

The Individual and the Collective in Open Information Communities

Philippe Aigrain

Founder, Society for Public Information Spaces (in creation)

© Philippe Aigrain, 2003. Use of this text is governed by the Creative Commons Attribution-NonCommercial license (<http://creativecommons.org/licenses/by-nc/1.0/>). This text is the extended abstract of an invited talk at the 16th BLED Electronic Commerce Conference, 9-11 June 2003.

Summary: The text focuses on 2 critical issues for open information communities: how lowering transactions costs linked to becoming an active contributor is an essential factor for their success; and how free (as in freedom) licenses enable new forms of relationships between the individual and the collective. I bring some evidence that this permits to overcome some of the traditional limitations of commons in the physical world. The analysis uses as examples the free Wiki-based encyclopedia Wikipedia and the peer-reviewed free encyclopedia Nupedia, the Slashdot technical news community (commercial), and Web sites using the SPIP free co-operative publishing software. I am very much indebted to Clay Shirky for his article: Social Software and the Politics of Groups¹.

In last year's conference, I presented a typology of free / open source software collaborations. This year, I would like to focus on information communities, proposing a specific angle to discuss one topic of this panel: the individual and its relationship with the collective. I will present some general principles, and discuss how they apply to concrete examples. I hope that what I will say will have also some relevance for those who want to develop commercial services, so that they can do it successfully while at the same time respecting the conditions for success and for ethical integrity of information communities.

Introduction: why information commons are different

In the past there were many objects of societal, common property. But these common goods were most often physical resources: grazing land, wood for heating, water, air. Physical resources are what economist call rival. If one uses them, they are exhausted, or at least depleted, and another can not use them, or gets less value from them. In such a situation, if one wants to keep these resources as public goods, one has to regulate the access to them, through instruments such as control of access, regulation or taxation. Physical commons are often restricted in access to a local community, and when access pressure builds up, for instance because of newcomers, demographic growth, technical change or simple search for profit, it becomes difficult to sustain them without setting up strong management mechanisms and protective regulation. In contrast, most information goods are non-rival entities, that can be reproduced at extremely low marginal cost. Often they have even positive externalities of usage, that is their value grows when more people use them. Free riders are mortal enemies for physical commons, while they may be friendly allies for information commons², provided of course they do not try to appropriate common goods for their exclusive usage, through strategies such as embrace-and extend-and-appropriate. This situation makes possible to manage by different means two aspects that are generally treated as one in physical commons: in information commons, one can separate access rights from governance of commons. One can have universal access rights granted through free licenses, and nonetheless use a pyramidal governance structure – for instance - within a particular community. However, we will see that there are important contexts in which the value of a flat open structure proves superior.

1 http://shirky.com/writings/group_politics.html

2 When information commons use some scarce resources such as bandwidth, for instance in peer-to-peer networks, free riding can represent a problem.

Transaction costs and the world of information

In “Coase's Penguin, or Linux and the Nature of the Firm” [3], Yochai Benkler makes a remarkable case for commons-based peer production as a superior production model for information artefacts in comparison to markets and organisational hierarchies such as corporations. Here is a quote from his own summary of the paper, in which I have emphasized the last sentence:

“First, [commons-based peer production] is better at identifying and assigning human capital to information and cultural production processes. In this regard, peer-production has an advantage in what I call “information opportunity cost”. That is, it loses less information about who the best person for a given job might be than do either of the other two organizational modes. Second, there are substantial increasing returns to allow very larger clusters of potential contributors to interact with very large clusters of information resources in search of new projects and collaboration enterprises. *Removing property and contract as the organizing principles of collaboration substantially reduces transaction costs involved in allowing these large clusters of potential contributors to review and select which resources to work on, for which projects, and with which collaborators.*”

This finding has far-reaching implications for those who try to set up information communities. For instance, it means that imposing high transaction costs to accommodate the needs of a business model will simply ruin the benefits that motivate the user involvement. Many commentators⁴ miss the real point here. They describe the Internet as a culturally-biased world having developed a free-of-charge mentality, that should be redressed to make successful business models possible. However, it is not so much cost that is rejected that the transactions costs that go with it. And such rejection is not the result of immaturity but of a deep understanding. The fact that peer-to-peer file sharing users pay significant amounts for bandwidth and hardware, and even for contents on physical carriers, speaks for itself. To understand why transaction costs are so adverse to the creation and exchange of information and contents, one should note that transaction costs are much more than the monetary costs attached to the transaction⁵. They include cognitive costs (for instance the cost of deciding whether or not to do an action that may lead to a charge), time and information costs (for instance navigating in the transaction management layers), privacy costs, uncertainty costs (when some rights are subject to further approval), locking-in (i.e. loss of freedom due to the fact that an information service will give you access only to specific sources, or will make it difficult for you to switch to another provider). In this sense, the proprietary and control aspects are more rejected than the cost itself.

Of course, there are transaction costs that are necessary to useful functionality. For instance, in a co-operative news comment community, if one wants to implement efficient moderation and make it possible for users to exploit the fact that they trust some people more than others, one must keep track of who is who, have some memory of what comment each user has produced. This has some time and privacy transaction costs. We will see that a service such as Slashdot is very careful about keeping these costs as low as possible, and tries to put them as little as possible in the way of access and interest for the service.

Business models importing the high transaction costs of the physical world, or building new ones in networked information services without giving significant new capabilities to their users are bound to fail, even if they mimic some technical or organisation aspects of open information communities. They might succeed only for media that are centralised by nature, or in limited niches of corporate customers, for which transaction costs are largely hidden in already high overheads. How many failures will we need of Bertlesmann-reengineered Napsters, or of on-line content services restricting users capabilities and rights in comparison to what they can do with physical carriers, before the obvious is recognised?

3 Yochai Benkler, *Coase's Penguin, or Linux and the Nature of the Firm*, 112 Yale L.J. (Winter 2002-03), <http://www.benkler.org/CoasesPenguin.html>

4 See for instance: Pierre Grosjean, *Internet: La fin de la Culture de la Gratuité*, Kapital, 17 mai 2001, <http://www.largeur.com/expArt.asp?artID=742>

5 Clay Shirky, *The Case Against Micro-Payments*, <http://www.openp2p.com/pub/a/p2p/2000/12/19/micropayments.html>

Universality of rights, diversity of roles

When one observes information communities, even as simple as a topical email discussion list, one notices that their success often depends on some individuals playing specific roles. One will be a “referrer”, pointing people to information posted in other sources, or to recent news. Another will be a “communicator” able to translate in accessible terms complex scientific or technical documents. Another will be a “problem solver”, assisting others in overcoming some difficulty. Another will be a “moderator”, intervening to try to overcome conflict, to bring debate to a higher level of mutual understanding of what is at stake in diverging views. Yet another will be a “challenger” (of arguments, of explanations). These roles may or may not be formalised in the community processes (in present information communities, it is mostly the notation and moderation – in the sense of control on posting and its visibility - that are made explicit). Of course, most users / contributors will flexibly move from one activity to another: maybe pure consumers of information at most times, at other times answering a question because they happen to know the answer from past experience, occasionally triggered by some contribution to elaborate a much more complex argument, etc.

Recognising this diversity of roles, defining tools and rewards that accommodate or support each of them is key to successful information communities. And this can be done – in open information communities – without restricting rights.

The community that looked impossible: Wikipedia

It sure is weird that the Wikipedia works

*Clay Shirky, op. cit. linking to “Wikipedia: Our Replies to our Critics”,
http://www.wikipedia.org/wiki/Wikipedia%3AOur_Replies_to_Our_Critics*

Wikis are server side Web software (written in a variety of scripting languages) that enable users to edit a Web page (using simplified control on HTML underlying structure) as they browse it. One can find a general introduction to Wikis in [6]. Most Wiki software is GPL-ed free software. The original Wiki software was defined in 1995 within the design pattern software engineering community for its own needs. When I was still working in research funding in this field, I often joked that this was the ultimate proof of the value of research funding. Design patterns, despite their merits, still struggle to penetrate software engineering practice. Who would have guessed that the community working on this topic would in the end achieve what is a true revolution in the information society through a simple tool that they produced “on the side”?

Wikis can be used in a variety of ways: as co-operative work tools for a restricted group, for instance the editors of a Web site, or “in the open” by allowing anyone to edit contents.

Free encyclopedias have been for quite a few years one of the highest priorities in open contents. Several projects have emerged. It is worth studying the respective fate of 2 of them: Nupedia⁷ and Wikipedia⁸. Both projects are creating articles subject to the same license, the GNU Free Documentation License. Both are producing articles in the form of HTML pages. But their ways of working are radically different. Nupedia defines itself as a peer-reviewed encyclopedia, with a classical peer-review editorial process, and the related organisation of roles (editors, reviewers, copy editors, etc.). It also emphasizes titles of academic recognition as a guarantee of quality. Meanwhile, Wikipedia lets anyone free to contribute or modify articles using an adapted Wiki. When it started, most people, include myself thought that it would simply not work. We have all witnessed email lists or discussion fora being overwhelmed by flames and noise, how could one let anyone come and “destroy” a piece, include spurious articles, inject erroneous facts, libellous statements? We were just all wrong, because we had underestimated 2 factors.

6 Bo Leuf, Ward Cunningham, The Wiki Way : Collaboration and Sharing on the Internet, Pearson Education, April 2001, ISBN: 020171499X

7 www.nupedia.org

8 www.wikipedia.org

The first factor is that there is more to laws than police enforcement. Wikipedia has what Clay Shirky calls a constitution, a clear vision of what it is trying to achieve, a related code of conduct (see its statement on neutrality of point of view⁹), The second factor is that Wikipedia has mechanisms and software tools to guarantee that if enough people actually work according to its constitution, it will not be destroyed by a limited number of hostile or noisy contributions. The key mechanism in this case is an advanced versioning system, allowing to easily go back to an earlier version of an article. But there are many other small technical features that play an important role such as an easy way to create links to non-existent articles that will appear as “missing and demanded” when people browse. Some people have also argued that the simplest fact that it is so easy to break in a Wikipedia article and modify it removes much of the motivation for negative contributions.

Not only are there today (5 June 2003) 130000+ articles in the English version of Wikipedia, but there are Wikipedia at some degree of elaboration in 40 languages, including Slovenian¹⁰. And these language versions are original, not translations of each other. There is even a coordination project to link them one of with the others, without assuming that the decomposition of subjects will necessarily be identical in various language versions. When one picks a random article, or an article in a recently created language version of Wikipedia, one is likely to find an empty shell, or a ridiculously elementary article. If you just go away thinking that it is not worth losing your time, you get it all wrong. Many of these empty shells will turn in blossoming trees. Many of these elementary articles will become elaborated presentations of complex subjects. And quality is contagious in open information communities. When one starts seeing it in other contributions, it becomes much more rewarding to invest into making a good one also.

In “dense regions” Wikipedia articles tend to be shorter (than classical paper encyclopedia articles) and extensively linked one with each other. As a set, they look closer to Vannegar Bush's dream in “As you may think” that anything else that I have seen.

It is too early to say if and how Wikipedia will keep scaling up, how for instance it will be able to redesign itself when the evolutionary process produces sub-optimal structures. There are signs that it does scale up, but it could be that new instruments will become necessary to decide about certain major structure redesigns. It is also too early to say that Nupedia will not in the end reach critical mass. It sure builds up much slower, but many free software projects have also taken a long time to reach critical mass. And finally, Nupedia and Wikipedia do not think of each other as competitors: Nupedia links on its top page to Wikipedia as a “complementary encyclopedia project”.

Valuing some to provide value for all: Slashdot

Different constitutions encode different bargains. Slashdot's core principle, for example, is "No censorship"; anyone should be able to comment in any way on any article. Slashdot's constitution (though it is not called that) specifies only three mechanisms for handling the tension between individual freedom to post irrelevant or offensive material, and the group's desire to be able to find the interesting comments. The first is moderation, a way of convening a jury pool of members in good standing, whose function is to rank those posts by quality. The second is meta-moderation, a way of checking those moderators for bias, as a solution to the "Who will watch the watchers?" problem. And the third is karma, a way of defining who is a member in good standing. These three political concepts, lightweight as they are, allow Slashdot to grow without becoming unusable.

Clay Shirky, op. cit.

Slashdot is a specialised information technology (and related subjects) news and news comments community. Refer to Rol Malda (alias Cmdr Taco)'s “Inside Slashdot” article¹¹ for a technical historical presentation. Slashdot has hundred of thousands of registered users, and has become simply the primary general IT information source for the profession. It is also a profitable service, mostly through specialised

⁹ http://www.wikipedia.org/wiki/Neutral_point_of_view

¹⁰ <http://sl.wikipedia.com/>

¹¹ <http://www.symonds.net/~roshan/articles/Inside%20Slashdot%20-.html>

advertising (see below). Slashdot was founded in September 1997 by a bunch of individuals who still operate it now that it is owned by the Open Source Development Network, a subsidiary of VA Linux.

Slashdot mixes a centralised editorial structure with fully open contribution to news comments. Readers – the word is ill-chosen, since they provide most of the substance of Slashdot contents - can submit stories to editors, or editors just choose their own. The editors ensure that on each day, there is a good mix of subjects and stories (what they call the Omelette). A story is a very short news item (a few lines), generally linking to a one or more extensive or primary sources. The story is important in itself, but the “readers” provide key insight in their interpretation. The underlying software (called Slash and now also under the GPL license), provides what looks at first side as a classical threaded indented presentation, but in reality exploits the key feature in Slashdot: notation by moderators. Moderators are picked among registered users, by a relatively elaborated algorithm, monitored by a meta-moderation process to prevent possible biases.

When one is picked as a moderator, one can give or take away one point to five comments and help qualifying them by giving them one of a set of predefined attributes. Moderators are encouraged to reward quality rather than punishing its absence, but both contribute to the key effect: highlighting quality for the reader and making nuisance invisible. Slashdot is full of irrelevant or stupid comments, but one just does not see them, unless one specifically want to. There are some more personalised schemes such as the possibility for a reader to attribute additional value to points granted by people one esteems.

Overall, Slashdot achieves a delicate balance between avoiding barriers and unnecessary transactions costs, and associating the unavoidable ones with rewards. For instance, one does not need to be a registered user to post comments: they will then appear as coming from the slightly deprecatory “anonymous coward” identity. One can of course register under pseudos and multiple identities. But one need to be registered to moderate, and have a karma (a score trying to capture the value of a person to the community as a whole). It also combines centralised decision making on the top-level structure, with total freedom of contribution to detailed contents.

End-user development becomes real: SPIP

SPIP¹² is a GPL-ed free server side software (actually a set of PHP scripts interfacing with a MySQL database to generate dynamic HTML pages) enabling non-software specialists to quickly set, administer, customise and develop interactive Web sites. It was developed by 2 French individuals for the needs of a newszine they were editing, and as a tool for developing some sites in freelance service contracts. SPIP belongs to a line of tools such as PHPNuke, but it has made significant progress towards ease of use and customisability. It quickly spread principally in the French-speaking Web , but also in other languages. The number of sites under SPIP grew from 1500 to 2110 in the past 3 months in the French-speaking Web, while it grew from 380 to 700 in other languages (Arabic, Créole de la Réunion, Danish, German, English, Esperanto, Spanish, Gallician, Italian, and Vietnamese). SPIP operates large sites such as Le Monde Diplomatique or the site providing information on recruitment competitions for the French National and territorial civil service.

SPIP is primarily oriented towards co-operative publishing. It has made explicit (for the administrator of the site) a number of choices that we have discussed above regarding transaction costs and roles. It has 3 levels of roles: administrators (this role should probaly be split between technical and editorial administrators), authors and simple visitors. One can decide whether or not a visitor can become by simple registration an author, and there are many other facilities for tuning discussion fora, email lists, petitions, etc. It is a simple tool, that embodies much of the beauty of end-user led development enabled by free software, and is in my opinion very significant in terms of creating a continuum of positions between software programmers and pure “users”. It is not per chance that this occurred precisely for a piece of software devoted to information communities.

Which coupling between information commons and monetary-based economy?

One of the key questions regarding information commons is whether and how there can be a positive

¹² <http://www.uzine.net/rubrique91.html>

coupling with the monetary-based economy. Economy (in particular economy of physical goods and of services) seems to draw great benefits from the existence of information commons (education, access to information, externalising some costs), but these benefits are hard to measure. Due to the immensely reduced costs of information handling, only a limited share of direct economic value needs to be injected to sustain the information commons ecosystem. But how? This coupling can be either direct, through a business model, or indirect, if government or other social and economic actors fund the activity because they derive indirect benefits from it. The examples described above leave this question unanswered, but indicate some directions.

There are indications that indirect coupling is greatly preferable in the early stage of development of an information community, because it does not make the development of the community dependent on the ability to generate quickly a business model from it. Market fanatics would maybe find this a flaw, but there is in my opinion no indication that the direct possibility to generate quickly a business model is in any way linked to the future social usefulness and economic potential of a given form of information commons. Even more, the direct imposition of a business model at an early stage risks to significantly bias the information community organisation. When an information community (or for that purpose a software user community) has reached some critical mass, the ability to translate it into business opportunities ranging from services to support and documentation becomes much more open,

If an information community is centralised, it must, when it reaches a large scale, draw significant monetary resources, for instance to fund connectivity. Advertising is often used for this purpose, but this “solution” comes with extreme dangers. Slashdot is now profitable through advertising thanks to its position of unchallenged leader in its segment. It uses advertising in a moderate, low intrusivity mode, that is well tolerated by its users. However, there are severe macro-economic limitations of advertising, which never in 150 years represented more than 1.5% of GDP in any country. Slashdot is aware of these limitations, and has recently moved to offering subscriptions where one can get advertising-free access for a fee, using a business model similar to Opera's. Some fear that this initiate a process leading to more intrusive advertising on the free access service.

More generally, it seems unlikely that advertising can create a general positive coupling between free information and monetary-based economy. In our still early days of information commons, advertising represents a business model proxy for indirect coupling: the provider commercialises the attention that it was able to generate from the community. However, experience of past media – notably television – has shown that their development often remains locks in this initial business model, with devastating effects on quality and economic value. Will the much more active role of users / contributors be enough to counter-balance this tendency?