Mapping the common based peer production: A crowd-sourcing experiment

Salcedo J , Fuster Morell, M , Berlinger M, Martínez, R and Tebbens W.

Abstract

Commons-based peer production (CBPP) is an emerging and innovative model of collaborative production. It usually takes place through a digital platform (Benkler 2006). It is characterized by peer to peer relationships, in contrast to the traditionally hierarchical command and contractual relationships, and with limited mercantile exchange. It results in the (generally) open access provision of commons resources (P2Pvalue, 2014). Some well-known examples are Free and Open Source Software (FLOSS) projects and Wikipedia. From those first generation of cases, there has recently been an expansion of CBPP to other areas of activity, such as citizen science, open product design, management of common spaces and open data sources. The paper explains the criteria to map this emerging model of collaborative production. The map of CBPP cases is based on web observation, web scripts, interviews to experts (to have an initial set of areas of activity of CBPP), a survey between CBPP cases to an ulterior classification and analysis of 302 cases. The result is the biggest database of CBPP cases, the data from the CBPP cases include area of activity, main purpose of the case, language, country, relationship with the digital environment (from digitally based to digitally supported), type of resulting resource, type of license and software and more of 150 variables. To map this diversity of cases is a big methodological challenge with some constrains such as the absence of previous CBPP database and other features of this phenomena that we explain in the following document.

Introduction

Relevance

At the end of 80's of XX century in an article called "Hybrid organizational arrangements: new form or transitional development?", Powell identified a third emerging form of the organisation of production, (Powell, 1987) distinct from the traditional two models considered by economic theory: firms and markets (Coase, 1937; Williamson, 1975). Since then, a rich literature on new forms of economic enterprise has developed, especially in the sectors more intensively dependent on knowledge, creativity, and innovation and those more impacted by the Internet and digital revolution (Castells et al., 2007) Various challenges have been highlighted in the flows of value production, consumption, circulation, and distribution that characterize these new forms of production, that often escape from the traditional systems of accounts and statistical surveys (Brynjolfsson and Saunders, 2009)

To frame these new unconventional forms of production, several proposals have been advanced with concepts such as P2P networks (Bauwens, 2005),cloud culture (Leadbeater, 2010) produsage (Bruns, 2008), free culture (Lessig, 2004), open culture (Stalder, 2004), wikinomics (Tapscott and Williams, 2008), open source production(Anthony, Smith and Williamson, 2007), recursive publics (Kelty, 2008), and online creation communities (Fuster Morell, 2012) Peer-to-peer relations are also one of the scenarios for the Future Internet envisaged by the Pashmina project (FP7)[1].

However, Yochai Benkler - partly relying on the work on the traditional commons developed by the 2009 Nobel Laureate Elinor Ostrom (1991)-in 2002 proposed and in 2006 systematised a new concept aimed at grasping an emerging and distinctive model of production: *Commons-based peer production* (CBPP)(Benkler, 2002, 2006). Benkler created the term CBPP to describe forms of production in which, with the aid of the Internet, the creative energy of a large number of people is coordinated into large, meaningful projects without relying on traditional

---

[1] http://www.pashmina-project.eu/ (20/08/2014)

hierarchical organisations or monetary exchanges and rewards (Benkler 2006).

In this work based on the review of previous characterisations of CBPP, through a questionnaire to the P2P value project members and experts[2], we have come up with a set of criteria of delimitation and categorization of CBPP. These criteria also define our unit of analysis across the research:

1. Collaborative production. CBPP involve some form of "collaboration" and some "production" – a process among peers that in their interaction form, develop, produce or build something valuable not present before their interaction. What results from this process might be very diverse.

2. Peer based: How individuals relate to each other and in a community. Community interaction is not solely or mainly coordinated by contractual relationships, mercantile exchange or hierarchical command. In contrast, individuals are in an autonomous condition and there is a decentralization in the conception and execution of problems and solutions.

3. Commons based: CBPP is not only characterized by being a peer process and productive (it is not only a peer-to-peer production), but also a commons process. Commons refers generally to that which is not driven primarily by restrictive/private appropriation but to a process that is driven by general interest. In the digital environment, this tends to take the form of an open access (with a license that assures the right to use [but not necessarily the right to make derivative works] and technically availability to use the resulting products).

4. Reproducibility and Derivativeness: Peers autonomy and commonness through reproducibility and derivativeness of the process and outcomes. This feature when applied to the digital environment is referred to as "forkability" (the license allows derivative work).

Previous research has provided a characterization of the organizational principles and features of CBPP. Most of the research on these conditions has been developed only for the case of *Free Libre Open Source Software* –FLOSS- (Crowston and

---

[2] To see the list of experts and the members of the project go to http://www.p2pvalue.eu/consortium/stakeholder-board (10/08/2014) and http://www.p2pvalue.eu/people (10/08/2014)

Howison, 2005; Lanzara, Morner and others, 2003; O'Mahony, 2007; Scaria, 2013; Weber, 2004). The empirical research was then expanded to the case of Wikipedia(Kittur et al., 2007; O'Neil, 2009; Reagle Jr, 2010; Viégas, Wattenberg and McKeon, 2007; Ortega, 2011) Hill (2012) provided a qualitative analysis of why Wikipedia was able to succeed in contrast to other apparently similar attempts to build an encyclopaedia. However, the research up to now has failed to take into account the diversity of types of CBPP, concentrating mainly on FLOSS and, later, on Wikipedia.

Preceding work has mainly focus on the functioning of single cases. When analysing the functioning of more than single cases, it has focused on considering very similar cases (such as contrasting the diverse linguistic Wikipedias (such as work of Ortega comparing the 10 most popular Wikipedias), comparing wikis at "wikifarms" (see work of Hill at Wikia) or FLOSS projects (see work of English. This paper is part of one of the few studies to compare diversity of CBPP cases. It is also the first research about CBPP, to our knowledge, to combine different methodological approaches. The challenge is how you map this emerging type of organizations and how do you sample and categorized these diversity of organizations.

The aim of this paper is to present the methods and criteria that we follow to create the first map of the CBPP universe. In this paper we explain how we construct in the best of our knowledge the largest database of CBPP experiences, also in this paper we present some of the main descriptive features that we identify about this type of organizations. This research project is under development and here we show a first lecture of the data, in that sense on further papers developed on the framework of the project we expect to present most detailed and explanatory findings about these emergent type of organizations.

This research is pioneering on mapping a large sample of diverse CBPP cases. We build a sample of a sufficient number of diverse examples (302) to support a statistical analysis of the success factors of productivity for value creation, and the analysis of the communalities of CBPP as a unit mode of production applied to very diverse set of areas of activities.

The paper has the following structure, firstly we expose the criteria and methods of sampling that we follow to create the database of CBPP experiences, secondly we show some preliminary descriptive data of the organizations that populate the database and finally we discuss about some features of the cases identified and the likely bias and limitations of our database

Research question
How do we mapping, an emerging and unknown phenomenon, like the CBPP?

Criteria and methods of sampling the CBPP universe

There are different techniques for sampling. Possibly, the most known is the probability sampling, which is based on criteria of randomness. Using a random or probability sample has several advantages. The most important benefit is the possibility to make inferences about the population with a certain degree of confidence. Randomization increases the likelihood that a large sample reflects the characteristics of the underlying population by avoiding assignment or selection based on the value of the variables of interest.

Nevertheless, is extremely difficult (if not impossible) to apply randomness to a population that is highly diverse and has unknown size and boundaries (Poteete and Ostrom, 2010). The diversity of cases, the lack of boundaries and the unknown size of the universe are some of the challenges that we have to face when we are mapping the CBPP phenomenon.

Additionally, randomization does not guarantee a representative sample per se, some scholars argue that random selection involves the risk of "missing relevant cases" (King, Kehoane & Verba, 1994, p. 124). The use of probability samples requires knowledge of the population—for instance, a list or census of the population or at least a partial list in some level of the population. This is not the case in CBPP, which is diverse and whose "universe" is unknown. Previous

attempts to map CBPP are limited.[3] In this sense, the P2Pvalue project's creation of a CBPP database is an innovative contribution. However, given the lack of reliable records of CBPP experiences, a probability sampling in the new CBPP field is impossible.

Our sampling aims to support an analysis that allows us to compare diverse formulas of CBPP (i.e., a comparability goal). Given the lack of conditions and unsuitability of developing a probability sample of diverse CBPP experiences, we used **non-proportional quota sampling**. This is a non-random type of sample mainly used when there is not a previous census or list of the population under observation, one of the main criteria of this type of sampling is to warranty the representation of the diversity of groups (quotas) that are part of a population, because it is not knowledgeable the exact proportion that each group has on the population, based on previous studies about the CBPP phenomenon and the knowledge of the researchers the number of cases to be included on the final sample are assigned with the aim to guarantee at least a minimum representation of each group.

Because this sampling aims to guarantee diversity, we expected that we would be able to talk about even small groups in the universe of CBPP. We ensured the inclusion of a mixed type of CBPP experiences to reflect the heterogeneity of CBPP. From a departing list of cases identified (around 350) on p2p value directory[4], we used different criteria of "matching" to ensure diversity in the sample. Additionally, to improve the robustness of our sample, we ensured the systematization of the sampling. That means that we documented the set of steps we followed in case collection and selection to facilitate the reproducibility of the sampling.

First, the strategy for case selection for the sample was to filter out all cases that

---

[3] Most CBPP research is based on only one type of CBPP or on single cases. Surveys of large N samples of FLOSS cases have been made. For example, Schweik and English did a survey of FLOSS, building its sample from Sourceforce (2012). Moreover, Wikipedia has been qualitatively compared with similar wiki projects (Hill, 2012), and Wikipedia's linguistic versions have been statistically compared (Ortega, 2009), as have wikis in Wikia's "wiki farms" (Hill, Shaw & Benkler, forthcoming). However, these comparisons considered only one particular type of CBPP. Regarding the universe of diverse CBPP, Benkler, Shaw, and Rossini mapped 150 cases (2010), and Fuster Morell (2010) statistically analysed 50 cases.

[4] http://www.directory.p2pvalue.eu/ (10/08/2015)

failed to match the **definition of CBPP** (our unit of analysis). This pertains to the fulfilment of the criteria of delimitation of CBPP that we defined and that refer to the presence of four features: collaborative production, peer relations, commons, and reproducibility. The case selection was also based on exclusion for **methodological constraints**.[5] But these problems were observed only in some cases of the previous directory list. We had to take off only 20 cases that fulfilled none of the criteria.

Second, a variety of CBPP was selected to ensure the **variability of the independent variables**:

i) Elements that generally describe the case: Year of foundation of the case and scope (local, national, and international)

ii) Elements connected to the type of collaborative production (independent variable): Area of activity, type of collaboration involved, and type of common resource

iii) Elements connected to governance (independent variable): Type of legal entity, type of license of the content generated by the user, and license of the software

iv)Element connected to sustainability (independent variable): Type of legal entity

We ensured **variation in the independent variables** without regard to the values of the dependent variables. We selected those elements for variability because they were the ones we had data on. As much as possible, we avoided a likely bias of selection (see King, Kehoane & Verba, 1994, p. 136). We ignored variability regarding the independent variable internal system of recognition and reward of contributions because of a lack of data on this variable at the start of the project. However, a subsequent analysis of the data also points to the variability of this variable.

---

[5] We excluded cases with no contact information (email or contact form) because this made it impossible to send them the survey. Another criterion of exclusion was lack of online activity. Because most of the indicators were based on online aspects, if online activity was minimal, we would not be able to develop the analysis. To ensure data availability, we also prioritised cases that were mainly digitally based (according to our classification, around 70% of the sample was composed of cases that were "digitally based" as opposed to cases that were only "digitally supported"). Yet another criterion of exclusion was linguistic: we excluded cases that used languages not understood by the team (who knew only English, Catalan, Italian, Spanish, French, Dutch, German, and Portuguese).

To ensure the **variability of the dependent variable** (King, Kehoane & Verba, 1994, p. 139) we included in our sample a diverse range of experiences, some of which are well known and important in terms of the different dimensions of value that we considered (e.g., reputation and social use value), but we also included many experiences that were almost unknown. Thus, we guaranteed the variability of our dependent variable.

Regarding the case collection we used an innovative process of mapping CBPP cases. We applied the logic of the collaborative open production of CBPP to map CBPP cases. We did so by building a directory of CBPP cases. The directory of CBPP is an open web resource that allows users to add CBPP cases collaboratively (see http://directory.p2pvalue.eu/). The directory provides several data from each case. We built an initial list of cases of around 125 (December 2013 / January 2014). This initial database was based on our own knowledge of cases as CBPP experts (IGOPnet team) and on some previous directories and mapping experiences of P2P projects and digital innovation projects—for instance, the P2P WikiSprint,[6] the portal OurProject.org,[7] the European project Digital Social Innovation,[8] and the host of open-source projects, LIBRE.[9] The directory was then opened to any online volunteer willing to populate the directory with more cases. In first stage the This strategy ensured the diversity of sources in the case collection because the experts that added cases had diverse backgrounds—from partners from five European countries (the UK, Spain, Italy, France, and the Netherlands) to members based in other countries (e.g., India, Ecuador, and the US). Additionally, to ensure diffusion and populate the directory, we made a "data jam" or hackathon to develop a crowdsourcing process to include new cases (12th March 2014, http://www.p2pvalue.eu/blog/p2pdatajam-review). However, although we allowed anyone on the Net to insert cases, engagement was limited. The majority of cases was entered by members of the P2Pvalue project.

After the directory was built and the data jam was held, with a departing list of around 350 cases, we analysed the cases from this process to extract our sample.

---

[6] http://wikisprint.p2pf.org/?lang=en (04/04/2014)

[7] https://ourproject.org/ (04/04/2014)

[8] http://digitalsocial.eu/ (04/04/2012)

[9] http://libreprojects.net (04/04/2012)

We checked the quality of the data (generally high) and (as previously pointed out) analysed the diversity of the cases in terms of the year of foundation of the case, scope (local, national, and international), area of activity, type of collaboration involved, type of common resource, type of legal entity, type of license of the content generated by the user, and license of the software. The objective was to create a balanced database in terms of the variability of our independent variables.

We supplied the missing data, conducted some basic "data cleaning," and defined a plan to complete the sample in a way that would increase its diversity, such as by increasing the number of cases in weak areas of CBPP (e.g., P2P funding). Finally we obtain a sample of 302 CBPP cases.

In this 302 cases we complemented the information obtained by the directory with a web observation or collection of data on the web pages of the different cases, in addition we run some web scripts to obtain some measure of web reputation and visibility of this cases on the web. (eg: Alexa Global Rank, Alexa inlinks Google Page rank and inlinks, Kred Influence, Twitter followers among others).

During the web collection, the estimated time dedicated to each case was between 40 minutes and two hours. To guarantee the reliability of our sample, another team member (who collected no data on experiences) was assigned exclusively to randomly testing almost 30% of all the cases and verifying the data of some outliers, the level of reliability was of the 90%. In this way we control the quality of our data. As for the data obtained through scripts, almost 15% was manually contrasted, with a similar level of reliability, the problem with this data is that has a high level of volatility depending on the moment of the measurement, the value changes.

Finally we send a survey among the cases to obtain these type of information that was almost impossible to obtain by the web analysis (web observation and scripts). For instance, questions like the size of the community (how many people participate and contribute actively), the community budget, the governance of the community (the presence of systems to resolve controversies, how the community take decisions) the number of people hired by the community, the percentage of

women that are parts of the community boards, etc. Also we included some questions about the relevance that the community gives to privacy issues and in this way implements policies to protect members' privacy. The survey also allowed us to triangulate some of the information obtained by the web analysis. For instance, if the community has a system to visualize the contributions of other members, or if it has some policy of rewards according to level of contributions of the different community members or if there are different roles within the community (e.g.administrators, moderators, simple users...).Nevertheless, as it was expected we only have a rate answer of 20% of the initial number of emails sent, we have an **n** of 67 communities of CBPP cases.

Some descriptive data about the sample.

In this section, we present some of the descriptive features of the CBPP database. United Kingdom (35,8%), followed by United States (30,2%) are the countries with more cases. Then, Spain, France, and Germany are the countries with more cases. Most of the cases might be position in Europe - and Spain - due to the European bias of the project, and research group.
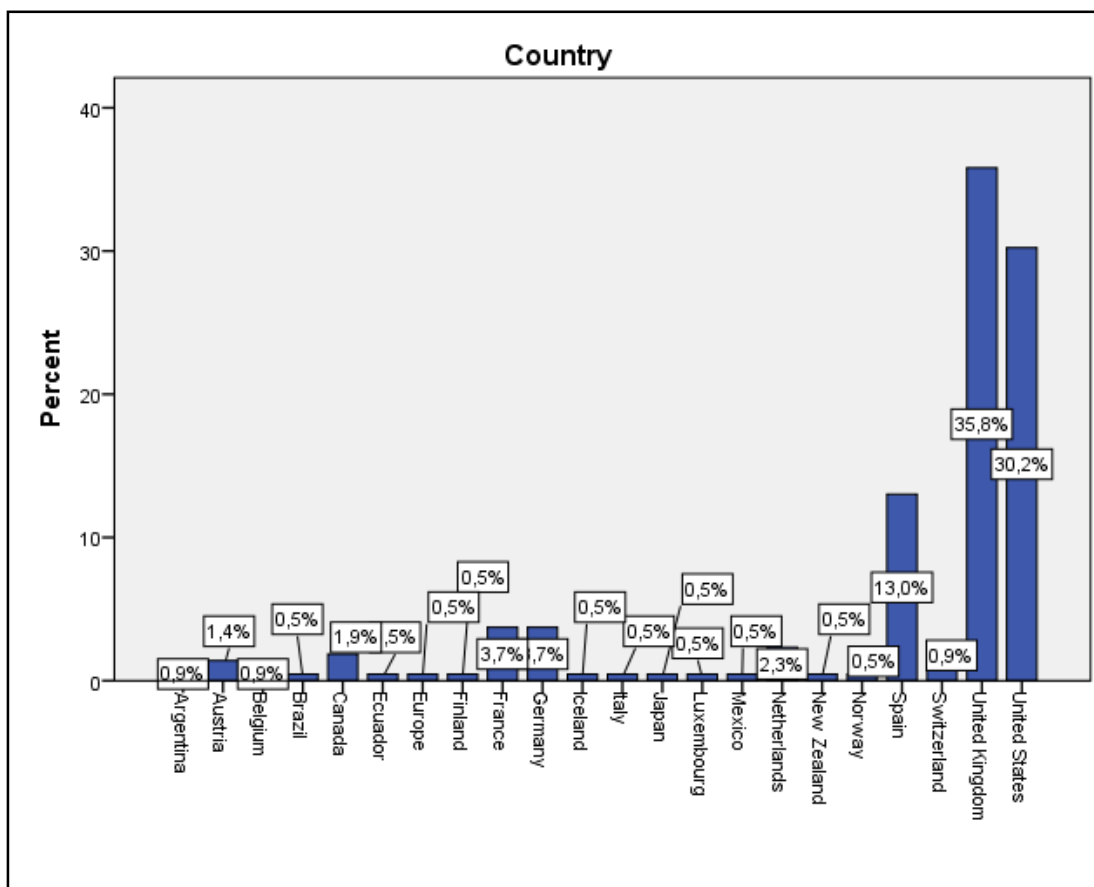

Figure 1, source: P2Pvalue project CBPP database.

The "older" case is from 1981. Actually, 1.7% of the sample started in the 80s. 9.5% in the 90s, 57.2% in the first decade of 2000, and 31.6% in 2010 to 2014. 1999-2000 seems to be the tipping point when CBPP started to growth and since then, it keeps growing. 2010 is the year in which more cases were founded (10% of the cases).
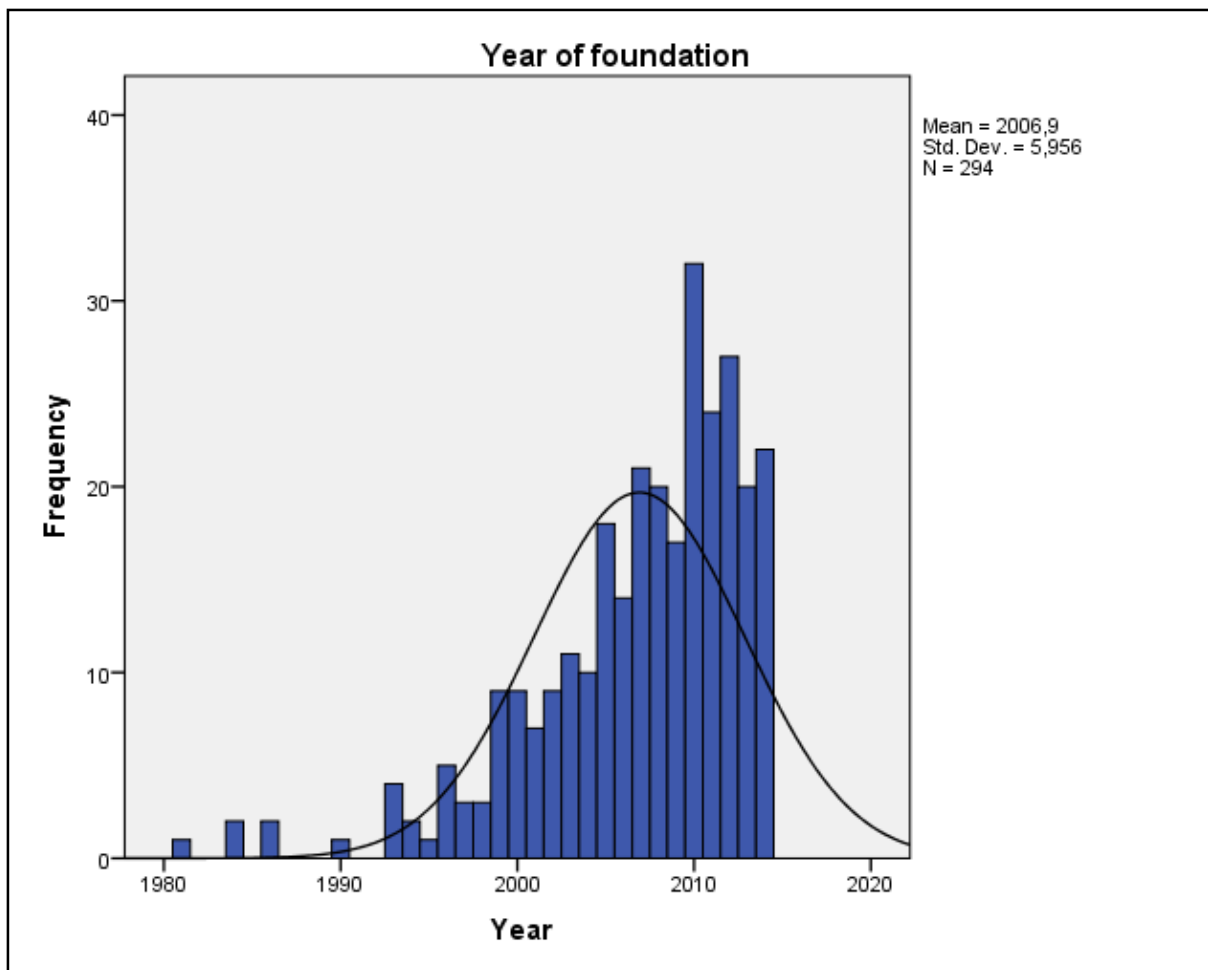
Figure 2, source: P2Pvalue project CBPP database

In relation to the area of activity, we can observe on the figure (3) that FLOSS (Free and Open source software projects) is the area of activity which involves more cases (29.4% of the cases in the sample). This is something expected with the fact that FLOSS was the first area of development of CBPP and it is the most developed. Additionally, the cases linked to areas of activities more connected to technology tend to be more frequent (open technology 8.7%). Locally oriented areas such as Community networks are 19.6% of the sample. Collaborative writing is also a popular area 11.9% (such as wiki communities) and collaborative research (8.9%) even if not being technically centred.
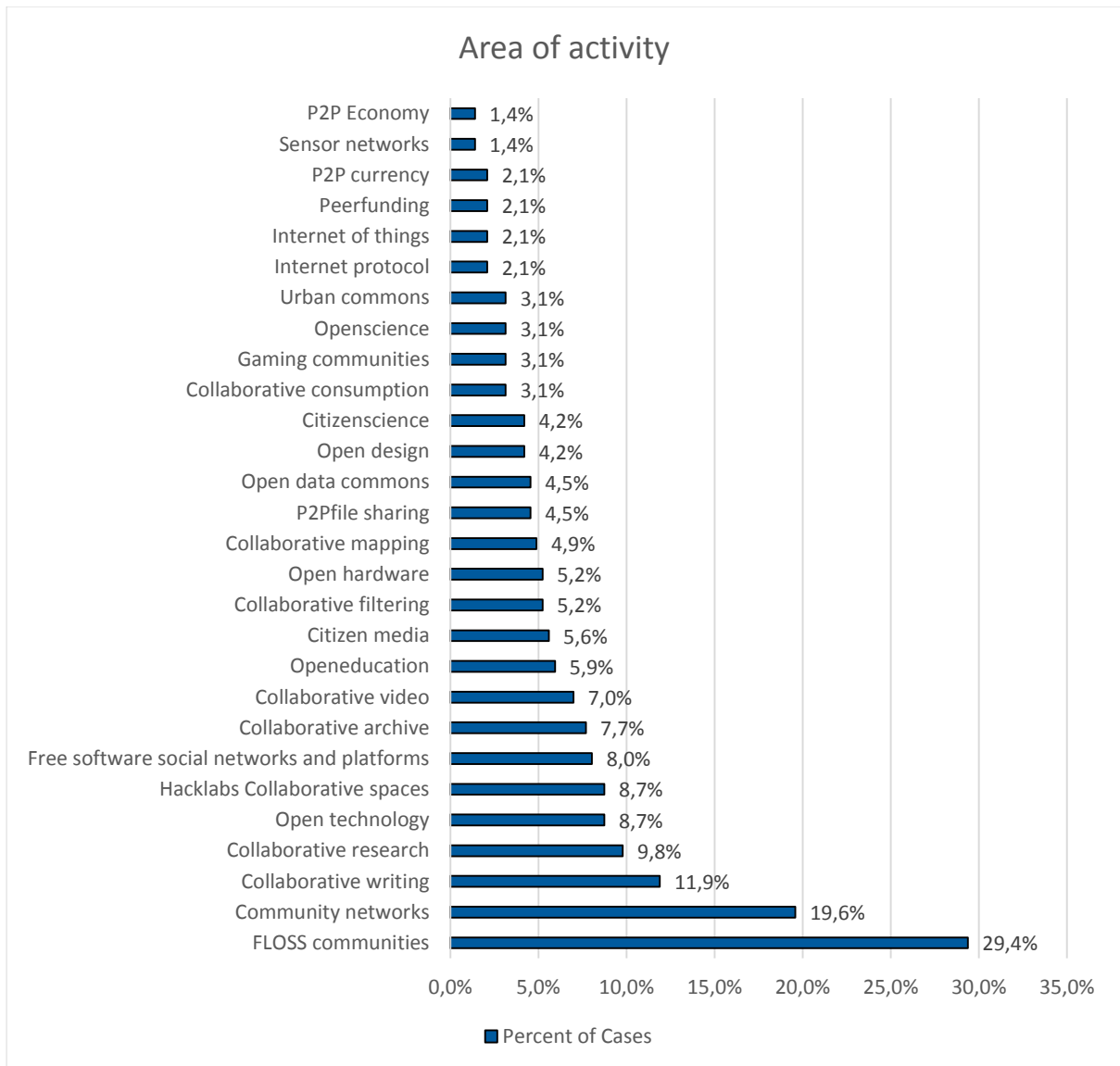
Figure 3, source: P2Pvalue project CBPP database.

On the database 73% of the cases are digitally based and 27% digitally supported. Digitally based means that the majority of CBPP mainly developed their activity on the Web, is the case of Wikipedia, Linux, Mozilla, etc…, digitally supported are the cases that use the Internet to promote offline services, it is the case of Hack labs, car, house or other services of sharing rival[10] goods.

---

[10] We use the categories of rival and non-rival goods use on political economy, the first one is a type of good that its consumption implies that other can consume or use at the same time the good (e.g.: a screwdriver), or clearly the use of the good have limits in terms of the number of people that could obtain benefit of it use. Also a rival good is called as subtractable good, the majority physical private goods are rival. A non-rival good or non-subtractable is a good that is marginal production is equal to zero, if most people decide to consume the good do not affect the individual level of consumption, even if more people use the good a network effect increase the value of the good it is the case of some services on the Web, for instance a software or a digital CBPP.

For what concerns the cases "digitally based", the most complex and integrated type of collaboration is classified as "Collage" (e.g. writing something together, as a dictionary or an encyclopaedia, or developing collaboratively a software package). And collage is the most frequent type of collaboration in the sample (45,7%). In terms of both complexity and frequency, it is followed by the type of collaboration we classified as "Album": that is putting together distinct and autonomous pieces (such as multimedia archives). This kind of collaboration concerns the 15,7% of cases in the sample. Finally, the least collaborative and the least frequent in the sample, is the collaboration based on "Exchange": it refers to spaces that allow an exchange between the individuals. It concerns the 11,7% of the cases.

For what concerns the cases "digitally supported", we distinguished the three types of collaboration: building a physical resource, sharing a space, and building of a collective process (such as, for example, a political movement). Each of these types of collaboration interests about the 9% of the sample.
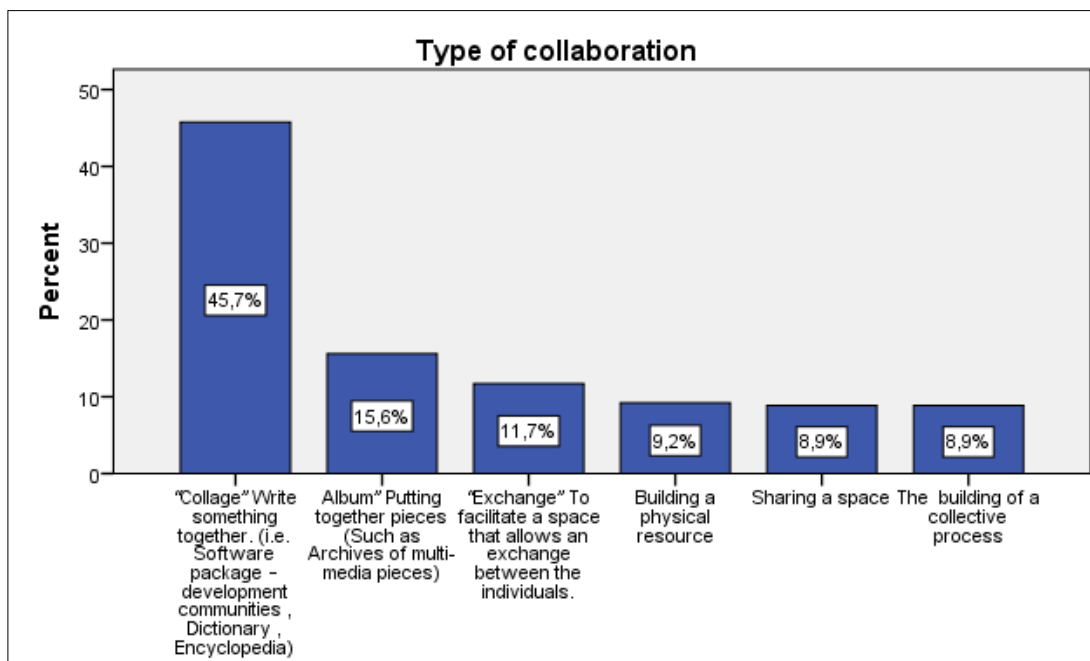

Figure 4, source: P2Pvalue project CBPP database.

According to the type platform that supports the case, we classified between platforms that are "one content centric platform" (that is supporting a specific content production, e.g. Wikipedia, GitHub) and platforms that are "multi content platform" (supporting multiple types of content, e.g. Google Drive, Kune…).

According to this categories, one content platforms characterize the 58% of the cases in contrast to 42% that rely on multi content platforms.(n=302)

In terms of the type of organization that supplies the infrastructure, 49,8% of the sample are Grass Roots Organization or Community Network as infrastructure provider. Then, 25,1% are businesses. 16,2% Academia and Research. 7,3% Social Enterprise Charity or Foundation. The less common are 1,6% Governmental and Public sector institutions.
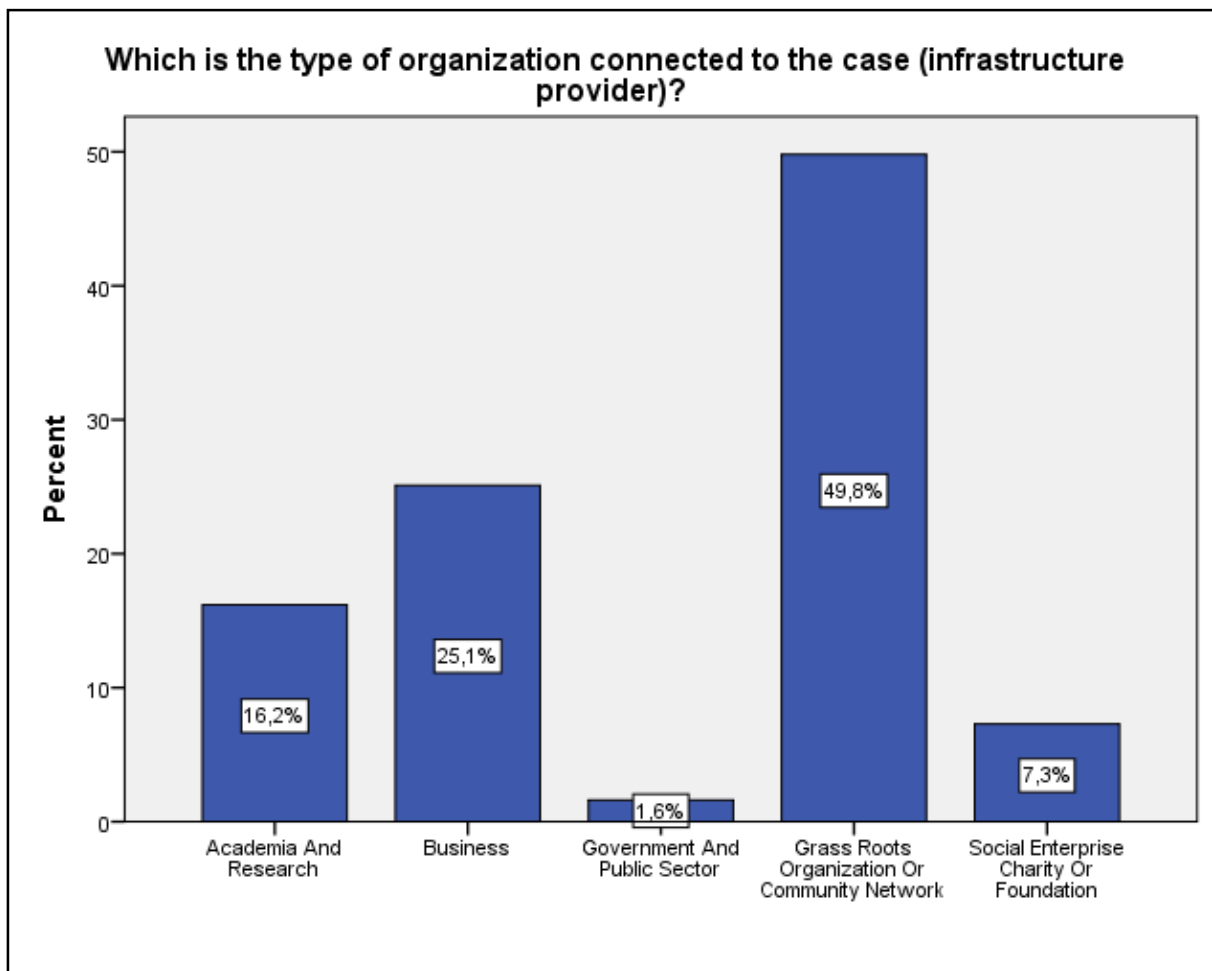

Figure 5, source: P2Pvalue project CBPP database.

About some governance issues of the CBPP cases, the mission of these organizations is a nodal point in defining the overall goal of the process. It contributes to give the process direction, and ultimately, define the process and its governance. We analyse of the most salient words in the description of the mission of the case (as available in the case website), the analysis shows

differences between the digitally based cases versus the digitally supported (see next word clouds). Digitally based cases use frequently words related to technology (the 5 most frequent terms: open, free, software, project and source). While digitally supported cases use words connected to "people" (5 more frequent terms: community, open, people, network and platform).

Figure 6, Word cloud digital based, source: P2Pvalue project CBPP database.

Figure 7, Word Cloud digital supported, source: P2Pvalue project CBPP database.

The following figure (7) presents data on the popularity of the indicators of

openness of the CBPP to people participation on the case development, we find that the functionalities linked to create¨ content are much more frequent than those linked to ¨communication¨ among participants. The possibility to add ”chips”/pieces of information (such as a new case in a directory) or actions (such as a new campaign in Goteo) is the most common form of participation (70% of the cases).
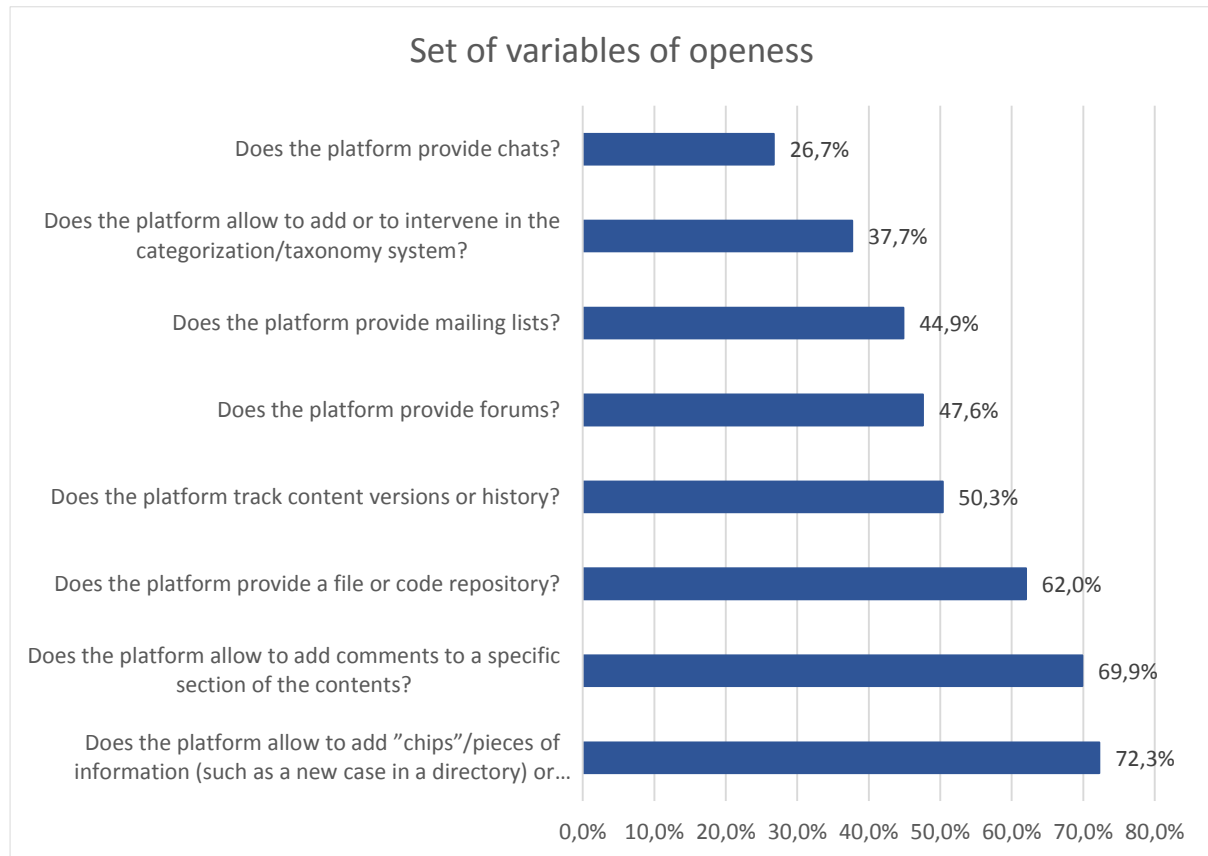
Set of variables of openess

Question	Value
Does the platform provide chats?	26,7%
Does the platform allow to add or to intervene in the categorization/taxonomy system?	37,7%
Does the platform provide mailing lists?	44,9%
Does the platform provide forums?	47,6%
Does the platform track content versions or history?	50,3%
Does the platform provide a file or code repository?	62,0%
Does the platform allow to add comments to a specific section of the contents?	69,9%
Does the platform allow to add ”chips”/pieces of information (such as a new case in a directory) or…	72,3%

Figure 8, source: P2Pvalue project CBPP database.

In terms of the size of the community of the cases included on the database, we can observe on the next figure the three indicators of community building that we used: How many people you estimate overall participate in the community? How many registered accounts are? (If applicable); and, How many people do you estimate that actively contribute to the community?.

Data suggest that the scale of the communities is extremely variable. There is not a very frequent range of number of people engaging or/and contributing. From 201 to 1000 (or more) is the most frequent range of people that overall participate in the community and of the number of registered accounts (but it is "only" around

20% of the cases for both indicators). In contrast, from 51 to 200 (or less) is the more frequent range (23%) of people that actively contribute to the community. It seems rational, and in line with the power laws dynamics, that the range of very active participants is lower than the regular participants.

Cases does not seem to be composed by very large communities. According to the two first indicators (people that participate and number of registered accounts) 50% (the median) of the cases are below the 1000 of participants, and the 60% of cases (cumulative percent) are below the 200 hundreds of people that participate actively.
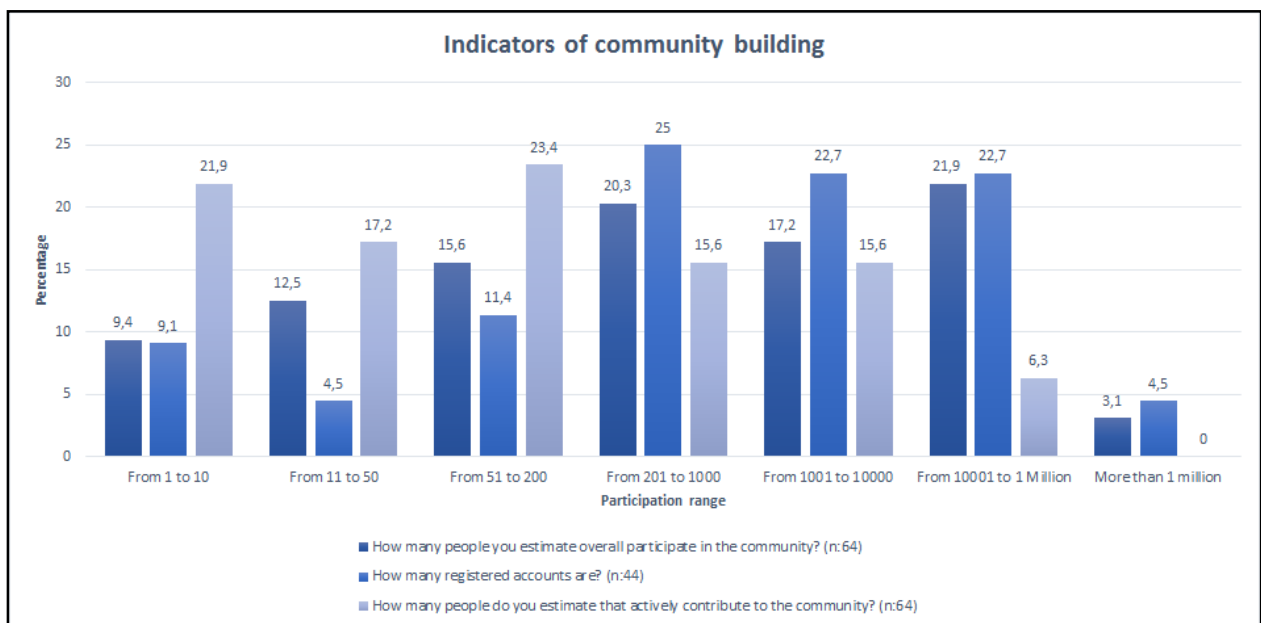

Figure 9, source: P2Pvalue project CBPP database.

In terms of the structure of the CBPP, ¨hierarchies¨ are very frequent. 88,8% of cases have **different types of account/role with diverse levels of permission**. The high presence of different types of accounts might be connected to the frequent role of administrators connected to the fact that in open online systems require often some level of "administration" to assure the operation of the platform and avoidance of undesired contributions (SPAM).

Regarding the license of the main digital commons resource. As the digital commons can be referring to software, hardware, content, networks or other

forms of commons resources, the licenses can be specific for those domains. One of the most frequent license is copyright all-rights-reserved. However, it "only" represents 18,4% of the cases. At first the 18,4% of cases presenting an all-rights-reserved license surprises as subject of being non-commons resource. This can be understood due to the wide inclusion criteria used to build the sample, to enable us to compare characteristics of cases that are near CBPP but aren't purely so. Or due to the fact that the common aspect of the case is connected to something else from which we collected the license.

Among the other licenses, the most frequent are CC BY-SA (18%), and General Public License (GPL) (18,9%). Then, BSD/MIT/Apache License, Lesser GNU Public License (11,9%), and CC BY (10,2%). Additionally, of all cases at least 48% includes a copy left (or "share alike") clause in the license, i.e. it requires users to continue applying the same license conditions down the stream.
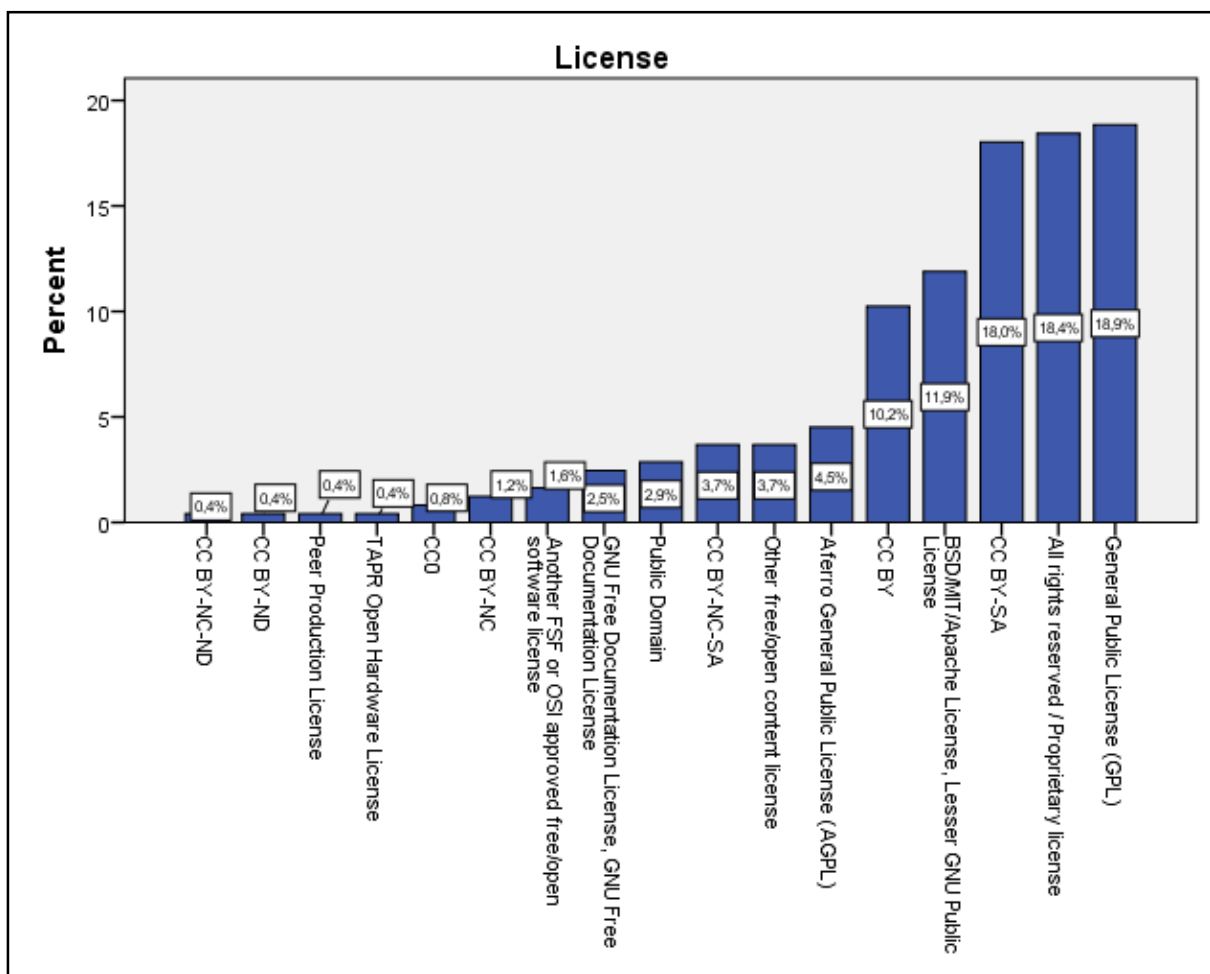

Figure 10,source: P2Pvalue project CBPP database.

When we analysed the different social media practices of the CBPP we identified that 15,2% of the cases does not use any social network. 18,2% use only one social network. 45% use two social networks. 15,9% use three. Only 5% use four, and 0.7% uses five (or more) social networks.
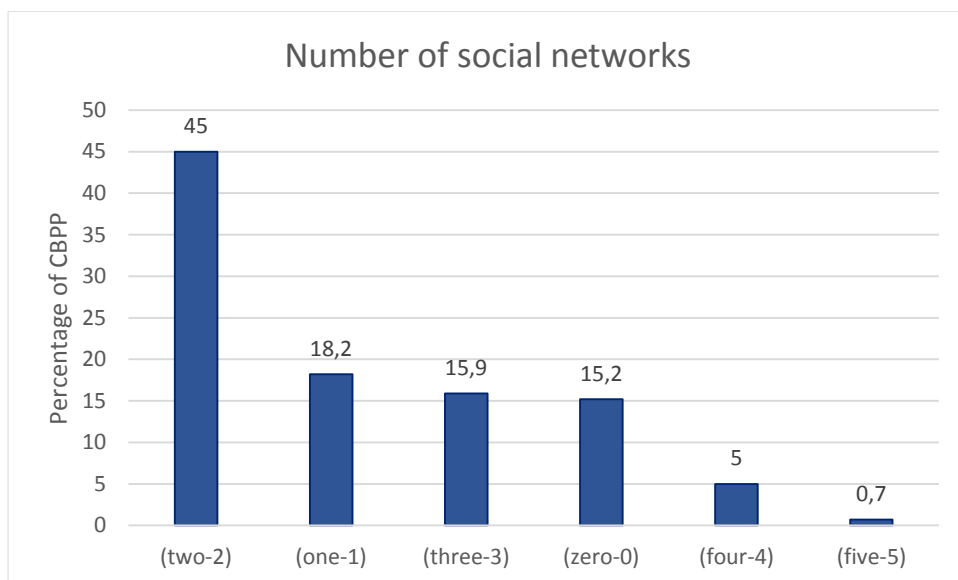

Figure 11, source: P2Pvalue project CBPP database.

Regarding the specific social network used. The most popular social network used by the CBPP is Twitter (89,8% of the cases) and, then, Facebook (76,6% of the cases). Other social networks are used, however they are much less popular. 18,4% use Github, 16,4% use Google plus, and 10,2% Youtube.
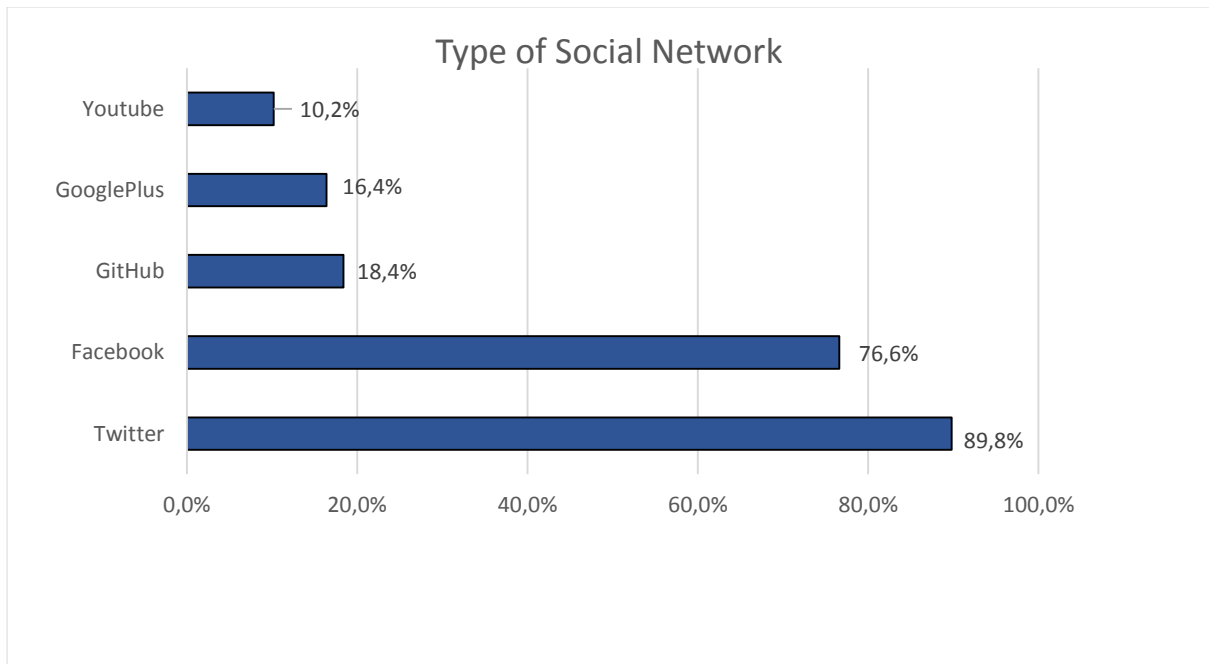
Figure 12, , source: P2Pvalue project CBPP database.

After this set of descriptive data about the CBPP sample, we present some considerations of this descriptive findings and about the development of this database, the possible bias, limits and the research ongoing to take advantage of all this information, we have identify more than 50.000 observations of the set of 302 experiences, in that sense there is a lot of work of systematization and analysis.

Discussion and possible source of bias, limits and dilemmas

Based on the previous data we consider crucial to highlight the strong emergence of CBPP that are more digitally supported than digitally based, in that sense these type of cases seem to be more focus on civic and grass roots organizations than on technology and software development.

About what is the most frequent institutional design in CBPP organizations, in terms of the licences the free and open licenses are predominant.

In spite of some ideological "contradictions" of using corporative social media platform to advertise the role of the CBPP organizations, the use of corporative

platforms such as Facebook, Twitter are very common by CBPP, it is expected that CBPP cases privilege the likely high network effect of a corporative media, in terms of the popularity of the platform and potential scope of their message. It is not so common a total transmedia practice in terms of use many type of platforms, the majority of CBPP use between one and two platforms.

About the potential bias of our sample, one possible source of bias is that the experts involved in the case identification are mainly related to the "first" generation of CBPP. That is, the well-known cases and "older" areas of CBPP, such as FLOSS and collaborative writing through wikis and the like, are the areas more widely studied. Another likely bias pertains to the country of the cases, given that a disproportionate number of cases came from Spain, where the research group is based.

We reduced this possible bias by privileging the criterion of heterogeneity to populate the database and thus have a representative sample of the diverse types of CBPP. We followed the different sampling criteria explained above. Different researchers from several countries and backgrounds were involved in different stages to control the type and quality of data. Finally, we did some intercoder tests to measure the reliability of the data.

The present research aims to extend the application of a CBPP framework of analysis to a larger range of experiences, compared with the most well-known and studied (such as FLOSS and Wikipedia)—including, among other new areas of application, experiences that are rather hybrids combining commercial aims and CBPP.

We adopted this strategy because we considered it more apt to give account of the most recent developments in CBPP, raise awareness of the increasing importance of such a model of production in information, knowledge, and the digital economy, and to address the research question whether the areas of applicability of CBPP could be considered a unified model of production.

However, to accomplish this objective, we had to refine our capacity of analysis to

recognize the layers of CBPP in different positions within the different cases—for example, with a more central or peripheral position in relation to the main objective of the project. To then focus on this selected layer in our analysis.

This highlights one of the main difficulties implied in the analysis of many cases of CBPP: the plurality of layers and dimensions that characterize this form of production and the following dilemma about selecting the main layer on which to focus the analysis.

This also applies to the outcomes. These projects often simultaneously produce a plurality of outcomes: various valuable contents (software, know-how, formal knowledge, data, etc.); valuable relationships; and valuable innovative rules, methodologies, norms, institutions, etc. And it is not always easy to select the relevant outcome on which to focus. Moreover, some outcomes can be produced indirectly and unintentionally—for example, by default or by a platform's architectural design.

Thus, in many cases, we simultaneously observe a plurality of license schemes. That is, each case can use a different license for each outcome (or even no license for some layers or outcomes).

This plurality can refer to a plurality of commons that are differently regulated within the cases. For example, some cases present a combination of software commons and commons of knowledge whereby members collaboratively construct a knowledge base related to the software. However, the software production process and license are generally different from the content production process.

However, while each dimension of this multiplicity of layers, outcomes, and licenses is important to understand the economy of each case and it can be hard to separate one from another, for methodological reasons, we focused on an individual outcome and layer to collect data. Clearly, the focus had to be directed to one common. However, which one was often neither easy to select nor immediately intuitively understandable from the point of view of an external observer. Thus, to reduce ambiguities, during the data collection, we introduced

the request to the data collectors to specify on which layer and outcome they were focusing the analysis and which license was applied to it.

Other challenge on the analysis was provided by the technical and organizational configuration of each project. In this sense, we could argue here along with Nobel Prize winner Eleanor Ostrom (1990): Each commons has its own features and uniqueness of configuration. More concretely, the challenge was posed by the multiplicity of platforms and online tools often used by each project and by the peculiar use of the different platforms by each case. For example, rather than relying on their own website, some projects are centred on the use of services provided by corporate social networks (such as Twitter, DataAnalysis15m, and GitHub for many FLOSS projects) or intensively use mailing lists, forums, and chat, which are, more often than not, hosted by external providers.

A subspecies of this dilemma (as well as of the previous) was also provided by cases in which the organizations have several subprojects (for example, the Open Source Robotics Foundation) that each have their own flexible and ad hoc techno-organizational configuration, including different partnerships. Finally, another subspecies category can arise in some cases that use multilingual platforms if they rely on multiple domains

We included in the database diverse cases, some of which are "pure CBPP" (closer to our departing definition of the unit of analysis though the criteria of delimitation of CBPP), while others are hybrids. To include hybrids in CBPP research—that is, projects that combine elements of CBPP with the commercial strategies of more conventional for-profit companies—is a consolidated approach, especially in FLOSS studies. This hybridism is possibly the main tendency characterizing the most recent developments in CBPP. However, we went further. We have intentionally included a little group of cases that are at the periphery of what is considered CBPP. The most extreme case is Skype, which has really few aspects of CBPP. It can be considered a CBPP, especially in the logic of the P2P protocol which its communication infrastructure relies on. However, the protocol is proprietary, as is the software (while the company owner is no less than Microsoft Inc.). Less extreme, albeit still at the periphery, are other projects run by for-profit

companies, like BlaBlaCar, Couchsurfing, Kickstarter, Eppela, Indiegogo, Gigatribe, BuddyBackup, Clickworkers, and Waze.

The reason for introducing these cases was twofold: to test the limits of a CBPP approach and to learn to visualize through multilayered analysis how these projects produce and rely on some form of CBPP—for example, a common or shared P2P co-produced system of evaluation, trust, reputation, information, etc. As already stated, such a multidimensional structure is crucial to understanding the functioning, success, and sustainability of many of these projects.

One last dilemma concerns with the informality that characterizes such forms of production. To a certain extent, the technological systems, communities, and resources of these projects are open, permeable, nested, and overlapped. That is, it can be sometimes difficult to mark the formal borders of the projects, to define the attributes of the systems, or to measure the outcomes of the communities.

One example among many others is the attempt we made to measure the "openness" of the project by looking at the possibilities of becoming a part of the **legal entity** presiding it. Beyond the difficulties of finding sometimes clearly specified rules or the diffused habit of allocating governance roles by merit to active contributors, many projects do not even have a clear legal entity. Moreover, in many projects, because the commons resource can be freely reproduced instead of the rights becoming a part of the legal entity, "**forkability**"—the right to fork and replicate that is in the hands of the community—plays the function of an ultimate tool that enforces commons governance. For example, most FLOSS projects are mainly organized by distributed version control systems (like "git"), which are exactly founded on a design that enormously simplifies the logic of forks.

Similarly, it could be argued in relation to projects that were relevant in the past but have been losing their momentum and are no longer that active (e.g., Morpheus). These cases provide difficult challenges at the time of collecting significant data, measures, and attributes, especially through actual web observation. Yet their census can be important to understanding the ecosystem of the CBPP because of the persistent value they disseminated that still constitutes a building block of more recent projects and because CBPP to a large extent relies

on, produces, and reuses goods and resources that are **non**-**rival** (which is also one reason why outcomes in the digital sphere are increasingly shared as commons, making their use value freely accessible and reusable, and is thus one reason for the increased informality typical in these forms of production). In general, these characteristics point towards the importance in CBPP of monitoring and evaluating what we call "**ecological value**" (the value that goes beyond the borders of the case by utilizing resources generated by it). This refers to flows, exchanges, and derivative works that cannot be easily located on specific platforms, community confines, and time frames and, as such, produce several challenges in terms of data collection.

As we mention, the database and its analysis has different constraints, nevertheless it is a first step to systematize and analyse the diversity of CBPP, in that sense we consider that is a great contribution that needs be considered and improved. Here we only present a small fraction of all the research project and only some descriptive data, on further papers and documents we expect to present more explanatory findings and reflections about this emerging phenomena, called CBPP.

References

Anthony, D., Smith, S. W. and Williamson, T. (2007) 'The quality of open source production: Zealots and good samaritans in the case of Wikipedia', *Rationality and Society*. Available at: http://www.cs.dartmouth.edu/~sws/pubs/TR2007-606.pdf (Accessed: 28 August 2014).

Bauwens, M. (2005) 'The political economy of peer production', *CTheory*, 1. Available at: http://www.informatik.uni-leipzig.de/~graebe/Texte/Bauwens-06.pdf (Accessed: 28 August 2014).

Benkler, Y. (2002) 'Coase's Penguin, or, Linux and" The Nature of the Firm"', *Yale Law Journal*, pp. 369–446.

Benkler, Y. (2006) *The wealth of networks: how social production transforms markets and freedom*. New Haven [Conn.]: Yale University Press.

Bruns, A. (2008) 'The future is user-led: The path towards widespread produsage', *Fibreculture journal*, (11). Available at: http://eprints.qut.edu.au/archive/00012902 (Accessed: 28 August 2014).

Brynjolfsson, E. and Saunders, A. (2009) *Wired for Innovation*. MIT Press. Available at: http://books.google.com/books?hl=es&lr=&id=WBYeChNzVo8C&oi=fnd&pg=PR5

&dq=Brynjolfsson+%26+Saunders&ots=72Vi-RBoZp&sig=_ZhBI3vJiEadrltRwR1gL-i4qIs (Accessed: 28 August 2014).

Castells, M., Tubella, I., Sancho, T. and Roca, M. (2007) *La Transición a La Sociedad Red*. Barcelona: Ariel.

Coase, R. H. (1937) 'The nature of the firm', *economica*, 4(16), pp. 386–405.

Crowston, K. and Howison, J. (2005) 'The social structure of free and open source software development (originally published in Volume 10, Number 2, February 2005)', *First Monday*. Available at: http://ptwich2.lib.uic.edu/ojs/index.php/fm/article/view/1478 (Accessed: 28 August 2014).

Hill, B. M. (2012) *Almost wikipedia: What eight early online collaborative encyclopedia projects reveal about the mechanisms of collective action*. Working Paper, Berkman Center.

Kelty, C. (2008) 'Two bits', *The Cultural Significance of Free Software. Duke*. Available at: http://twobits.net/pub/Kelty.pdf (Accessed: 28 August 2014).

Kittur, A., Suh, B., Pendleton, B. A. and Chi, E. H. (2007) 'He says, she says: conflict and coordination in Wikipedia', in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, pp. 453–462. Available at: http://dl.acm.org/citation.cfm?id=1240698 (Accessed: 28 August 2014).

Lanzara, G. F., Morner, M. and others (2003) 'The knowledge ecology of open-source software projects', in *19th EGOS Colloquium, Copenhagen*. Available at: http://ifiptc8.org/events/esrcseminars/LanzaraMorner.pdf (Accessed: 28 August 2014).

Leadbeater, C. (2010) *Cloud culture: promise and danger*. British Council: Counterpoint:< www. counterpoint-online. org/cloudculture-promise-and-danger.

Lessig, L. (2004) *Free culture: How big media uses technology and the law to lock down culture and control creativity*. Penguin.

Fuster Morell, M. (2012) 'Acción colectiva a través de redes online: Comunidades de Creación Online para la construcción de bienes públicos digitales', *Revista de Estudios para el Desarrollo Social de la Comunicación*, 1(6). Available at: http://www.revista-redes.com/index.php/revista-redes/article/viewFile/202/196 (Accessed: 5 February 2014).

Fuster Morell, M. (2010). Governance of online creation communities. Provision of platforms for participation for the building of digital commons (Unpublished dissertation), Department Social and Political Science, European University Institute, Florence

O'Mahony, S. (2007) 'The governance of open source initiatives: what does it mean to be community managed?', *Journal of Management & Governance*, 11(2), pp. 139–150.

O'Neil, M. (2009) 'Cyberchiefs. Autonomy and authority in online tribes'. Available at: http://works.bepress.com/mathieu_oneil/7/ (Accessed: 28 August 2014).

Ortega, F. (2011) 'Wikipedia: producción colectiva de conocimiento libre'. Available at: http://medialab-prado.es/mmedia/8/8602/8602.pdf (Accessed: 28 August 2014).

Ostrom, E. (1991) *Governing the commons: The evolution of institutions for collective action*. Cambridge Univ Pr.

Powell, W. W. (1987) 'Hybrid organizational arrangements: new form or transitional development?', *California management review*, 30(1).

Reagle Jr, J. M. (2010) '"Be Nice": Wikipedia norms for supportive communication', *New Review of Hypermedia and Multimedia*, 16(1-2), pp. 161–180.

Scaria, A. G. (2013) 'Schweik, Charles M. and Robert C. English (2012). Internet Success: A Study of Open-Source Software Commons. Cambridge (MA): The MIT Press', *International Journal of the Commons*, 7(2), pp. 577–578.

Stalder, F. (2004) 'Open Cultures and the Nature of networks', *Science, Technology and Society*, 22(2), pp. 165–182.

Tapscott, D. and Williams, A. D. (2008) *Wikinomics: How mass collaboration changes everything*. Penguin.

Viégas, F. B., Wattenberg, M. and McKeon, M. M. (2007) 'The hidden order of Wikipedia', in *Online communities and social computing*. Springer, pp. 445–454. Available at: http://link.springer.com/chapter/10.1007/978-3-540-73257-0_49 (Accessed: 28 August 2014).

Weber, S. (2004) *The Success of Open Source*. Cambridge, Mass: Harvard University Press.

Williamson, O. E. (1975) 'Markets and hierarchies', *New York*, pp. 26–30.