



# Imagining the Wikipedia community: what do Wikipedia authors mean when they write about their ‘community’?

new media & society

XX(X) 1–18

© The Author(s) 2010

Reprints and permission: sagepub.

co.uk/journalsPermissions.nav

DOI: 10.1177/1461444810378364

<http://nms.sagepub.com>



**Christian Pentzold**

Chemnitz University of Technology, Germany

## Abstract

This article examines the way Wikipedia authors write their ‘community’ into being. Mobilizing concepts regarding the communicative constitution of communities, the computer-mediated conversation between editors were investigated using Grounded Theory procedures. The analysis yielded an empirically grounded theory of the users’ self-understanding of the *Wikipedia community* as *ethos-action community*. Hence, this study contributes to research on online community-building as it shifts the focus from structural criteria for communities to the discursive level of community formation.

## Key words

computer-mediated communication, Grounded Theory, online community, Wikipedia

## Introduction

In his essay on the concept of community Zygmunt Bauman started with the remark that ‘Words have meanings: some words, however, also have a “feel”. The word “community” is one of them. It feels good’ (2001: 1). This particular quality, he argued, provokes the discrepancy between the perceived omnipresence of communities and the ambiguity of the concept’s meaning. Thus, Bauman points to a perplexing development – ‘never has the word “community” been used more indiscriminately and emptily than in the decades when communities in the sociological sense became hard to find in real life’, as Hobsbawm (1994: 428) put it.

---

## Corresponding author:

Christian Pentzold, Thueringer Weg 11, D-09126 Chemnitz, Germany.

Email: [christian.pentzold@phil.tu-chemnitz.de](mailto:christian.pentzold@phil.tu-chemnitz.de)

The advent of modern information and communication technologies like the internet has further promoted the proclamation of so-called 'communities'. The idea of 'on-line communities' (Licklider and Taylor, 1968: 37) coming to life via computer-mediated communication extends as far back as the late 1960s and was popularized by Rheingold's (1993) seminal book animating a burgeoning body of literature discussing contexts and processes of online community-building (for amplification see e.g. Jankowski, 2002; Jones, 1998; Preece and Maloney-Krichmar, 2005). Nevertheless, such line of reasoning has been confronted with the sociological origin of the term as a specifically space-centred concept. Tönnies (1889/2001) saw the community as an organic form of fellowship within a shared physical space allowing face-to-face interactions. This narrow understanding has been promoted by critics of a generous usage of the label 'online communities' who wished to assign the term only to locally confined, multiplex social relations (e.g. Calhoun, 1998).

One way of addressing such criticism is to adjust the concept. For example, Wellman and Gulia (1999: 333) proposed that 'communities do not have to be solitary groups of densely knit neighbors but could also exist as social networks of kin, friends, and workmates who do not necessarily live in the same neighborhoods'. In the same vein, Castells (2001: 127) explicated: 'the major transformation of sociability in complex societies took place with the substitution of networks for spatial communities'. In their attempt to set new definitions both re-interpretations aim at the structural level regarding the necessary conditions collectives should exhibit so to be termed 'communities', like the number and strength of ties, the stability of contact or the reciprocity of exchanges.

The argument put forward in this article, however, points to another direction. A second way of engaging with the criticism regarding potential communities is to look at the creation of the idea that there are such things as communities. More precisely, I will examine the communication between participants of the online project Wikipedia in order to find out how they construct their self-understanding and self-description as 'community'. Hence, the analysis follows an argumentation stressing the importance of the symbolic constitution of communities in communication besides and above their structural characteristics. In doing so, the focus is shifted to discourse and social meanings.

## **Background: the communicative construction of communities**

The turn away from structural or functional criteria to the symbolic dimension of community-building is reflected in Anderson's study on *Imagined Communities* (1983/2003). Although he was mainly concerned with the rise of nation-states, his approach had an extensive influence on thinking about the creation of collectives generally. Anderson defined the nation as an 'imagined political community' (1983/2003: 6) because its members would never know most of their fellows. Thus, the historical key to its existence, he postulated, was the production and distribution of print products. In fact, Anderson declared that communities need not be underpinned by spatial coexistence and social intimacy to establish a sense of togetherness. In order to form a community, locally separated individuals do not necessarily have to rely on face-to-face interaction, therefore, but on the 'subjective feeling ... that they belong together', as Weber (1922/1978: 40) had already noted. Such an imagined community is fictitious in the sense that its

perception unfolds largely apart from direct interaction. Yet, to say that it belongs to an imagined world does not compromise its social relevance – people act on their understanding of social relations. Because Anderson assumed widespread anonymity and because he claimed a crucial role for communication technologies in connecting the geographically dispersed community members his concept has been widely received in the writings on online communities (Baym, 1999; Feenberg and Bakardjieva, 2004).

The character of communities as imagined entities has been further explored by Cohen (1985). To understand the importance of a community in peoples' experience he focused on the awareness collectives have of themselves with respect to others. He argued that the consciousness of community is largely symbolic. This symbolic dimension can take an explicit shape in rituals, totems, or memorials. Yet, Cohen also pointed to a more mundane form when he noticed that 'much of our symbolism does not have a special vocabulary or idiomatic behaviour: it is, rather, part of the meaning which we intuitively ascribe to more instrumental and pragmatic things in ordinary use – such as words' (1985: 14). Consequently, the word 'community' itself becomes an object of interest when it is understood as 'something more than a rhetorical figment' (1985: 13). In other words: The term 'community' not only denotes an entity but it also transports associated social meanings and attitudes. As a symbol 'community' not only represents something else because '[s]ymbols do not so much express meaning as give us the capacity to make meaning' (1985: 15).

Taken together, these two approaches contribute to the study of community-building with the insight that a collective's self-understanding and self-description as 'community' is constructed in interaction and communication rather than residing in structures only. The investigation of these processes thus transcends any attempt to juxtapose concepts like 'community' or 'network' regarding their scholarly significance in capturing a particular social phenomenon. To look at the creation of a community on the semantic level does not foreclose, for instance, to view and analyse the same social constellation in different terms because this approach does not aim at the doctrinal level of mutually exclusive theoretical definitions. Instead, the examination looks at the understanding of 'community' developed by the actors themselves. In doing so, the study follows Fernback (1999: 205; see also Jones, 1998) who, in reference to Cohen, suggested a turn toward the evolving formation of online communities in 'the communicative process of negotiation and production of a commonality of meaning, structure, and culture'.

## **The creation of the *Wikipedia* community**

When accessing the *About Wikipedia* website, the visitor is told several times that this resource originates from the efforts of the 'Wikipedia community' (Wikipedia, 2010c). Moreover, to receive further advice on how to contribute, the potential author is pointed to a so-called 'community portal' (Wikipedia, 2010f). And on the page 'Wikipedia is a community' the user is assured that Wikipedia is 'more than *just* an encyclopedia, *Wikipedia is a community*' (Wikipedia, 2010b, original emphases).

This study was designed to examine the collective interpretation of Wikipedia as *Wikipedia community*. In consequence, the leading question was: What particular meaning do the Wikipedia editors attach to the term 'community'? The study analysed the

communicative construction of this particular self-description as the way the Wikipedia authors write their community into being.<sup>1</sup> In focusing on the semantic dimension of community-building the analysis followed the theoretical implications in two ways.

First, it was assumed that the belief that the particular term is an accurate name for the relationship is not a natural given but must be negotiated – the editors could have supposedly employed other expressions like ‘group,’ ‘network,’ or ‘team’ in order to understand and describe themselves. Instead, the positively connotated ‘community’ prevails. Consequently, the investigation did not focus on mere verbal decoration or accidental word choice. Instead, it supposed ‘semantic and programmatic concerns’ (Rose, 1999: 167) which mattered in establishing the meaning of the symbol ‘community’ in this specific context. These concerns, it was posed, generally revolve around its basic character as ‘the warmly persuasive word to describe an existing set of relationships’ which ‘seems never to be used unfavourably’, as Williams (1976/1983: 76) diagnosed. Hence, it does not come as a surprise that the term is widely employed to name the customers of a service or brand a ‘community’ (Fernback, 2007).

Second, it was held that ‘community’ is an abstract term without a clear reference object (unlike, e.g. ‘tree’ or ‘apple’). This open extension causes its perceived elusiveness. Yet, it does not mean that the expression conveys no meaning at all – instead, it is open to various different definitions, often in connection with value statements evoking some positive connotations of good will, communion and sympathy. In order to analyse the acceptance of the word, its particular usage has to be explored. Thus, the examination builds on the insight prominently voiced by Wittgenstein (1953/2003: § 43): ‘the meaning of a word is its use in the language’ which was elaborated in Williams’ *Keywords* (1976/1983) using the example of ‘community’. An inquiry into vocabulary is a promising enterprise, he argued, because social issues ‘cannot be focused unless we are conscious of the words as elements of the problems’ (1976/1983: 16). The formation of meanings emerging in the contextual usage of the expression ‘community’ cannot be clarified by consulting dictionaries only. Because it involves values and beliefs we find ‘a history and complexity of meanings’ (1976/1983: 21) that need to be reconstructed. Consequently, in order to examine the word’s contextual meanings, the study looked at the communication between the editors with a focus on the occasions when they alluded to the ‘community’ of Wikipedia authors. In fact, the users’ computer-mediated, text-based conversations were seen as the instances where their shared understandings become manifest.

In its focus on the communicative construction of the social meanings of ‘community’, this study differs from other discourse-centred research on online community-building. For instance, in her work on the communicative practices of a fan newsgroup, Baym (1999) studied a broader range of beliefs, purposes and aims constructed in language activities. Different to that, McLaughlin, Osborne and Smith (1995) as well as Reid (1998) looked at the ways newsgroup subscribers and MUD players discursively worked out their interaction standards.

## Methodology and method

The analysis was based on Grounded Theory procedures (GT: Glaser and Strauss, 1967; Strauss and Corbin, 1990). This qualitative methodology was chosen for two reasons:

First, its strength lies in the ‘analytical power to theorize how meanings, actions, and social structures are constructed’ (Charmaz, 2006: 151). Consequently, it lends itself to an interpretative examination of social meanings. Second, although GT is first of all a methodology and not a definite method, it nevertheless enabled valid and transparent interpretations of the data through a systematic set of procedures combining the reciprocal and continuous collection, coding and comparison of the material.

The analysis was conducted with the support of the CAQDAS Atlas.ti which has specifically been designed in accord with GT methodology. Different to the linear composition of the following report, the analytical work moved back and forth between its different stages. A second tension is posed by the report’s formal restrictions and the methodological necessity to capture the fullness of the analytical renderings and to provide vivid descriptions in order to convey the credibility of the discovered theory. In presenting such a theory it does not suffice to state the final results because the ‘canon for judging the usefulness of a theory is how it was generated’ (Glaser and Strauss, 1967: 5). Here, this demand has been addressed in the following manner. Firstly, the material is introduced. Secondly, I will reflect on the coding procedures discussing the major categories while omitting a thorough overview of the complete set of codes and categories. To substantiate the argument, examples are given. Thirdly, building on that, the empirically grounded theory of the *Wikipedia community* is presented.

While coding in GT traditionally relies on a set of rather general questions to characterize a phenomenon, the present analysis employed a more comprehensive catalogue which aimed to grasp community-related aspects. Hence, the first step of coding was guided by a predefined set of queries.<sup>2</sup> This register provided a heuristic starting point for the coding process and thus supported the ‘awareness of the subtleties of meaning of data’ (Strauss and Corbin, 1990: 41) as demanded by GT.

**Table 1.** List of coding questions

Q 1. In which social processes does the community play a role?	Q 13. What is its scope?
Q 2. What are its roles/functions?	Q 14. What are its origins?
Q 3. What are its parts/manifestations?	Q 15. What are circumstances for its dissolution?
Q 4. What people are interested in the community?	Q 16. What are possibilities for its destruction?
Q 5. In which parts can people get involved?	Q 17. What are typical phases of existence?
Q 6. How are people involved?	Q 18. What are circumstances for its spread?
Q 7. In which activities can people partake?	Q 19. What are circumstances for its usual existence?
Q 8. In which roles/functions can people partake?	Q 20. What are its names?
Q 9. What are advantages for members?	Q 21. What are the intentions, needs, aims and reasons leading to its existence?
Q 10. What are prohibitions?	Q 22. In which contexts does it play a role?
Q 11. What are duties?	Q 23. What are its roles/functions in these contexts?
Q 12. What other characteristics does the community have?	

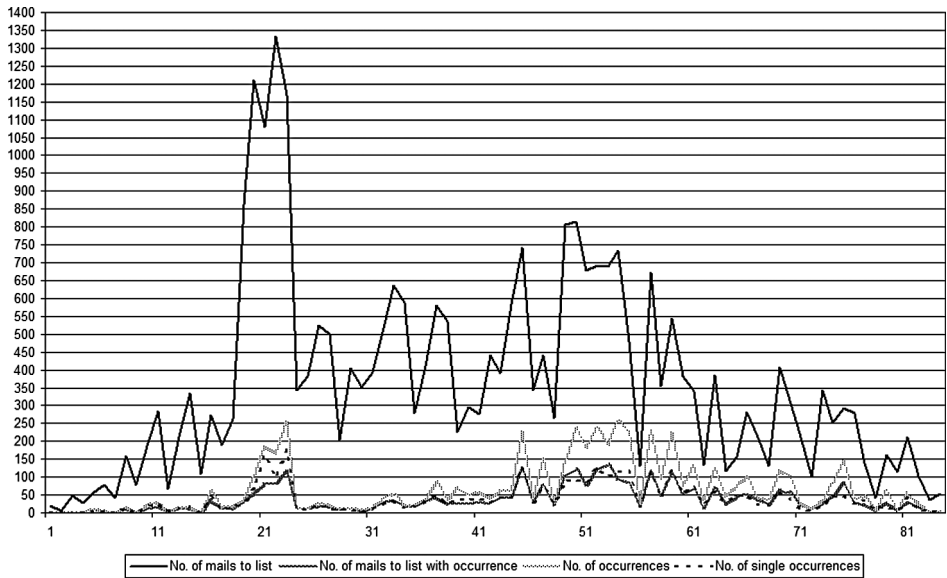
Source: Adapted from Konerding, 1993.

In GT, coding means to single out passages representing ideas, to ask questions about each unit as well as to compare and name them. Thereby, the material is conceptualized. Based on that, the emerging concepts are grouped and linked to each other, i.e. they are categorized. This phase of open coding is completed by steps of axial and selective coding where categories are developed along their characteristics, dimensions and relations. Hence, the GT approach here differs from a content analysis in that its focus is not on counting elements in collections of documents but on an in-depth interpretation and reconstruction of social meanings.

## Data

The material for the analysis was taken from the Wikipedia-l mailing list (Wikipedia, 2010a). This list was firstly chosen because it is the central public discussion forum for the overall Wikipedia project in all its language versions. According to Jimmy Wales (2001), the project founder, this mailing list should be 'regarded as the place for meta-discussions about the nature of Wikipedia'. While most of the other 170-odd active lists are dedicated either to single language editions or technical, editorial and organizational purposes, the Wikipedia-l is open to the general public of authors to debate issues of common interest. Hence, it presents a place where users can raise matters of wider concern and more fundamental importance apart from their editing activities in the wiki. Postings in all languages are permitted, but most of them are written in English. Because of its project-wide relevance it pools users who otherwise work separately in their respective language versions. This is not to say that the discussion about the understanding of Wikipedia as 'community' is limited to this particular list. There are supposedly numerous wiki pages where the authors exchange their opinions on that matter but the messages posted to the Wikipedia-l are an ideal starting point because the list was already in use when Wikipedia was just a small but rapidly growing set of websites where the initial users shared their views on the project. Arguably, the interpretations developed in these discussions were then diffused and thus recur in other places of the wiki, too. Secondly it was chosen because it is the oldest mailing list of Wikipedia. It was initiated with a post on 22 January 2001; the project itself was only founded one week earlier on 15 January. Although it has lost traffic due to the growing prominence of other lists, the Wikipedia-l thus covers the project's entire existence.<sup>3</sup> Thirdly, a practical argument supports the choice: All messages are completely archived in text-files facilitating the data access, retrieval and editing for the coding process.

In the 84 months between its initiation and the end of the period investigated on 31 December 2007, 30,500 messages were sent to the list. Yet, not all of them contain the word 'community'. Given the study's focus, the first task was to filter the occurrences and associated context. This context was operationalized as the complete text of an email in which the word had been used at least once. Out of 30,500 posts 3105 contained the word. Altogether, the authors used the expression in 5663 passages. Of these, 3221 are unique single occurrences where the authors employed the word in a certain context for the first time, whereas 2442 of the occurrences are quotations in later emails (see Figure 1).<sup>4</sup>



**Figure 1.** The total number of monthly emails and the distribution of the occurrences of 'community'

In GT, data collection is neither based on the complete population of cases nor on a representative partial sample. Instead, theoretical sampling is introduced as a set of procedures connecting the coding with a continuous sampling of further incidents. Thus, the sampling is not a self-contained step of analysis before the proper examination. Both are rather understood as interconnected and cumulative phases. The study relied on all three coding and corresponding sampling procedures (Strauss and Corbin, 1990: 180–193). First, open sampling was employed to discover as many potentially relevant categories as possible. Consequently, the collection was rather indiscriminate: all relevant messages of the second month in each succeeding year were sampled so to include the mails sent in February 2001, April 2002, June 2003, and so forth. During the progressing analysis, additional material was selected according to open sampling and, in later phases, also to axial and discriminative sampling procedures.<sup>5</sup> However, not all occurrences in these monthly collections were automatically and completely sampled, but their incorporation into the corpus was always based on their theoretical relevance. Hence, to say that a total of 27 months with 1660 occurrences was part of the sample is no meaningful statement because the coding did not aim at a complete evaluation of all instances. Instead, the judgement when to stop sampling was based on the development of the theory. When no additional relevant material was found for one of the categories, no further material was coded ('theoretical saturation': Glaser and Strauss, 1967: 61). Thus, better information would be the number of applied codes (250) and coded text passages (1312). However, they only become relevant in an account of the coherent analytical process which is discussed below.

## Analysis

### a) Coding

The analysis started with the register of 23 questions (Table 1). They were used as concepts, i.e. as labels placed on all discrete phrases interpreted as relevant passages. This first phase of semi-structured coding began with the systematically selected sample containing the archived email texts from seven months covering the complete timeframe from 2001 to 2007. For instance, passages like ‘Personal attacks, though, have no place in our community’ (April 2002)<sup>6</sup> or ‘DO NOT CHANGE PICTURES!!’ (October 2005) were coded as possible prohibitions for community members (Q 10). During this process of breaking down the material into phrases and conceptualizing them, the data showed that, on the one hand, some concepts were often employed while others remained without quotations, and, on the other, that the coded material was rich in detail. Hence, despite its value for enhancing theoretical sensibility and opening up the data, the list of questions was for some aspects too specified while for others not particularized enough. Especially the concepts regarding the groups of interested people (Q 4) and the way they could become involved (Q 6), the parts of the social entity (Q 5), prohibitions for (Q 10) as well as functions of its members (Q 8), phases (Q 17), circumstances for its existence (Q 19), and possibilities to destroy it (Q 16) were well grounded. Furthermore, the actors were creative in finding other names for the collective (Q 20) to highlight certain aspects of their understanding of the *Wikipedia community* like ‘community of nice people’ (April 2002), ‘rational beings who understand the purpose of Wikipedia’ (April 2002) or ‘big pool of passionate people who believe in the freedom of information and sharing knowledge’ (August 2004).

In order to account for the empirical richness of these 10 concepts, they were further specified in a step following the basic analysis of naming and conceptualizing. Thus, the examination went beyond the predefined questions toward evolving concepts.

The data showed that the editors referred to different parts of the *Wikipedia community* when they talked about the ‘Finnish wikipedian community’ (February 2007) in contrast to the ‘(international) Wikipedian community’ (October 2005). Moreover, the authors related these concepts, for instance, when they wanted to ‘consult with the wider community’ (October 2005) in order to discuss problems between ‘ro.wikipedians’ (October 2005) from the Romanian version. Therefore, the concept regarding the parts of the social entity (Q 5) was elaborated to accommodate the discovered plurality of meanings. Labels not only for the two ideas of different *language communities* and an *international community* but also provisional concepts referring to the additional ideas of a *core community* (“sysops” or better terminology might be “community member”, April 2002) and of a *community of users with access rights to the MediaWiki software* (‘volunteer developer community’, April 2002) were introduced. This process of examining via constant comparing resulted in a collection of 64 basic conceptual units mapping initial parts of the communicatively constructed social meanings of the *Wikipedia community*.

Based on that, possible connections between the codes were proposed and they were organized into categories representing particular phenomena. Again, labels to name these categories were formulated. Thereby, the four categories *ethos-community*, *language*



*community*, *technical community* and *expert community* seemed to play an integral role – integral in the sense that they rendered the data most effectively. The first of them referred to the group of concepts joined around an understanding of ‘community’ as open body without clearly demarcated boundaries. Interested people can join it by committing themselves to a set of norms, behavioural standards and attitudes – by sharing its ethos. Sometimes, the authors referred to a specific rule (‘exhibiting the behaviour that we expect of good Wikimedians, tolerance and NPOV’, December 2006), sometimes a more general set of expectations and standards was only implicated (‘Much of my rant was based on the fact that [username] does not follow community standards and etiquette’, April 2002). Some of the authors also seemed to balance the status of rules and moral standards. For example, one argued that ‘A rule is a rule when it’s clearly expressed and people are aware of it. A rule that is not written is ... a desire, a convention’ (February 2007). The second category, *language community*, identified the cluster of meanings containing the idea that there are many Wikipedia versions and related communities which may differ with and correspond to each other in a range of variables – e.g. language, culture, rules, number of articles and editors. The third and fourth, *technical community* and *expert community*, have been devised so to consider, on the one hand, the idea that there is a community only comprising a core group of technical access rights holders (‘the developer group de facto determines what changes are made to the Wikipedia software’, April 2002) and, on the other, the idea of a community of experts contributing their special knowledge to the encyclopaedia (‘there are several scientists who work on wikipedia and are not scared away’, August 2004). Yet, at this point of the analytical work the connections between these categories and subcategories were still vague.

Consequently, in the following step, these categories were developed and possible relations between the major categories and their subcategories explored. Configuring potential characteristics and their dimensions helped to recognize and explicate the supposed connections. This procedure of axial coding can be demonstrated by taking a section from the subcategory *language boundary* that classified all concepts considering the differences between single language versions: It was expanded along attributes like *consequence* (What results from the differences?) or *contact* (What are the relations between the versions?). These characteristics were then complemented with significant dimensions – *consequence* was, for instance, located on continuums from ineffective to revolutionary, from destructive to beneficial, from total to zero; *contact* fluctuated between locked and completely open, constant exchange and no exchange, or balanced and unidirectional. Using such contrasting dimensions, further material was sampled and coded.

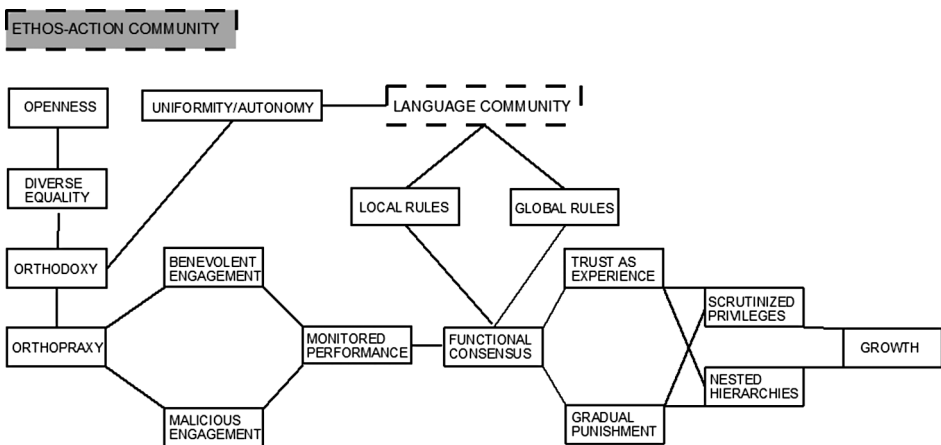
This phase of coding toward the properties pertaining to the phenomena yielded another 125 codes. The data indicated that *ethos-community* should be raised to the core category. Not only did it account for a substantial part of the data and thus was the most evident category, in comparison with the other categories it also had the strongest generic power to systematically integrate them. Yet, other hitherto peripheral attributes gained importance: Although the pivotal aspect of the evolving theory remained the users’ commitment to an ethos, categories of benevolent and malicious actions had to be established according to the coded material. This led to a reformulated *ethos-action community*. In that sense, the *Wikipedia community* is formed by individuals sharing basic views on aims, values, and norms. These attitudes become apparent and assessable in their performances.

Furthermore, the four major categories could no longer be seen as referring to different entities and as forming incompatible understandings of the *Wikipedia community*. Instead, the analysis of their relations suggested a fundamental and dominating function of the *ethos-action community*: Contributors from all different backgrounds can join Wikipedia in one of its *language communities*. These versions vary to a certain degree and regarding a range of criteria. People can become part of the *Wikipedia community* by an active commitment to a set of aims, norms and rules whose status, however, is controversial. In contrast, *technical community* and *expert community* lost their importance because of their limited significance in the data. The first seemed to play a role in setting and controlling the technical basis for the establishment of new language versions, the second formed a special author segment with an ambiguous status in terms of its commitment to the project’s ethos and position in Wikipedia’s hierarchy.

**b) Wikipedia as ethos-action community**

The last step of the analysis focused on elaborating the connections between the categories and subcategories. Thus, the theory emerged as ‘concepts related through statements of relationship, which together constitute an integral framework that can be used to explain or predict phenomena’ (Strauss and Corbin, 1998: 15).

Based on the discovered network of categories an empirically grounded theory of the *Wikipedia community* can be narratively laid out (see Figure 2): The collective responsible for the existence and growth of Wikipedia is perceived as *ethos-action community*. Membership and thus the boundaries are defined by adherence to a set of standards regarding the project’s purpose, norms, values, and valid actions. The ethos itself is not an equally accepted block of statements. Its loosely coupled elements rather consist of parts which are presupposed while others need to be explained when used by the authors. Hence, editors alluded to general ‘ideas of openness, fairness, objectivity, community



**Figure 2.** The relations of categories. Although there are some procedural interrelations, it does not visualize a causal process but the most relevant connections

consensus' (October 2001) but also to 'existing NPOV guidelines' (July 2003) addressing a formalized Wikipedia policy regarding the neutrality of entries.

The *Wikipedia community* is open for any volunteer willing to accept the code of conduct. In turn, this *openness*, as the concept was labelled, is seen as warranting the influx of a broad range of capable contributors intending to add their knowledge to the various and growing parts of the encyclopaedia. Despite their potentially different backgrounds (social, cultural, religious, educational, etc.), these individuals are seen as equal in their fundamental rights to edit the wiki. Yet, this *diverse equality* is constrained in one aspect – the set of required attitudes and standards. The individuals' commitment to this *orthodoxy*, however, cannot be checked before their entry because this would already violate the project's openness. Hence, what are assessed are their actions which can be performed in accord or in conflict with the expectations. Therefore, because orthodoxy cannot be evaluated, *orthopraxy* becomes the basis for deciding upon the status as *Wikipedia community* member.<sup>7</sup> The actions individuals might choose are not only understood as compliance or contradiction but also as being *benevolent engagement* or *malicious engagement*. Whereas the first is seen as contributing to the growth of the encyclopaedic content, the second is perceived as possibly requiring disciplinary reactions. The judgement if a certain action is rated as conducive or obstructive is based on the automatically saved record of all activities in the wiki. For example, one author stated: 'We'll track it for one week, to see how it goes' (January 2004). Every action is, therefore, *monitored performance*. If members act in accord with the ethos, they can accumulate trust. Thus, although everybody is welcome to contribute, *trust-as-experience* only gradually builds up, based on benevolent engagement. In other words, users are not trusted prior to their actions but, rather, they have to convince others of their trustworthiness. Moreover, trust-as-experience lies at the heart of Wikipedia's *nested hierarchies*. Despite the idea that the editors are equal, the authors differentiated groups of users with special rights. The levels of this hierarchy are understood as nested in the sense that a user whose monitored performances are valued by fellow members and who is therefore trusted can gradually work their way up the command structure. 'If we want to give "reliable" users privileges that we don't immediately give to everyone, that's great' (October 2001), as one user put it.

Although members of the ethos-action community can be promoted to higher functions, their authority depends on *scrutinized privileges*. Consequently, if they violate the standards, this malicious engagement is registered too and might be *gradually punished*. Hence, not all sorts of actions are disciplined with the same penalty. Instead, these procedures are understood as being adapted to the perceived gravity of the malicious engagement. Thus, a granted status can be successively withdrawn. The decision, if someone takes up a position or if a certain rule should be installed, revised or abolished, is seen as *functional consensus* of the ethos-action community members. They are understood as acting on these matters without external intervention in a form of self-governance. This process is consensual because decisions should only be taken after discussing them. So one user declared: 'At the end of that week, we'll evaluate, including a community vote as to whether to include it or not' (January 2004). It is functional because this process is assumed to react to an emerging problem or task, not to predetermined issues.

From this background, the category of *language community* addresses the idea that there are separate collectives and Wikipedia versions which differ from each other in the

language of their entries and authors. Yet, this is seen as only the most apparent difference. Apart from that, the language communities vary in their *uniformity* and *autonomy* regarding the ethos. Moreover, they are supposed to have different sizes, establish different working mechanisms, standards of evaluation, and so forth – dynamics confirmed in corresponding studies (e.g. Ortega et al., 2007; Pfeil et al., 2006). A user can be a member of a language community – although people might be multilingual, the authors consider language communities as formed by native speakers who not only distinguish themselves from other participants in their mother tongue but also in a broader set of cultural aspects which sometimes cause ‘Quarrels between contributors from different countries’ (January 2004). Consequently, Wikipedia language communities are seen as not only linguistically different but also as diverse in terms of certain rules, working mechanisms and social arrangements. In that sense, local language communities have *local rules*. In contrast, the ethos-action community does not refer to a separate entity detached from any language community. Rather, it serves as a foil to evaluate the language communities’ degree of autonomy and uniformity with respect to *global rules* and as a framework connecting the separate editions. As such, the critical status of the ethos-action community becomes evident. In fact, it is understood as not being connected to any specific language community. Members of a language community have to share some principles and norms with users in other language communities to become and stay members of the encompassing ethos-action community. Their autonomy is limited in the sense that the code of conduct consists of local rules which can be adapted in individual language communities and global rules which should govern the interactions in all versions if they want to call themselves Wikipedia and Wikipedians – which means that the users acting correspondingly are part of the overall ethos-action community. The corresponding statements thus ranged from ideas like ‘The wikimedia foundation is for keeping the servers running ... but not for enforcing rules (or a however defined code of ethics) upon all projects’ (January 2004) to ‘neutrality policy ... That’s one point about Wikipedia that is non-negotiable’ (December 2002).

Yet, although the ethos-action community refers to a common ground shared by all language communities, its principles have to be drafted in the context of one particular language version. This context seems to be virtually always formed by the English-language Wikipedia whose standards are taken as guidelines for other communities. Assumably, several circumstances contribute to this role: It is the initial version, it is the largest edition in size and registered users, its working language is the *lingua franca* of the web and thus it is seen as the most inclusive Wikipedia edition. Nevertheless, this status is also questioned. For instance, some authors stressed the importance of the proportional representation of members from other communities in the decision-making process. In this line, one user reminded that ‘It is very important that steps be taken to ensure that those who don’t speak English have their interests well represented, there is no question about that’ (January 2004).

## Conclusion

The investigation started with the argument that a study of online community-building must not necessarily focus on structural criteria like the number of members or the

quality of their exchanges. Instead, it was proposed to turn toward the processes in which a collective establishes the conviction that 'community' is an apt description of its relationship and to look at how the notion is deployed. Thus, the investigation followed the assumption that the word 'community' itself becomes a symbol and therefore a critical device in the sense-making processes of Wikipedia contributors.

In particular, the study focused on the semantics of the term 'community' and examined the communication between Wikipedia editors. The analysis looked at how the participants employed the expression when alluding to the *Wikipedia community*. In order to discover the set of negotiated meanings the authors established in writing about themselves, the analysis was based on GT procedures because they allowed for interpreting and piecing together the meanings produced in the authors' conversations resulting in an inductive theory of the *Wikipedia community*. It emerged that the users primarily understand their collective as an *ethos-action community* tying community membership not to admission procedures but to the personal acceptance of a set of moral obligations and rules of conduct. The editors' commitment materializes in actions that are, in turn, evaluated, rewarded or sanctioned. The right action and the right thinking then become crucial for determining the community's boundaries. The normative standards guiding the members' behaviour in the projects are principles like 'Assume good faith' and the 'Neutral Point of View' (Reagle, 2010) as well as explicit policies (Wikipedia, 2010d). In fact, while a shared set of normative beliefs and common activities are standard features of sociological definitions of communities, the present study showed what form they take in the routines of a particular community-building process.

Hence, the study adds to the existing literature regarding the analysis of online community-building in two ways: First, it demonstrated how a rigorous GT analysis can be employed not only to classify data but to discover the contextual formation of a collective's shared self-understanding as community irrespective of any structural or functional criteria. Therefore, it differs from other Wikipedia studies using GT mainly to produce categories, like Pfeil, Zaphiris and Ang's (2006) examination of editing patterns, Beschastnikh, Kriplean and McDonald's (2008) work on policies, or Kriplean and colleagues' (2007) analysis of discussion pages. Unlike such approaches, this study applied the steps of coding provided by GT not only as ways of creating classificatory categories or of tagging data but as a method of theorizing. Hence, the analytical process unfolded as flexible accessing, sampling, structuring, linking, tentative conceptualizing and reviewing that resulted in the empirically grounded theory of the *ethos-action community* of Wikipedia authors. The interpretative procedures focused on material taken from the communication between the editors. Thus, the examination built on Berger and Luckmann's (1966) classical argument that verbal conversation is pivotal for the social construction of reality.

Nevertheless, such restricted data runs the risk of privileging one dimension of community-building and ignoring the multiple ways communities are enacted in practice. Hence, Jones (1998: 4) asks how the complex interpretative momentums which constitute the individual attachment and commitment to communities are to be ascertained when electronic discourse is captured and taken out of context. In consequence, pursuing further research into the formation of online communities should look at and engage with the manifold relevant practices in their processual (Schatzki, 2002). The social meanings

attached to a specific community neither manufacture themselves nor do they form a homogeneous ideational background. Rather, community-building happens partly as strategic and conscious manoeuvres, partly as habitual and unconscious accomplishments on the semantic as well as structural level. Moreover, an examination of these diverse practices would have to engage with their spatial and material arrangements, like the software code, that afford the establishment, maintenance and transformation of communities.

Second, the study showed how an inquiry into the vocabulary of 'community' sheds light into the collectivization of loosely connected individuals in an online environment via computer-mediated communication. The Wikipedia authors constituted their particular self-understanding and self-description as ethos-action community by eliminating other ways of thinking and talking about the collective, its activities and relationship as well as by demarcating the differences with other groups. Such boundaries encapsulate, as Cohen (1985: 12) remarked, the identity of a community. Reflecting on this insight, Bauman (2001: 14) argued that every communal understanding like that of an ethos-action community is 'only an achievement, attained (if at all) at the end of a long and tortuous labour of agreement'. Its homogeneity and unity needs to be made and other options have to be elided. Therefore, invoking a unified ethos-action community is more than just denominating a social entity but involves, as Joseph (2002: xix) stated, the enactment of communion which entails power struggles, hegemonic interpretations and practices as well as mechanisms of exclusion.

In this respect, thinking of Wikipedia as an ethos-action community involves the insight that this particular set of meanings not only orders the thinking and saying but an even broader range of activities of its members. Being committed and emotionally attached to a community not only grounds in a sense of togetherness and a similar way of making sense of things (Cohen, 1985: 16) but ultimately rests on shared actions. Correspondingly, the ethos-action community is, in Bauman's (2001: 72) terms, an 'ethical community' weaving a web of responsibilities between its adherents. Because it is of vital importance for the project to secure active support, its members have to be bound into long-term commitments. Arguably, Wikipedia's need for obligations, loyalty, moral bonds and responsibilities for conduct is high because the project's openness ensures the exit option and thus complicates any penalization. Hence, in future research Wikipedia's 'government through community' could be developed along Rose's (1996, 1999) Foucauldian-inclined analysis which considers community as a territory of administration because it addresses its members as 'subjects of allegiance to a particular set of community values, beliefs, and commitments' (1996: 331). Its particular quality with respect to alternative forms of governance is that the community appeals to its subjects as self-responsible 'moral individuals' who are supposed to identify with the relevant norms and thus govern their actions (and thinking) to become and stay community members. Turning to Wikipedia, it can be argued that the autonomous authors are made amenable to administrative actions by a language of 'community' which ties community membership to compliance with a set of norms and values – orthodoxy and orthopraxy. The decision to join the *Wikipedia community* by practicing benevolent engagement rests with the authors themselves who have to embrace the local and global rules so to become a trusted member. The *Wikipedia community* binds its members in a network of allegiance. The project's self-governance is, in this respect, attained by self-governing

editors. Therefore, it does not come as a surprise that some of the facets of the theory resemble Ostrom's (1990) principles of successful self-governance regimes guarding commons resources. That Wikipedia's governance can be interpreted in the light of Ostrom's work has already been demonstrated in some earlier studies (Forte et al., 2009; Viegas et al., 2007). Yet, this investigation additionally shows that governance principles like monitored performances, gradual punishment or nested hierarchies have also found their way into the shared self-understanding of the *Wikipedia community*. Building on that, research should concentrate on integrating different approaches to investigate online governance. While the present study highlighted social meanings and normative obligations it only took a glance at the ways governance is enacted via the formation of hierarchies or the specification of rules.

Finally, if a community is not naturally given but artificially crafted, its members have to be made aware of their allegiance (Rose 1996: 334). To be able to identify with the community and to commit themselves to its ends, norms and procedures, prospective members have to be socialized. Consequently, Wikipedia contributors do not only have to learn to use the software tools, but they also have to acquire the appropriate beliefs, values, common understandings and practices (Bryant et al., 2005). People can be active in Wikipedia without being a member of the community because it takes more than edits to be a *Wikipedian* – membership is based on compliance. Thus, the English-language Wikipedia welcomes novices with 'Welcome aboard! We're glad you wish to help develop Wikipedia, and hope you enjoy participating in its community as much as we do' (Wikipedia, 2010e).

## Acknowledgments

I thank Malte Ziewitz, the editors and anonymous reviewers for their comments as well as Christian Groß and André Wulf for their research assistance. The financial support from the German National Academic Foundation and the British Engineering and Physical Science Research Council ('Networks for Web Science grant', EP/FO/3701/1) are gratefully acknowledged.

## Notes

- 1 I owe this phrase to boyd (2006).
- 2 A methodical remark: The questions were derived from research in linguistic semantics. In particular, scholarship into the communicative contextualization of knowledge provided the background for the analysis of the meanings attached to the term. Following Fillmore (1982), such relational knowledge can be modelled as a list of questions. For elaboration see Fillmore et al. (2003) and Konerding (1993: 425-34). Doing so implies a shift from the GT standard of maximized openness to a semi-structured analysis. This has caused some debate in GT. See Hood (2007).
- 3 Since July 2009 there are 1052 list members. For more information <http://www.infodisiac.com/Wikipedia/ScanMail/Wikipedia-l.html>.
- 4 The texts were parsed with a (Java-based) Processing script for the truncated character strings \*ommunit\*, \*ommuniz\* and \*ommuny\* covering, according to the Oxford English Dictionary, the lexical field of community-related English words.

- 5 In these progressive steps of sampling, data with an above-average, below-average and typical usage of the term was selected. Hence, months were chosen according to statistical evidence about mean and median numbers, standard deviation as well maximum number of emails containing the term per month ( $m = 36.96$ ,  $mn = 29.00$ ,  $SD = 35.098$ ,  $max = 135$ ,  $R = 135$ ) and of single occurrences of the term per month ( $m = 38.35$ ,  $mn = 25.00$ ,  $SD = 39.534$ ,  $max = 177$ ,  $R = 177$ ). To further broaden the sample, the engagement of identifiable key actors (J. Wales, Larry Sanger, Wikipedia's co-founder, paid staff and the Board of Trustees) was traced. Hence, it is correct to say that the data has a bias toward Wikipedia's power elite.
- 6 For quotes from the material, the relevant month is stated.
- 7 Both expressions were derived from theology: orthodoxy – 'correct belief'; orthopraxy – 'correct action'.

## References

- Anderson B (1983/2003) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.
- Bauman Z (2001) *Community: Seeking Safety in an Insecure World*. London: Polity.
- Baym NK (1999) *Tune In, Log On: Soaps, Fandom, and Online Community*. Thousand Oaks, CA: Sage.
- Berger PL and Luckmann T (1966) *The Social Construction of Reality. A Treatise in the Sociology of Knowledge*. Garden City, NY: Doubleday.
- Beschastnikh I, Kriplean T and McDonald DW (2008) 'Wikipedian Self-governance in Action: Motivating the Policy Lens'. Paper presented at the International Conference on Weblogs and Social Media, Seattle, WA, 30 March–2 April 2008.
- boyd d (2006) *Friends, Friendsters, and Myspace Top 8: Writing Community into Being on Social Network Sites*. *First Monday* 11(12). Available at: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1418/1336>
- Bryant S, Forte A and Bruckman A (2005) 'Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia'. Paper presented at the ACM SIGGROUP, Sanibel Island, FL, 6–9 November 2005.
- Calhoun C (1998) Community without propinquity revisited: Communications technology and the transformation of the urban public sphere. *Sociological Inquiry* 68(3): 373–397.
- Castells M (2001) *The Internet Galaxy: Reflections on the Internet, Business, and Society*. Oxford: Oxford University Press.
- Charmaz K (2006) *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Thousand Oaks, CA: Sage.
- Cohen AP (1985) *The Symbolic Construction of Community*. London and New York: Routledge.
- Feenberg A and Bakardjieva M (2004) Virtual community: No 'killer implication', *New Media & Society* 6(1): 37–43.
- Fernback J (1999) There is a there there. Notes toward a definition of cybercommunity. In: Jones SG (ed.) *Doing Internet Research*. Thousand Oaks, CA: Sage, 203–220.
- Fernback J (2007) Selling ourselves? Profitable surveillance and online communities. *Critical Discourse Studies* 4(3): 311–330.
- Fillmore C (1982) Frame semantics. In: Korea LO (ed.) *Linguistics in the Morning Calm*. Seoul: Hanshin, 111–137.



- Fillmore CJ, Johnson CR and Petruck MRL (2003) Background to framenet. *International Journal of Lexicography* 16(3): 235–250.
- Forte A, Larco V and Bruckman A (2009) Decentralization in Wikipedia governance. *Journal of Management Information Systems* 26(1): 49–72.
- Glaser BG and Strauss A (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine.
- Hobsbawm E (1994) *The Age of Extremes: A History of the World, 1914–1991*. London: Michael Joseph.
- Hood JC (2007) Orthodoxy vs. power: The defining traits of grounded theory. In: Bryant A and Charmaz K (eds.) *The Sage Handbook of Grounded Theory*. London: Sage, 151–64.
- Jankowski N (2002) Creating community with media. histories, theories, and scientific investigations. In: Lievrouw LA and Livingstone PS (eds) *Handbook of New Media*. London: Sage, 34–49.
- Jones SG (1998) Information, internet, and community in the information age: Notes toward an understanding of community. In: Jones SG (ed.) *Cybersociety 2.0*. Thousand Oaks, CA: Sage, 1–34.
- Joseph M (2002) *Against the Romance of Community*. Minneapolis: University of Minnesota Press.
- Konerding K.-P (1993) *Frames Und Lexikalisches Bedeutungswissen*. Tübingen: Niemeyer.
- Kriplean T, Beschastnikh I, McDonald DW and Golder SA (2007) Community, Consensus, Coercion, Control: CS\*W or How Policy Mediates Mass Participation. Paper presented at the Group 07, Sanibel Island, FL.
- Licklider JCR and Taylor RW (1968) The computer as a communication device. *Science and Technology* 76: 21–31.
- McLaughlin M, Osborne KK and Smith CB (1995) Standards of conduct on Usenet. In: Jones SG (ed.) *Cybersociety*. Thousand Oaks, CA: Sage, 90–111.
- Ortega F, Gonzalez-Barahona JM and Robles G (2007) Quantitative Analysis of the Top Ten Wikipedias. Paper presented at the ICISOFT/ENASE 2007, Barcelona.
- Ostrom E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Pfeil U, Zaphiris P and Ang CS (2006) Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer-Mediated Communication* 12(1): 88–113.
- Preece J and Maloney-Krichmar D (2005) Special issue: Online communities: Design, theory and practice. *Journal of Computer-mediated Communication* 10(4).
- Reagle J (2010) *Good Faith Collaboration: The Culture of Wikipedia*. Cambridge, MA: The MIT Press.
- Reid E (1998) Power and the evolution of social control in cyberspace. In: Kollock P and Smith MA (eds) *Communities in Cyberspace*. London and New York: Routledge, 107–134.
- Rheingold H (1993) *The Virtual Community: Homesteading on the Electronic Frontier*. New York: Addison-Wesley.
- Rose N (1996) The death of the social? Re-figuring the territory of government. *Economy and Society* 25(3): 327–356.
- Rose N (1999) *Powers of Freedom: Reframing Political Thought*. Cambridge: Cambridge University Press.
- Schatzki TR (2002) *The Site of the Social: A Philosophical Account of the Constitution of Social Life and Change*. University Park: Pennsylvania State University Press.

- Strauss, AC and Corbin J (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Thousand Oaks, CA: Sage.
- Strauss AC and Corbin J (1998) *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (2nd edn). Thousand Oaks, CA: Sage.
- Tönnies F (1889/2001) *Community and Civil Society*. Cambridge: Cambridge University Press.
- Viegas FB, Wattenberg M, Kriss J and van Ham F (2007). Talk Before You Type: Coordination in Wikipedia. Paper presented at the HICSS.
- Wales J (2001) *Statement of Principles*. Available at: <http://en.wikipedia.org/wiki/index.php?curid=16257> (consulted 20 January 2010).
- Weber M (1922/1978) *Economy and Society: An Outline of Interpretive Sociology*. Berkeley: University of California Press.
- Wellman B and Gulia M (1999) Net surfers don't ride alone: Virtual communities as communities. In: Wellman B (ed.) *Networks in the Global Village*. Boulder, CO: Westview Press, 331–367.
- Wikipedia (2010a) *Wikipedia-L*. <http://lists.wikimedia.org/pipermail/wikipedia-l/> (consulted 20 January 2010).
- Wikipedia (2010b) *Wikipedia Is a Community*. Available at: [http://en.wikipedia.org/wiki/Wikipedia:Wikipedia\\_is\\_a\\_community](http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_a_community) (consulted 20 January 2010).
- Wikipedia (2010c) *Wikipedia\$: About*. Available at: [http://en.wikipedia.org/wiki/About\\_Wikipedia](http://en.wikipedia.org/wiki/About_Wikipedia) (consulted 20 January 2010).
- Wikipedia (2010d) *Wikipedia:List of Policies and Guidelines*. Available at: [http://en.wikipedia.org/wiki/Wikipedia:List\\_of\\_policies\\_and\\_guidelines](http://en.wikipedia.org/wiki/Wikipedia:List_of_policies_and_guidelines) (consulted 20 January 2010).
- Wikipedia (2010e) *Wikipedia:Welcoming Commitee*. Available at: [http://en.wikipedia.org/wiki/Wikipedia\\_welcome](http://en.wikipedia.org/wiki/Wikipedia_welcome) (consulted 20 January 2010).
- Wikipedia (2010f) *Wikipedia\_Community Portal*. Available at: [http://en.wikipedia.org/wiki/Wikipedia:Community\\_portal](http://en.wikipedia.org/wiki/Wikipedia:Community_portal) (consulted 20 January 2010).
- Williams R (1976/1983) *Keywords: A Vocabulary of Culture and Society*. London: Fontana.
- Wittgenstein L (1953/2003) *Philosophical Investigations*. London: Wiley-Blackwell.

Christian Pentzold is a doctoral candidate at the Institute for Media Research, Chemnitz University of Technology, Germany. In 2009 he was a visitor to the Oxford Internet Institute, University of Oxford. His research interests are in the areas of media sociology, media economy and governance in digitally mediated environments.