; in a second section we focus on the distribution of roles within the discussion spaces. The last section is a synthesis of roles performed by participants in order to stress specific profiles of individual participants.

6.1.Participation at the boundaries during the process

We analyzed the distribution of participants in the implementation spaces (coding and documentation), in the discussion spaces and between these spaces (boundary spaces). [FIGURE 8 ABOUT HERE]
Figure 8 represents the number of actions performed by participants in the implementation spaces (revisions of code and of PEP document) and in the discussion spaces (messages posted in the use-oriented or design-oriented mailing-lists) during the whole PEP 327 process (all the 51 discussions considered). To ensure readability, we only display in Figure 4, the 4th quartile of participants in terms of posted messages in the discussions spaces (corresponding to 12 participants (out of 47) posting 311 messages (out of 405) in the design space, and 24 participants (out of 95) in the use space posting 258 (out of 340). For the documentation space, we represent the two participants.

[FIGURE 8 ABOUT HERE]

Figure 8 Distribution of participation in the discussion and the implementation spaces for the PEP 327 corpus (only the 4th quartile of participants in mailing-lists is represented, excepted for documentation)

As illustrated by [FIGURE 8 ABOUT HERE]

Figure 8, no participant contributes simultaneously in every space (coding, documentation, design and use). Most participants in the PEP 327 process are pure discussants evolving either in the design or in the use discussion spaces, as only 7 participants acts in the implementation spaces (7 different participants, 6 in the code production and one added in documentation). Moreover, there are no pure coders –all participants in the coding space act in another space- and a few participants in documentation, which is classical in an OSS project (Mockus et al. 2002; Lakhani & Von Hippel, 2003).

Only 9 participants contribute *"at the boundary"*, i.e. in two or more spaces. Participation in these boundary spaces are represented in Figure 10 which presents a synthetic map of participation in the various spaces. All participants are represented by their initials. As in [FIGURE 8 ABOUT HERE]

Figure 8, only the fourth quartile of participants in terms of posted message (24/95 participants in use; 12/47 participants in design space) are represented, in all the spaces they act in[14]. If a participant posts a message in python-list or python-dev only, he/she is represented respectively in the use or design space (e.g. MC or BR). If he/she posts in both lists, he/she is represented at the boundary between design and use (Aa, AM, SH). If a participant made a revision on code, he/she represented in the coding space. Actually, all the participants making revisions of code also post in one or another list, they are positioning at the boundary between the coding space and the considered list (e.g. NC is making revisions of code and posting in python-dev).

[FIGURE 9 ABOUT HERE]

Figure 10 Map of participation in the discussion and implementation spaces

This way, Figure 10 reveals five trends of participation at the boundary: (1) *design-use and coding boundary space* (occupied by FB the user-champion and TP a super-expert) ; (2) *design, coding and documentation boundary space* (with only RH, a super-expert); (3) *design-use boundary space* (with two super-experts Aa, AM, and a "user" SH); (4) *design and coding boundary space* (GvR and NC); (5) *Use and coding boundary space* (one "user" RB).

Participants in boundary spaces adopt different types of activities, suggesting a complementarity in their roles: some are only present in the discussion space (Aa, SH, AM) others are also present at the boundary with coding (FB, TP, RH, GvR, NC). These participants may be distributed between spaces on the basis of their competencies, availability or interest (e.g. interest in code or not, interest in some specific techniques, interest in animating mailing-list). For instance, participations

[14] Data represented are the same data than in figure 4.

of the user-champion and TP in *design, use and coding spaces* can be explained by their positions and history in the project: FB, the *champion* exhibited a strong implication during the entire process[15], in line with what can be expected from the status of champion: "championing" his proposal and helping the participants in reaching a consensus. TP[16] is a major contributor especially in the design space (whatever the PEP process). Indeed, TP is a super-expert on the technical side of this design process (he was the developer and maintainer of one of the solutions that was used previously by the community to address the problem solved by the new functionality introduced by the PEP). His activity in the discussion may reflect his wish to support this new proposal. A similar explanation can be proposed to explain Aa's presence (in the design-use boundary space) as he was the "owner" of another previous solution, but interestingly he is not present on the technical side. Moreover, the strong participation of RH in the coding space[17] is confirmed by interview with the user-champion who recognized that RH provided "*strong technical support and mentoring*" to the user-champion proposal, while remaining in the background (coding space), such as AM was supporting him in the design-use discussion space. These both supports add an impact on acceptation of this PEP (Barcellini et al., 2010b). Finally, the project leader is present *at design and coding* boundary space, the relay with the use space being ensured by other participants (for instance TP), as he is not an expert in the problem addressed as he stated during interviews (Barcellini et al., 2010b). Indeed, GvR participated far less in this PEP process than in the other ones we studied (Barcellini et al., 2009). Indeed, he posted around 5% of messages (19/405) in the design space compared to 20 to 25% of messages in other Python design processes.

In sum, the observed patterns of participation in boundary spaces show:

- that identified key-participants contribute to create boundary spaces (TP, Aa, AM ,RH and GvR) and the user-champion (FB). Indeed, they make up 6 over 9 participants in the boundary spaces.
- but that there is a room for participation for other participants *as* simultaneously, 3 other participants had an activity at boundaries (SH, NC, RB), which suggest that.

This confirms that crossing analysis in several spaces of actions is relevant to embrace participation, as what is occurring in the implementation space is only one part of the OSS design process. A following issue is to more finely characterize what the contributions of some participant or other are: i.e. what their roles and profiles are.

---

[15] He posted around 25% of messages in both lists and performed 9% (4/44) of code revisions.

[16] He is the 2nd poster just after the champion (52/405, 13% of messages) and the third contributor in the coding space (1/44, 2% of modifications).

[17] RH contributes to 77% (34/44) of all the code revisions, and he is the fourth contributor in the design space (7%, 27/405 messages).

6.2. Distribution of roles and participation between and across spaces: A focus on the use discussion space and the design discussion space

In order to examine what is happening in and between the use and the design spaces, we investigate roles when parallel discussions occurred about same topics. Consequently, this section reports on the content analysis of the 5 discussions about PEP 327 process that occurred simultaneously in the use discussion space and in the design discussion space.

Distribution of interactions: "y-forks" structured around a central hub

To refine the description of the distribution of participation in the discussion spaces, we look at reciprocal citations between participants in the previous five same-topic discussions ([FIGURE 10 ABOUT HERE]
Figure 11), in both the design (top left of the figure) and use discussion spaces (bottom right if the figure), as well as at the boundary between these discussion spaces (middle of the figure).

This figure embraces the distribution of interactions between the group of participants involved in the design and use discussion spaces (corresponding to the design and use-oriented mailing-lists) and at the boundary between these spaces (*design-use* boundary space). Participants are symbolized by circles, except the project leader (triangle) and the champion (star). The three sizes of symbols are representing high, mean and low intensities of participation (in terms of number of activities performed)[18]. The arrows represent the occurrences of citations between participants, the intensity of their interaction being represented by the thickness of the arrow[19]. On [FIGURE 10 ABOUT HERE]

Figure 11, we represent the upper quartile of citations over all individuals (Q3= 2 citations; the distribution considered is a matrix of "who is quoting whom?"). This choice excludes 19 participants who quote each other only once. Interested readers may refer to [FIGURE 13 **ABOUT HERE]**

Figure 15 in appendix to have a picture of all participants and their interactions. This complete graph (elaborated "by hands") remains very difficult to interpret being very fuzzy, which explains our choice to represent only main trends of interactions neglecting single ones.

[FIGURE 10 ABOUT HERE]

---

[18] Small size for the participants involvement corresponding to the 1st and 2nd quartile (0 to 14 activities), mean size for the median quartile (15 to 87 activities) and larger size for the 4th quartile (equal to more than 88 activities).

[19] Only the relations greater or equal to the 3rd quartile of the distribution of citations are represented (75% of discussions). The smallest arrows represent 2 mutual citations between participants, to represent superior strength we use arbitrary thickness.

Figure 11 Distribution of interactions (a) and activities (b) of the fourth quartile of participants PEP 327 corpus

[FIGURE 10 ABOUT HERE]

Figure 11 reveals the structure of the network of communication between participants:

- Six out of 20 participants (FB, AM, TP, AA, JR and IB) have a "high" interactive role, as we defined it in Figure 7. They occupy a central position in the network of interactions, which refers to a bridge position from a SNA perspective. Four out of these six are in the design-use boundary space (FB, AM, TP, AA) - they are all identified key-participants (super-experts and the champion). The two remaining participants were acting in the use space (JR and IB) and were not initially identified key-participants;
- Four participants (SH, RA, BR and MW) are mid-central, i.e. they have a mean interactive role and they interact with two other participants. SH is at the boundary space (he is the only participants who is not an initially identified key-participant) and the others are in the use space;
- Finally, about half of the participants (9/20) are peripheral in the network of interactions, i.e. they have a "low" interactive role as they interact only with one other participant.

Consequently, the network of interactions can be structurally characterized as a *central hub held by the champion (FB) and a set of forks organized around the super-experts*.

The champion (FB) holds indeed a central position (central hub) as three out of the four other participants in the design-use boundary space interact highly with him and do not interact with each other. He personally interacts with 11 out of the 19 participants considered (72/124, 60% of all citations). The principal axis of interactions is composed by AM and FB interacting strongly together (28/124, 23% of citations). This strong interaction has been confirmed by stakeholders during interviews (Barcellini et al., 2010b) in which AM explained that he provided a strong support to FB. Both of them are also high contributors to the discussions: the champion contributing to 26% of activities (320/1194) and AM to 19% (224/1194).

Around this hub, interactions are structured by several poles, "y-forks", each led by one participant, who appears to be a super-expert and a boundary actor evolving in the design-use space:

- A first pole is constituted around the champion. He is the only one linking the design-use boundary space with the design space since he is interacting with two design discussants. Furthermore, these two discussants do not interact with each other. The champion interacts less directly in the use space (24/72, 33% of champion citations) and more in the design-use boundary space (45/72, 62% of champion citations);

- The others poles link the boundary space with the use space. There are three simple "y-forks" (structured around AA, SH and TP). A fourth structured pole around AM is more complex – mixing forks and circle- since there is no unique central position: participants interact with AM but also between them, even if a principal axis seems to be composed by AM and JR. We can assume that a content analysis of discussed themes would reveal that these poles are structured around specific design themes.

In synthesis, analyzing distribution of interactions shows:

- The centrality of the *design-use* boundary space for these discussions. Indeed, there are only two participants in the strict design space whereas the five other participants in the design discussions are also in the use discussion, i.e. design-use boundary space.
- That communication between design and use spaces are supported and distributed among the super-experts: all of them, except the champion, interact with the use space around different poles. These poles can be seen as "y-forks": relations between external participants appear to be exclusive and the forks are linked to the *design-use* boundary space by super-experts (being *bridges*). As we outlined previously (Barcellini et al., 2008a), these poles relay information and knowledge from the use domain to the boundary space.
- That the champion provides a relay with the design space that facilitates knowledge exchanges between use and design. Additionally, our results clarify that super-experts and the champion are not only present in various boundary spaces but have also an activity to structure interactions inside of the discussion spaces.

The importance of partially re-localizing use and design discussions in the boundary space could be explained by the fact that the discussions occurred at the beginning of the design process in which the use community may be more involved. Indeed, our previous studies showed that discussions remained concentrated in the design space in the following steps of the design process (Barcellini et al., 2009).

Distribution of design-oriented roles and concentration of group-oriented roles

[FIGURE 11 **ABOUT HERE]**

Figure 11 refines results of the previous section by outlining how some types of roles (i.e. design-oriented and group oriented) are distributed among participants in the discussion spaces. As described in Figure 7, colors represent tendency of participants to perform activities: black is used to visualize participants that tend to perform design-oriented or group-oriented roles more than other participants; white is used to visualize participants who tend to perform design-oriented or

group-oriented roles less than other participants and grey is used to visualize participants with no specific tendencies. This color code combined with the size of the symbols described for Figure 6, gives an indication of the strength and the coloration of participation (design-oriented or group-oriented).

[FIGURE 11 ABOUT HERE]

Figure 11 Distribution of roles among participants

[FIGURE 11 **ABOUT HERE]**

Figure 11 highlights the fact that design-oriented roles are distributed between all participants whereas group-oriented activities are concentrated by few participants.

Indeed most participants (18/20, i.e. 90%) are contributing to design-oriented activities at the same ratio. Only two participants –one in each space- (GvR and RA) avoid design-oriented activities. It is worth note that design-oriented activities appear here to be an irrelevant criterion for constructing profiles of participants since all participants are performing these activities: these activities are not useful to discriminate between participants.

On the contrary, group-oriented activities are concentrated on one participant at the design-use boundary (FB) and five participants (5/20; 20%), two in the design discussion space (GvR, MC) and three in the use discussion space (MW, RA, El). Only the champion (FB) tends to perform group-oriented activities at the boundary, these activities being relayed by other participants in the use and design spaces.

6.3.Synthesis: Individual profiles of participants

Profiles of participation may be outlined regarding the space in which participants act, their centrality in the interactions and activities which they tend to perform more than other participants (*coloration*) (Figure 13).

In Figure 13, individual profiles are organized according to the spaces in which participants are acting: at the boundaries (implementation-design-use boundary, or implementation-design boundary), or in the design or use space only; for the record there are no participants acting in the implementation space only. Intensity and coloration of participation are then visualized according to the color defined in Figure 7:

1. Black is used to represent “high” contributors according to the four dimensions of roles (implementation, interaction, design and group). We represent a “+” sign as a reminder that black is for representing an *attraction* to perform design or group activities.

2. Grey is used for mean contributions. We represent a "=" sign as a reminder that grey is for representing *no tendency* to perform design or group activities more than observed on average across all participants;
3. White is used for low contributors. We represent a "-" sign as a reminder that white is for representing a *repulsion* to perform design or group activities

To define each participant's profiles, we apply the following criteria considering only high participation as the main specificity of participants compared to other ones:

(1) Firstly, we consider two specific positions in the network of interactions: the central position in interaction (*hub)* and the edge positions (*bridges*[20]*)*. Because these positions may be adopted within a space, i.e. locally or at a boundary space, we will thus refer to as *boundary hub*, *boundary bridge* or *local bridge;*

(2) Secondly, we color this position depending on the discussant activity that a participant tends to perform more than others. Furthermore, we consider only group-oriented roles, since we showed in the previous section that design roles were similarly distributed among all the participants. Subsequently, all participants are (implicitly) designers. Nevertheless, participants who have an attraction for the group-oriented role are called "*group orienter*" referring to group coordination, socio-relational activities and energizer, encourager, harmonizer and orienter roles in the terms of Benne and Sheats (1948). Once again, participants could be group orienter at the boundary or locally.

(3) Finally, we consider participation within implementation spaces. We use the term "coder" to refer to the few participants that exhibited a high participation in implementation; as for the design, we do not specifically provide a label for discussion participation because all participants were simultaneously discussants.

Applying these rules enables us to define the profiles synthesized in Figure 13. Among the 20 participants in the analyzed discussions, nine exhibited no specific profile (mean or low participation) (9/20, 45%). Among the remaining participants, we have identified the 5 following profiles of participation: boundary hub group orienter and coder, boundary group orienter, boundary bridge, local bridge, and local group orienter.

[FIGURE 12 ABOUT HERE]

Figure 13 Observed profiles among the 20 main participants

[20] As defined in the theoretical section an edge is said to be a bridge if deleting it causes its endpoint to lie in different components of a network.

The champion : a unique profile covering the entire spectrum of activities and spaces

The champion FB exhibited a unique profile of *boundary hub-group orienter and coder*. This profile is characterized by:

- a central interactant position in the network (as illustrated by the central hub in [FIGURE 10 ABOUT HERE]
- Figure 11) ;
- a high contribution in implementation spaces;
- a high coloration towards group-oriented activities;
- activities at the boundaries between implementation and discussions spaces, and between design and use sub-spaces.

His profile can also be interpreted as a combination of an *initiator-contributor* role coupled with an *encourager-harmonizer* role, as referred to by Benne and Sheats (1948). He may be seen as a supervisor of the design process, for this design situation

His profile of participation may have helped him to become a recognized participant in the project. Indeed, at the time of our study, the champion was not yet a "well-recognized" participant. The PEP 327 process was the first place for him to prove his competencies to other participants. We can assume that being present in the design-oriented and group-oriented discussion activities as well as in the implementation space helped him to evolve in the community, being in line with meritocratic and peer recognition rules in OSS communities in which one's power is evaluated on the basis of technical and discursive activities.

The project leader : a boundary group orienter and technical supervisor

The project leader GvR has a unique profile strongly oriented toward group coordination and moderately oriented toward code production. He is one of the two participants who has a repulsion toward design.

A quarter of participants : boundary and local bridges

There are five bridges, corresponding to 25% of all participants.

Among them, there are *three boundary bridges* (AM, TP, AA) and two local bridges. Boundary bridges evolve within different boundaries. AM and AA are at the design-use discussion space boundary whereas TP is also present in the implementation space (even if he has a "low" contribution). These profiles are characterized by the following features:

- they are summits of "y-forks" in the network of interaction (Figure 7);
- they are "mean" designers and do not tend to perform group-oriented activities more than other participants.

This profile is held by 3 participants all recognized super-experts of the project. Their profile appears to be the one of a boundary spanner of the project, helping in transferring knowledge between design and use (Barcellini et al., 2009) but also

in structuring interactions by a specific position in discussions and presence in boundary spaces.
The two local bridges (IB and JR) are characterized as:

- evolving in the use discussion space;
- being edge of the network;
- "mean" designers that do not tend to perform group-oriented activities more than other participants;

Local bridges are characterized by their mid-central positions reflecting the fact that these participants may impact in the design process, but in a "less important" way than bridges.

Local group orienters

The profile of four *local group orienters* (El, MC, MW and RA) is characterized are by :

- evolving in one space only: MC is in the design space, the three others being in the use space;
- being peripheral participants;
- a tendency to perform group-oriented activities more than others.

Three of them have strong interaction links with the champion (as displayed in Figure 8). This suggests that these local orienters support the champion in his coordination and socio-relational activity.

7. Discussions and further works

This research proposes an approach to convey clear and accurate pictures of situated collaborative activities in an online design community, i.e. how participants and their actual activities are distributed according to time and space with regard to solving a specific design problem within a particular OSS community. We propose a socio-cognitive approach based on the analysis of actual roles and profiles to understand how participation is organized in the (discussion and implementation) spaces of actions in which participants may act in online design communities.
Following this approach, this research has two main contributions: *(1)* the first one concerns the way participation is understood in OSS research, as a model of roles distribution in collaborative design process and *(2)* the second one is more general as it concerns implications of the approach concerning the way participation may be apprehended in other design contexts thanks to this approach.

7.1.A new approach of participation in OSSD

Insights

Concerning the improvement in understanding OSS participation, we propose an alternative view to the "core-periphery" model of participation (e.g. Mockus et al., 2002; Gacek and Arief, 2004; Crowston and Howison, 2006; Jensen and Scacchi, 2007; Sowe et al., 2006). Indeed, we stress that this model offers a static view of participation proposing more or less exclusive categories of participation (e.g. active users vs. developers) that are inherent mostly to actions in the implementation side (coding, bug reporting) or relative to coordination in OSSD. To enrich this view, we propose to understand participation:

- as *distributed* in both spaces of actions, not only implementation but also discussion;
- as *situated*, i.e. related to specific design problems and not on large-scale corpus of undifferentiated online exchanges (and thus ill-delimited) regarding the tasks performed by participants;
- as *grounded in collaborative design activities* (Olson et al. 1992; d'Astous et al, 2001; 2004; Détienne et al., 2004; Détienne, 2006) performed by participants whatever their statuses in the community, and not to participation in one or another design' task (commit, bug report…).

Applying this approach to our case study– an enhancement proposal process concerning the addition of a specific function in the Python community- has enabled us to understand the mapping of participation in and between spaces of actions (implementation, discussion and boundary spaces), the distribution of roles among these spaces, and specific profiles of main participants in these spaces.

Our result confirms that participation in OSSD is hybrid as participants may act in both implementation and discussions spaces (e.g. Ducheneaut, 2005; Sack et al., 2006), but that pure coding participation may not exist. Indeed, all coders act in the discussion space and in this space collaborative design activities – at least regarding generation-evaluation of design solutions- are distributed among a broad number of participants whatever their statuses in reference to the "core-periphery" model. In other words, what happens on the implementation side (coding, bug reporting, documenting) is only the "tip of the iceberg" of OSS participation whereas "core-periphery" model does not emphasis explicitly forms of participation related to discussion space. For example, some participant that would have been qualified of "active users" in the "core-periphery" model are actually *co-designers* of the OSS as they are engaged in collaborative design activities in discussion.

We have also revealed the presence of some key-participants in and between spaces (at the boundaries) defining *five individual profiles of participation* (boundary hub-group orienter and coder, boundary group orienter, boundary bridge, local bridge, and local group orienter) on the basis of three criteria:

- the place of participants: local space (design, use) or boundary space (design-use, design-implementation…);
- the position of the participants in the network of interactions (central hub, summit of a fork);
- the coloration of participation taking into account only the tendency to perform group oriented activities (coordination, support), since design-oriented activities are well distributed among participants.

Moreover, these results highlight the dynamics of interactions between participants and clarify the profiles of key participants, such as the super-experts recognized in the project or the champion. Yet, a "core-periphery" model grounded in exclusive ring cannot embrace those actions at the boundary – for instance between design discussions and use discussions. We assume that the existence of participants with such profiles are key elements for the performance of the design process, such as previously revealed in software design studies about boundary spanners (Sonnewald, 1996) and super-experts (Sonnentag, 2001).

Moreover, our results extend general description concerning super-expertise in software design (Sonnentag, 1998; 2001) as well as boundary spanning (Sonnenwald, 1996), clarifying the nature, place and forms of participation of super-experts and boundary spanners, their roles and profiles in a specific design process. On the one hand, results have shown their preeminence as well as the preeminence of other key-participants (e.g. super-experts) across activity spaces to sustain knowledge exchanges, whereas a small subset of these "known" participants also provided group-oriented support. On the other hand, there is (although small) room to "unknown" participants to develop an activity in these boundary spaces, such as being the "champion" of a specific problem and design process. Moreover, it may stress the fact that project leader's implication also involves in guaranteeing the "philosophy" of the Python language by his interventions, as an *energizer* and a *coordinator* (Benne & Sheats, 1948).

7.2.Implications

This research provides a picture of a single design process. Our objective was to understand a piece of the reality of what is occurring in some OSS design process to enrich the current model of participation by a complementary perspective. We assume that there are three potential implications of the approach we develop. (1) The first one concerns the continuation of the work in the context of OSS seen as a model of successful and efficient design organization, (2) the second one concerns the replication of same kind of studies in other OSS community, online

communities or design process organization in order to produce more general results and formalize collaborative design processes in order to improve and transform it, and finally (3) the third one concerns the potential automation of the approach to support the replication of this work.

Understanding more deeply a successful design organisation

OSSD process can be seen as successful organization of design (in terms of quality of the artifact design), but also as an organization attractive for participants, in the sense that it supports them in developing their skills and learning for instance (ref). In this sense, we assume that the modeling of participation proposed by our approach can be used to better understand what may explain the quality of the software designed and how the organization of design process support development and learning of participants. This implies to replicate this work in a more longitudinal perspective, in order to address the following questions:

(a) how profiles of the same participants vary depending on contexts. This may help to clarify the way key-participants construct their position in the project (e.g. super-experts), and the way they develop their skills by participating to OSSD ;

(b) what are the nature of resources mobilized by participants and what are the dimensions of the developmental processes in which participants are engaged. For instance, one issue concerns how the evolution of participants' profiles in the processes correlates with the development of their (actual vs. perceived) competencies (e.g. Ducheneaut, 2005);

(c) what the effects (if any) of the presence or the absence of specific profiles (boundary hub, coder, orienter) are on the performance of the design process and the quality of the artifact (designed and maintained). Indeed a previous study on Python (Barcellini et al., 2009) comparing two episodes of design-proposal related to the same-topic led us to suggest that success may depend on the profile adopted by the champion supporting the proposal.

Generality of the approach and its potential applications to other design contexts

Beside implications oriented to the understanding of OSS design, we assume that focusing on roles and profiles performed by participants in spaces of actions is an innovative approach to address forms of participation in other design situations. Indeed, understanding participatory design settings to better understand and organize it is one of the key issues of HCI practitioners.

Our approach makes more explicit the aggregation of activities performed by participants all along a specific design process. However, the potential extension

of this approach to other contexts of design has to be discussed, in particular regarding:

1. the definition of spaces;
2. the definition of a situated design problem;
3. and the way profiles are determined (e.g. Détienne, et al. 2012).

We are currently extending this work in three different contexts: one is related to the design of a Wikipedia article (Fréard et al., 2010; Détienne, et al., 2012); another is related to online co-elaboration of work practices of French ergonomist (Barcellini, Delgoulet & Nelson, 2012), and the last one is related to the design of a software in agronomics context (Barcellini, Prost, & Cerf, in revision). We will mainly focus on Wikipedia in the following to illustrate how the approach can be extended to other situations.

(1) A first issue is to determine the *spaces of actions* in which participants may act. In the case of Wikipedia, a discussion space is encompassed by discussion pages associated with Wikipedia articles and the equivalent of the implementation space – called *production space* in this case- is the page containing all the editions made to the article. More generally, the production space should be the space in which design solutions are crystallized after discussions.

(2) A second issue is to determine *a situated design problem*, defining the goal pursued by participants. In the case of Wikipedia, the design of a single article within a project (e.g. Astronomy) is assumed to define a design problem (as a PEP in the case of Python).

(3) Finally, a third issue is to determine the dimensions of roles underlying profiles, a metric to identify profiles. In the other contexts investigated, we have operationalized the four dimensions of roles linked to discussion (interaction, task-oriented and group-oriented) and production (coding, documenting, editing…). However, the approach had to be adapted depending on: *modalities of interactions* (face-to-face or mediated), *goals of participants* and *research question*. *Modalities of interactions* impact the way interactive roles are defined (quotation between participants in the case of OSS, turns in discussion in the case of face to face interactions). *Goals of participants and research question* impact the way task oriented roles may be defined and analyzed. Indeed, in this research we are interested in collaborative design activities and we could refer to a stabilized literature describing them (e.g. Olson et al., 1992, d'Astous et al, 2001; Détienne, 2006). In the case of Wikipedia, we have elaborated a slightly different coding scheme of roles combining discursive activities with type of knowledge exchanged by participants (e.g. Fréard et al. 2010; Détienne et al., 2012).

Concerning the metrics, we have pointed that the notion of profile has been already operationalized in face-to-face design situations –collaborative architectural design- (Baker et al., 2003; 2009). However, in this former study profiles of participants (3 actually) were determined on the basis of frequencies of contribution. In this research and the following (Fréard et al., 2010; Barcellini et

al., in revision), we have refined our understanding of the notion of profile as referring to the specific contributions of participants compared to others and have consequently proposed to use Relative Deviation as an indicator of specific contributions of participants.

Nevertheless, one can wonder to what extent roles, profiles and their distribution identified in this case study may be generalizable. We assume that roles (interactive, task-oriented, group-oriented) are quite generic to any design context. The profiles, as combination of roles are more specific and some of them could possibly be found in other contexts, e.g. the boundary hub group orienter. The distribution of roles and profiles is more related to the design context. However we could advance that some particular balance of roles, e.g. balance between task-oriented and group-orienter roles, may be similar in various contexts and may be an indicator of the quality of collaboration within the team (Détienne et al., 2012).

To conclude, our approach may be applied also to other design contexts than *computer-mediated* ones. For example, participatory design or user-centered design situations could be analyzed to reveal how distribution and types of participation are actually associated with taking users and usage into account in design. In this direction, we could question to what extent the notion of boundaries remains relevant and how these boundaries are "materialized" (how spaces and boundary spaces could be defined). Moreover, distribution of profiles and roles of various stakeholders might be analyzed with respect to the quality (i.e. usefulness as well as usability) of the final design product.

Automation of the approach

The third implication of this work regards the development of new tools based on specifications derived from our approach. The target users are twofold : researchers who would like to conduct in-depth investigation into online activities and participants in communities. The methodological approach might be partly used to characterize participation both in other online epistemic community contexts and in other design situations (either face-to-face or not). Analyses of roles are time-consuming and thus a part of the analyses might be automated in order to progress from fine-grain human analyses to semi-automated analyses at a more large level. In this direction, we combine the approach used to analyze co-elaboration of knowledge in Wikipedia with Natural language Processing techniques (Fréard et al, 2010). For participants, computer-assisted management tools dedicated to online communities could be specified on the basis of our findings. For example, they could provides monitoring function, i.e. reveal "what is going on" at the boundaries (who is participating), awareness functions, i.e. what topics are discussed, in which spaces they are discussed and even if user/usages topics have been actually conveyed in the design space. Finally,

because such tools provide potentially supports to reflexive activity within design community, an interesting issue would be to assess their impact to this activity.

8. Notes

Background. *This article is based on the Ph.D. of the first author.*

Acknowledgements. The authors wish to thank all the members of the Python community for their kind welcome.

Support. The authors wish to thanks French Ministry for Research and INRIA for funding this research.

Authors' addresses. **Flore Barcellini**, Le Cnam, CRTD, Ergonomics team , 41 rue Gay-Lussac 75005 Paris France. flore.barcellini@cnam.fr . **Françoise Détienne**, LTCI- UMR 5141 - CNRS - Telecom Paris Tech, 46 rue Barrault 75634 Paris Cedex 13 France, francoise.detienne@telecom-paristech.fr . **Jean-Marie Burkhardt**, IFSTTAR LPC, 25 allée des Marronnier 78 000 Versailles-Satory France, jean-marie.burkhardt@ifsttar.fr

9. References

Antikainen, M., Aaltonen, T & Väisänen, J. (2007). The role of trust in OSS communities : the case of Linux Kernel community. In *proceedings of IFIP conference*, 223-228. Boston, USA : Springer.

d'Astous, P., Détienne, F., Visser, W., & Robillard, P. N. (2004). Changing our view on design evaluation meetings methodology: a study of software technical evaluation meetings. *Design Studies, 25*, 625-655.

d'Astous, P., Détienne, F., Robillard, P. N., & Visser, W. (2001). Quantitative measurements of the influence of participants' roles during peer review meetings. *Empirical Software Engineering, 6*, 143-159.

Bach, P. & Twidale, M. (2010). Social Participation in Open Source : What means for designers. *Interactions, 17 (3),* 70-74

Baker, M., Détienne, F., Lund, K., & Séjourné, A. (2009) Etude des profils interactifs dans la conception collective en architecture. In F. Détienne & V. Traverso (Eds) *Méthodologies d'analyse de situations coopératives de conception: le corpus MOSAIC*, Nancy : PUN.

Baker, M. (2004). *Recherches sur l'élaboration de connaissances dans le dialogue*. Unpublished dissertation Thèse d'Habilitation à Diriger des Recherches. Université de Nancy 2.

Baker, F., Détienne, F., Lund, K., & Séjourné, A. (2003) Articulation entre élaboration de solutions et argumentation polyphonique. In J.C. Bastien (Ed.) proceedings of EPIQUE'03, 235-240. Rocquencourt, France : INRIA.

Bales, R.F. (1950). Interaction process analysis : a method for the study of small groups. Cambridge : Addison-Wesley.

Barcellini, F. Prost, L. & Cerf, M. (in revision). Understanding emergence of an software online community :an analysis of roles. *Design Studies*

Barcellini, F., Delgoulet, C. & Nelson, J. (2012). Pratiques d'échanges dans une communauté d'ergonomes francophones : vers la définition d'un cadre d'analyse. In P. Béguin, V. Pueyo & M-F. Dessaignes (Eds.), *Actes du 47ème congrès de la SELF*, 6-12. Lyon, France : Editions du Gerra.

Barcellini, F., Détienne, F. & Burkhardt, J.M. (2010a). Role-based analysis of collective activity in online communities. Workshop on *studying online behavior*, *CHI'10*, avril 10-15 2010, Atlanta, USA

Barcellini, F., Détienne, F. & Burkhardt, J.-M. (2010b). Distributed design and distributed social awareness: exploring inter-subjective dimensions of roles. In M. Lewkowicz, P. Hassanaly, M. Rodhe & V. Wulf (Eds) *Proceedings of the COOP'10 conference*, 3-24. London, UK : Springer Verlag.

Barcellini, F., Détienne, F., & Burkhardt, J.M. (2009). Participation in online interactions spaces: design-use mediation in an Open Source Software community. *International Journal of Industrial Ergonomics, 39,* 533-540.

Barcellini, F., Détienne, F., & Burkhardt, J.M. (2008a). Users and developers mediation in an Open Source Software Community: boundary spanning through cross participation in online discussions. *International Journal of Human Computer Studies, 66 (7)*, 558-570.

Barcellini, F., Détienne, F., Burkhardt, J.M., & Sack, W. (2008b). A socio-cognitive analysis of online design discussions in an Open Source Software community. *Interacting with Computers, 20(1)*, 141-165.

Barcellini, F., Détienne, F., Burkhardt, J.M., & Sack, W. (2005). Thematic coherence and quotation practices in OSS design-oriented online discussions. In K. Schmidt, M. Pendergast, M. Ackerman, & G. Mark (Eds) proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work, 177-186. New-York, USA : ACM Press.

Benne, K.D., & Sheats, P. (1948). Functional roles of group members. *The journal of social issues, 4(2),* 41-49.

Bernard, J.-M. (2003). Analysis of local or asymmetric dependencies in contingency tables using the imprecise Dirichlet model. In proceedings of the 3d International symposium on imprecise probabilities and their applications *ISIPTA'03*. Carleton Scientific.

Bryant, S. L., Forte, A., & Bruckman, A. (2005) Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In proceedings of the *2005 ACM SIGGROUP conference on Supporting group work,* 1-10. New York City, USA : ACM Press

Bucciarelli, L.L. (1998). Engineering design process. In: Dubinskas, F. (Ed.), *Making time: culture, time and organization in high technology* (pp. 92-122). Philadelphia, PA, USA : Temple University Press.

Capilutti, A., Lago, P. et Morisio, M. (2003) Characteristics of Open Source Projects. In proceedings of the Seventh *European Conference on Software Maintenance and Reengineering CSMR'03,* 1-11. USA : IEEE Computer Society.

Cohendet, P., Créplet, F., & Dupouët, O. (2003). Innovation organisationnelle, communautés de pratique et communautés épistémiques : le cas de Linux. *Revue Française de Gestion, 29 (146)*, 99-121.

Crowston, K., & Howison, J. (2005). The social structure of free and open source software developement. – *First Monday, 10(2),* http://www.firstmonday.org/issues/issue10_2/crowston.

Crowston, K. & Howison, J. (2006). Hierarchy and centralization in free and open source software team communications. *Knowledge, Technology & Policy 18(4)*, 65–85.

Crowston, K., Qing, L., Kangning, W., Eseryel, U.Y., & Howison, J. (2007). Self-organization of team for free/libre open source software development. *Information and Sofwtare Technology, 49*, 564-575.

Crowston, K. (2011). Lessons from volunteering and free/libre open source software development for the future of work. *Researching in the future of information systems. In proceedings of IFIP Advances in Information and Communication Technology*, 215-229. London, UK : Springer

Crowston, K., Wei, K., Howison, J. & Wiggins A. (2012). Free/Libre Open Source Software Development : what we know and what we do not know. *ACM Computing Survey, 44(2)*, 7 :5-7 :35.

Curtis, B., Krasner, H. & Iscoe, N. (1988). A field study of the software design process for large systems. *Communications of the ACM, 31,* 1268—1287.

Curtis, B. (1986). By the way, did anyone study any real programmers ? In E. Soloway and S. Iyengar (Eds) *Empirical Studies of programmers*, pp 256-261. New Jersey, USA : Ablex publishing.

Darses, F., Détienne, F., Falzon, P., & Visser, W. (2001). *A method for analysing collective design processes*. Reseach Report n° 4258, INRIA, September 2001.

Détienne, F., Baker, M. & Burkhardt, J-M. (2012) Perspectives on quality of collaboration in design, Introduction. *CoDesign*: *International Journal of CoCreation in Design and the Arts*, *8(4),* 197-199.

Détienne, F., Barcellini, F. Baker, M., Burkhardt, J.M., & Fréard, D. (2012). Online epistemic communities: theoretical and methodological directions for understanding knowledge co-elaboration in new digital spaces. *Work, 41(Supplement 1),* 3511-3518

Détienne, F. (2006). Collaborative design: Managing task interdependencies and multiple perspectives. *Interacting with Computers, 18*,1-20.

Détienne, F., Boujut, J-F., & Hohmann, B. (2004) Characterization of Collaborative Design and Interaction Management Activities in a Distant Engineering Design Situation. In F. Darses, R. Dieng, C. Simone, M. Zaklad (Eds) *proceedings of the 2004 COOP conference*, 83-98. Amsterdam, the netherlands : IOS Press.

Ducheneaut, N. (2005). Socialization in an Open Source Software Community: A Socio-Technical Analysis. *Journal of Computer Supported Collaborative Work, 14,* 323-368.

Fréard, D., Denis, M., Détienne, F., Baker, M, Quignard, M. & Barcellini, F. (2010). The role of argumentation in online epistemic communities: the anatomy of a conflict in Wikipedia. In M. Neerincx and W-P Brinkman (Eds) *proceedings of the European Conference on Cognitive Ergonomics ECCE'10*, 91-98. New York City, USA : ACM Press.

Gacek, C., & Arief, B. (2004). The Many Meanings of Open Source. *IEEE Software, 21*, 34-40.

Gasser, L., Scacchi, W., Ripoche, G., & Penne, B. (2003). Understanding Continuous Design in F/OSS project. Communication at 16th International Conference on Software Engineering & its Applications (ICSSEA-03), Paris, France, December 2003.

German, D. (2003). The GNOME project: a case study of Open Source, Global Software Development. *Software process improvement and practice, 8*, 201-215.

Ghosh, R.A, Glott, R., Krieger, B., & Robles, G. (2002). *Free/Libre and Open Source Software: Survey and Study*. FLOSS Deliverable D18: FINAL REPORT. Part 4:

Survey of Developers. International Institute of Infonomics, University of Maastricht, The Netherlands, June 2002.

Gleave, E., Wesler, H.T., Lento, T.M., & Smith, M. (2009). A Conceptual and Operational Definition of 'Social Role' in Online Community. In proceedings of the *42nd IEEE Hawaii International Conference on System Sciences*, 1-11. IEEE Xplore

Gonzalez-Barahona, J.M., L. Lopez & G. Robles (2004): Community Structure of Modules in the Apache Project. In Proceedings of the *4h International Workshop on Open Source Software Engineering,* 44-48. IET digital library.

Grinter, R., (1999). Systems Architecture: Product Designing and Social Engineering. In proceedings of *the international joint conference on Work activities coordination and collaboration (WACC '99),* 11-18. New York City: ACM Press.

Gutwin, C., Penner, R., & Schneider, K. (2004) Group Awareness in Distributed Software Development. In *Proceedings of CSCW 2004*, 72-81. New York, USA : ACM press.

Hedberg, H. & Livari, N. (2009). Integrating HCI Specialist Into Open Source Software Development Projects. In C. Boldyreff et al. (Eds.) proceedings of *OSS 2009, IFIP AICT 299*, 251–263. London, UK: Springer.

Hendry, D.G. (2008). Public participation in proprietary software development through user roles and discourse. *International Journal of Human-Computer Studies, 66*, 545–557.

Herbsleb, J. (2007). Global Software Engineering: The Future of Socio-technical Coordination. In proceedings of *FOSE '07 2007 Future of Software Engineering*, 188-198. New York, USA : ACM Press.

Iivari, N. (2011). Participatory design in OSS development: interpretive case studies in company and community OSS development contexts. *Behaviour & Information Technology, 30(3)*, 309-323.

Jensen, C. & Scacchi, W. (2007). Roles Migration and Advancement Processes in OSSD projects : a comparative cas study. In proceedings of *International Conference on Software Engineering (ICSE'07)*, 364-374. IEEE digital library.

Jergensen, C., Sarma, A., & Wagstrom, P. (2011). The onion patch : Migration in Open Source Ecosystem. In proceedings of the *ESEC/FSE'11 conference*, 70-80. New York City, USA : ACM Press.

Krasner, H., Curtis, B., & Iscoe, N. (1987). Communication breakdowns and boundary spanning activities on large programming projects. In G. Olson, S. Sheppard, and E. Soloway, E. (Eds.) *Empirical Studies of programmers: Second Workshop*, 47-64. New Jersey, USA : Ablex publishing.

Madey, G., V. Freeh & R. Tynan (2002): The Open Source Sofware Development Phenomenon: An Analysis based on Social Network Theory. In R. D. Banker, H. Chang, & Y-C. Kao (Eds.) proceedings of the *Americas Conference on Information Systems* (AMCIS2002), 1806-1813. Dallas, USA: University of Dallas- AIS elibrary.

Mahendran, D. (2003). *Serpents and Primitives: An ethnographic excursion into an Open Source community*. Unpublished Master's Thesis, School of Information Management and Systems, University of California at Berkeley.

Mockus, A., Fielding, R. T., & Herbsleb, J. (2002). Two Case Studies of Open Source Software Development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology, 11(3)*, 309–346.

Nakakoji, K., Yamamoto, Y., Nishinaka, Y., Kishida, K., & Ye, Y. (2002). Evolution patterns of open-source software systems and communities. In *IWPSE '02: Proceedings of the International Workshop on Principles of Software Evolution*, 76-85, New York, NY, USA : ACM Press.

Oberlé, D. (1995). L'approche interactionniste des rôles. In G. Mugny, D. Oberlé, J-L. Beauvois (Coords) *Relations humaines, groupes et influence sociale (*pp101-110). Grenoble, France : PUG, collection « La psychologie sociale ».

Olson, G.M., Olson, J.S., Carter, M. R., & Storrosten, M. (1992). Small Group Design Meetings: An Analysis of Collaboration. *Human-Computer Interaction, 7,* 347-374.

Olson, G. M., & Olson, J. S. (2000). Distance Matters. *Human-Computer Interaction, 15*, 139-178.

Preece, J. (2000) *Online communities : designing usability and supporting sociability*. New York, USA : John Wiley and sons.

Raymond, E. S. (1999). The cathedral and the bazaar [page web] http://www.tuxedo.org/esr/writings/cathedral-bazaar/ [20 june 2005].

Rouanet, H., Le Roux, B., & Bert, M.-C. (1987). *Statistiques en sciences humaines*. Paris, France : Dunod.

Ripoche, G. & Sansonnet, J.-P. (2006). Experiences in Automating the Analysis of Linguistic Interactions for the Study of Distributed Collectives. *Journal of Computer Supported Cooperative Work, 15(2-3),* 149-183.

Sack, W, Détienne F, Burkhardt, J.M., Barcellini F., Ducheneaut, N., & Mahendran D. (2006). A Methodological Framework for Socio-Cognitive Analyses of Collaborative Design of Open Source Software. *International Journal of Computer Supported Collaborative Work, 15(2-3)*, 229-250

Sarant, S.A. (2004). *The role of organizational boundary spanners in industry/university collaborative relationship*. Unpublished Dissertation Thesis. North Carolina State University, 2004.

Scacchi, W. (2001). Understanding the requirements for developing Open Source Software Systems. *IEEE Proceedings--Software*, *149(1)*, 24-39.

Sonnentag, S. (2001) High Performance and Meeting Participation: An Observational Study in Software Design Teams. *Group Dynamics, 5(1),* 3-18

Sonnentag, S. (1998) Expertise in professional software design : A process study. *Journal of applied psychology, 83(5)*, 703-715.

Sonnenwald, D.H. (1996). Communication role that support collaboration during the design process. *Design Studies, 17*, 277-301.

de Souza, C., Froelich, J. & Dourish,P. (2005). Seeking the source: software source code as a social and technical artifact. In K. Schmidt, M. Pendergast, M. Ackerman, G.Mark (Eds) proceedings of the *GROUP'05* conference, 197-206. New York, USA: ACM Press.

Sowe, S. , Stamelos, I. , and Angelis, L. (2006). Identifying knowledge brokers that yield software engineering knowledge in OSS projects. *Information and Software Technology , 48*, 1025–1033.

Sowe, S. , Stamelos, I. , & Angelis, L. (2008) Understanding knowledge sharing activities in free/open source software projects: An empirical study. *Journal of Systems and Software, 81(3)*, 431–446.

Stempfle, J., & Badke-Schaub, P. (2002). Thinking in design teams - an analysis of team communication. *Design Studies, 23*, 473–496.

de Visscher, P. (1995). Différenciation des rôles et animation de groupes. In G. Mugny, D. Oberlé, J-L. Beauvois (Eds.) *Relations humaines, groupes et influence sociale (*pp 131-138). Grenoble, France : PUG, collection « La psychologie sociale ».

Visser, W. (2006). *The cognitive artifacts of designing*. Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Wellman, B. & Berkowitz, S.D. (1988). *Social Structures: A Network Approach*. Cambridge, USA: Cambridge University Press.

Welser, H.T., Cosley, D., Kossinets, G., Lin, A. Dokshin, F., Gay, G., & Smith, M. (2011). Finding social roles in wikipedia. *Proceedings of the IConference'11*, 122-129.

Zhou, M., & Mockus, A. (2010). Developer fluency: achieving true mastery in software projects. In Proceeding of *FSE '10 Proceedings of the eighteenth ACM SIGSOFT*

international symposium on Foundations of software engineering, 137-146. New York, USA : ACM Press

10. Appendix

[FIGURE 13 ABOUT HERE]

Figure 15 Distribution of interactions (a) and activities (b) of participants PEP 327 corpus

Comparing Figure 10 and 13 reveals that in the design space, only one added participant "JF" might have a relevant position, being a local center of interactions. In the use space, 6 added participants (AC, DB, BP, DC, PM and JPC) might have a relevant position. All other participants have no noticeable positions (there are peripheral).

Figures

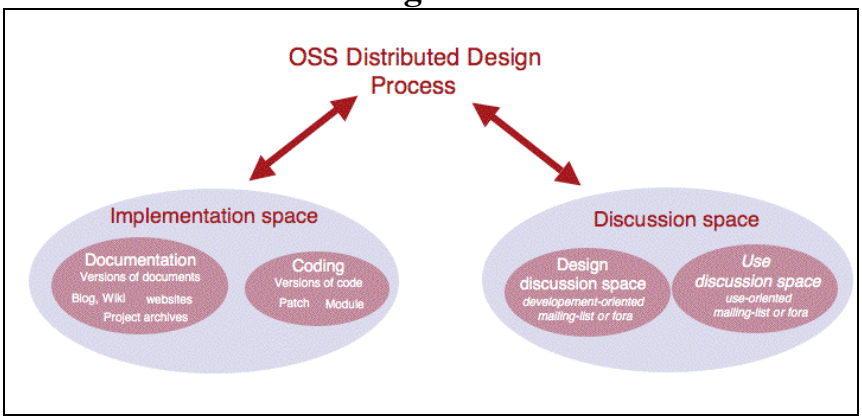

Figure 1 Distribution of OSS design process in discussion and implementation spaces

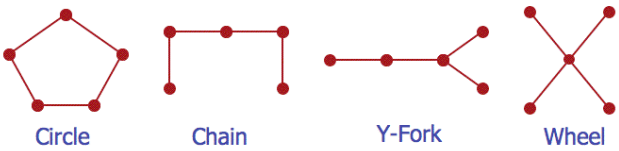

Figure 2 Structures of communication network (from Oberlé, 1995)

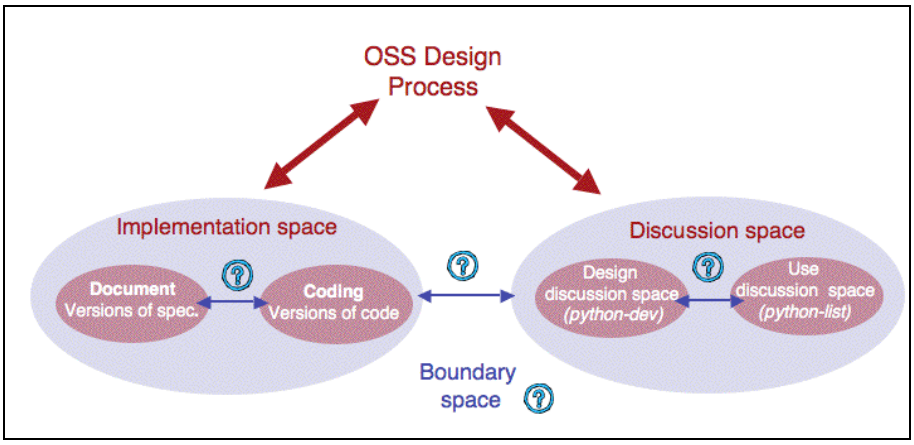

Figure 3 Distribution of the OSSD process in the 2 spaces of actions and their boundaries

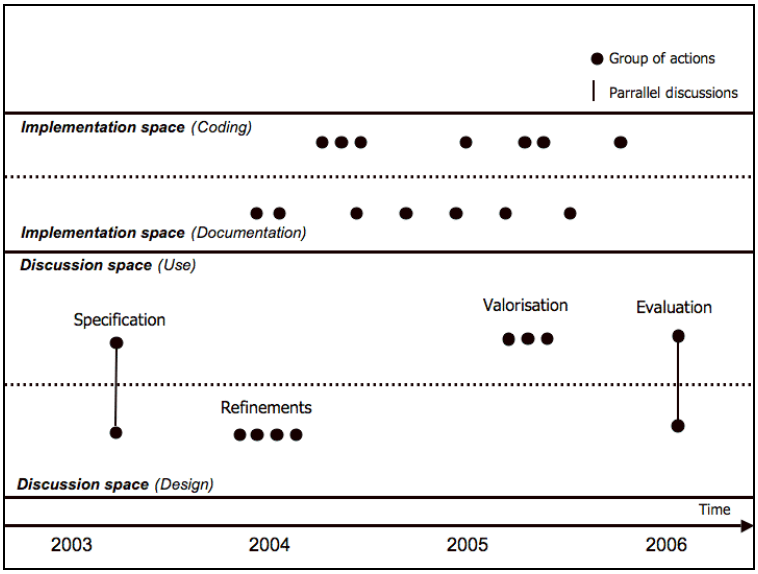

Figure 4 Simplified overview of PEP 327 process among the spaces of actions (adapted from Barcellini et al., 2009)

Spaces	Actions	Participants	Size
Discussion	5 discussions /219 messages	58	12000 lines of text produced by participants
Implementation coding	44 commits/revisions of module decimal.py	7	176 lines of comments, 14393 bytes modified in module decimal.py
Implementation document	9 revisions of PEP 327 documents	2	11 lines of comments, 18373 bytes modified in PEP document

Figure 5 PEP 327 corpus analyzed in the discussion and implementation spaces

Spaces of actions	Roles	Descriptors	Examples
Discussion (design discussion and/or use discussion)	*Interactive role*	Positions in the network of communication (centrality-hub, periphery…) Number of participants interacting (citing and being cited) with another one Number of mutual citation between participants	Not relevant
	Design-related role	Discursive function related to design (proposition, evaluation, clarification) Epistemic function related to providing knowledge (application domain, computer science, scenario of use …)	Proposition *I'd like to add (to previous proposals) (name_function) How about doing it both ways… ?* Evaluation *-1'1. It looks like a noun to me Now, that's a great idea* Clarification *As I said earlier in the msg, (…) I had in mind above Is this about the example below?*
	Group-oriented role	Discursive function related to group management (coordination, decision, synthesis) Discursive function related to interpersonal relationship: acknowledgment of others' work, support	*I suggest that you register with the PEP editor right away to get a PEP number. Okay, here's what we have so far: Enum (y). The code of XX class is based in large part on the code of Y: thank you for your (very) valuable ideas*
Implementation (coding and documentation)	*Implementation role*	Initial writing of the PEP document or referenced contributions in the PEP document Number of direct commits in documentation or coding spaces Number of commits in which the committer explicitly referenced the work of another participants	Not relevant

Figure 6 Coding scheme of roles

Type of role		Implementation roles	Interaction roles		
			Interactive	*Design-oriented*	*Group-oriented*
Metrics used		*Intensity*	*Intensity*	*Coloration*	*Coloration*
Intensity of participation	*High*	*Q3 ≤ n° of revisions (Q3=4)*	*Q3 ≤ n° participants interacting with (Q3=3)*	*RD between epistemic and cognitive activities and participants ≥ 0,2*	*RD between coordination and social activities and participants ≥ 0,2*
	Mean	*Q1<n° of revisions< Q3*	*Q1< n° participants interacting with < Q3*	*-0,2 < RD between epistemic and cognitive activities and participants < 0,2*	*-0,2 < RD between coordination and social activities and participants < 0,2*
	Low	*n° of revisions ≤ Q1 (Q1=1)*	*n° participants interacting with ≤ Q1 (Q1=1)*	*-0,2 ≤ RD between epistemic and cognitive activities and participants*	*-0,2 ≤ RD between coordination and social activities and participants*

Figure 7 Synthesis of rules used to construct profiles of participation

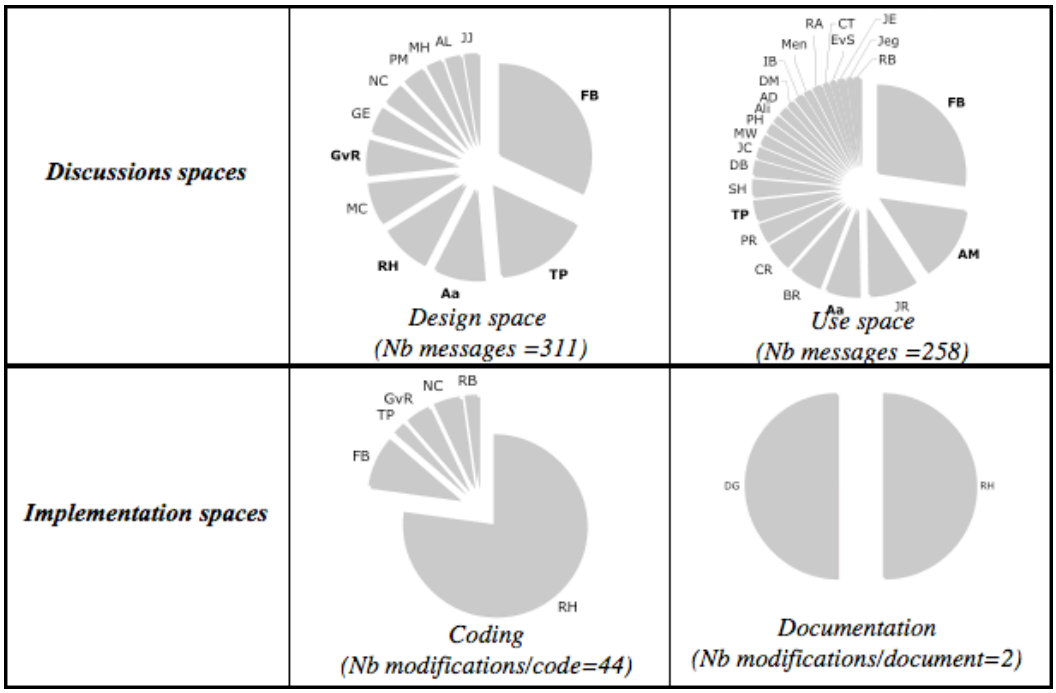

Figure 8 Distribution of participation in the discussion and the implementation spaces for the PEP 327 corpus (only the 4th quartile of participants in mailing-lists is represented, excepted for documentation)

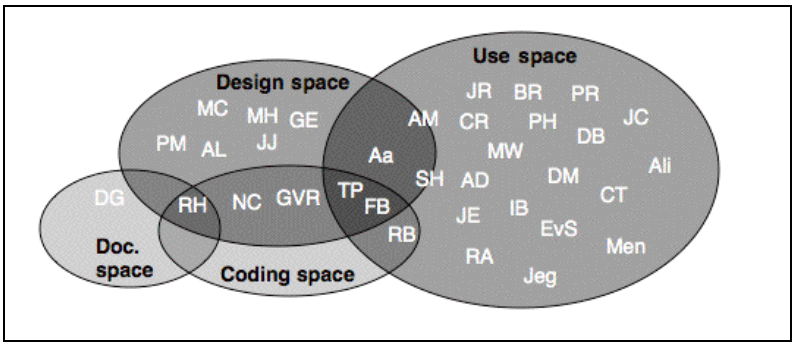

Figure 9 Map of participation in the discussion and implementation spaces

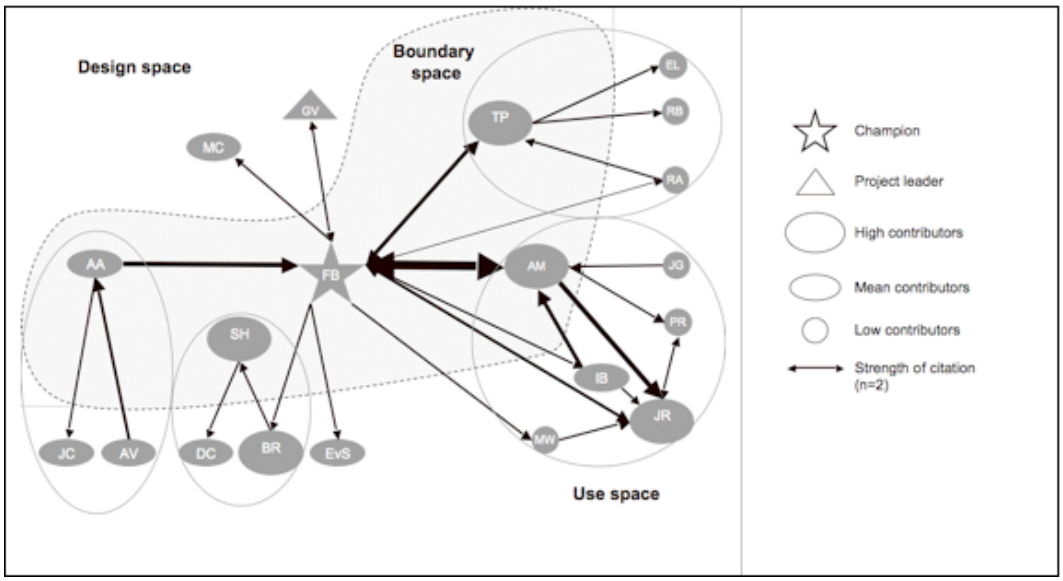

Figure 10 Distribution of interactions (a) and activities (b) of the fourth quartile of participants PEP 327 corpus

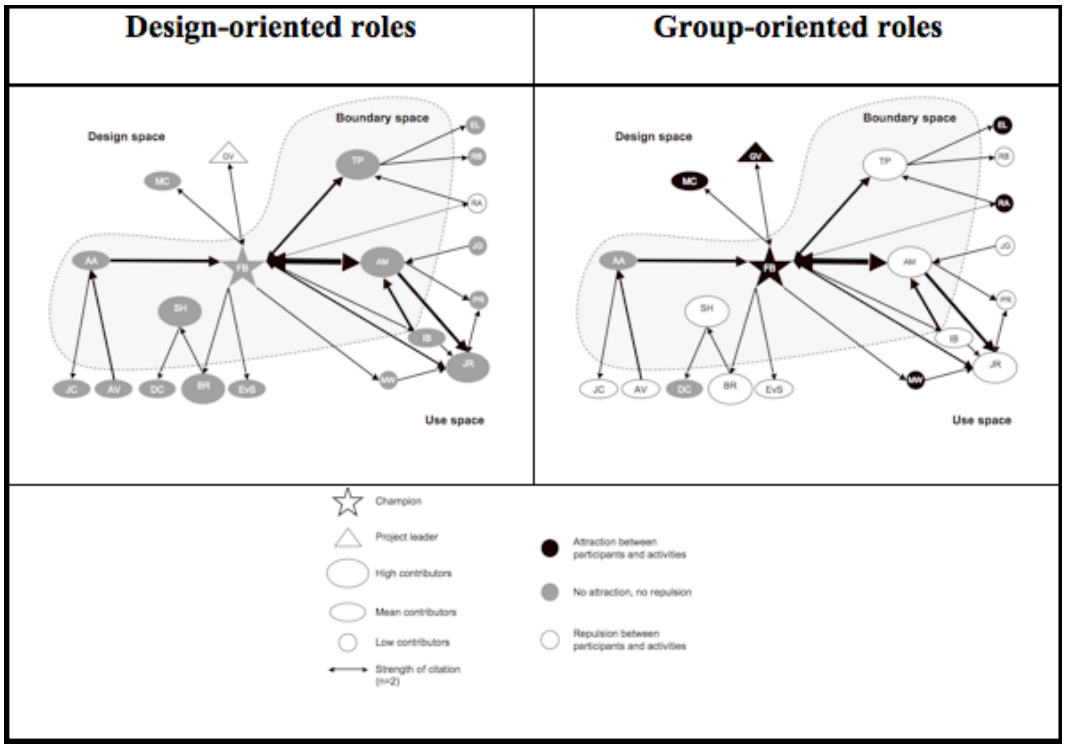

Figure 11 Distribution of roles among participants

Spaces	Roles Participants	Interact.	Design	Group	Implem.	Profile
Boundary *(Imp. design use)*	**FB**		=	+		Boundary hub, group orienter & coder
	TP		=	-		Boundary bridge
Boundary *(imp-design)*	**GvR**		-	+		Boundary group orienter & coder
Boundary *(design-use)*	**AM**		=	-		Boundary bridge
	AA		=			Boundary bridge
	SH		=	-		*No specific profile*
Design	**MC**		=	+		Local group orienter
Use	**MW**		=	+		Local group orienter
	IB, JR		=	-		Local bridge
	BR		=	-		*No specific profile*
	RA		-	+		Local group orienter
	EI		=	+		Local group orienter
	AV, EvS, PR, JC, JG, RB, DC		=	-		*No specific profile*

Figure 12 Observed profiles among the 20 main participants

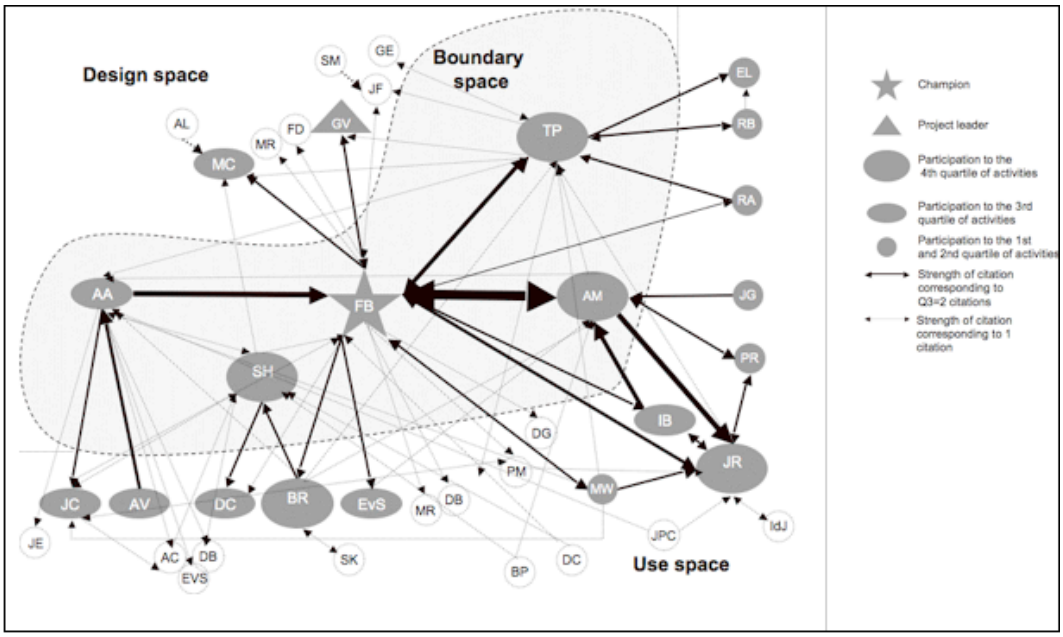

Figure 13 Distribution of interactions (a) and activities (b) of participants PEP 327 corpus

FOOTNOTES

1. **Flore Barcellini** is an ergonomist with an interest in cooperative and mediated activities and online communities; she is a associate professor in the Research Center on Work and Development (CRTD) of Le Cnam. **Françoise Détienne** is a cognitive psychologist and ergonomist with an interest in collaborative design, technology-mediated collaboration and online communities; she is a research professor at CNRS in the Department of Social, Economics and Human Sciences of LTCI (CNRS - Telecom ParisTech). **Jean-Marie Burkhardt** is a psychologist and ergonomist with an interest in collaborative design, mixed realities and learning technologies; he is a research professor in the Laboratory of Driver Psychology of IFSTTAR - Institut Francais des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux.
2. a group of people connecting via the Internet with a common goal (Preece, 2000)
3. Coding activity may be mostly individual whereas it may be collective in some specific methodology such as pair programming for instance.
4. We previously made a distinction between three spaces named discussion, documentation and implementation. However, we believe that it is ambiguous to refer to a single space embracing the proper implementation composed of documentation and coding.
5. Supported by the Concurrent Version System (CVS) or by Subversion
6. The PEP document is written to describe a new language feature: it is intended to provide a concise technical specification of the feature, the rationale for the feature, and a

reference implementation.

7. The real difference in OSS and proprietary software situations may lie in more open coordination practices, i.e. auto-attribution of tasks (Crowston et al., 2007).
8. Contrary to other studies analyzing online discussions, we have used the quotation link rather than the reply-to link to reconstruct the discussions. In a previous paper (Barcellini et al 2008b), we have shown that the organization of messages according to the quoting link is relevant to reconstruct the thematic consistency of online discussions and to understand the interactions between participants in terms of verbal turns.
9. The initiator of the discussion.
10. Subversion systems that trace all revisions of code, their content and authors.
11. For the record, broad traffic of these lists is as follows: python-list (5370 participants posting 51495 messages); python-dev 451 participants posting 8955 messages).
12. These commits and revisions are one part of the implementation space: we do not consider actions into issue tracking systems as the Python project did not use it at the beginning of the process. Moreover, we do not consider branches in the decimal.py modules.
13. This requirement is filled in our case as we have 42 cells (21 participants vs. design and group-oriented activities) and 4 "0".
14. Data represented are the same data than in figure 4.
15. He posted around 25% of messages in both lists and performed 9% (4/44) of code revisions.
16. He is the 2nd poster just after the champion (52/405, 13% of messages) and the third contributor in the coding space (1/44, 2% of modifications).
17. RH contributes to 77% (34/44) of all the code revisions, and he is the fourth contributor in the design space (7%, 27/405 messages).
18. Small size for the participants involvement corresponding to the 1st and 2nd quartile (0 to 14 activities), mean size for the median quartile (15 to 87 activities) and larger size for the 4th quartile (equal to more than 88 activities).
19. Only the relations greater or equal to the 3rd quartile of the distribution of citations are represented (75% of discussions). The smallest arrows represent 2 mutual citations between participants, to represent superior strength we use arbitrary thickness.
20. As defined in the theoretical section an edge is said to be a bridge if deleting it causes its endpoint to lie in different components of a network.