AI and Society, Open Forum, May 15, 2012.

Information overload and virtual institutions

Daniel Memmi

University of Quebec in Montreal, Canada
memmi.daniel@uqam.ca

Abstract

The internet puts at our disposal an unprecedented wealth of information. Unfortunately much of this information is unreliable and its very quantity exceeds our cognitive capacity. To deal with the resulting information overload requires knowledge evaluation procedures that have traditionally been performed by social institutions, such as the press or universities. But the internet has also given rise to a new type of social institution operating online, such as Wikipedia. We will analyze these virtual institutions to understand how they function, and to determine to what extent they can help manage the information overload. Their distributed and collaborative nature, their agility and low cost make them a very interesting social model, but also a rather fragile one. To be durable, virtual institutions probably need strong rules and norms, as well as an appropriate social framework.

Keywords

Information overload – Knowledge management - Social Institutions – Virtual institutions – Wikipedia

1 Introduction

It is by now evident that the internet has ushered in a new era, offering a previously unimaginable wealth of information. Electronic networks coupled with various mobile devices mean that anybody, anywhere, anytime can access practically any information quickly and at little cost (given the right infrastructure, which is usually available in developed countries). This is of course an extremely beneficial development, for individuals as well as for society as a whole. Because so much information has already been digitized, we have at our fingertips most of human knowledge, useful daily information and unlimited human contacts.

But this abundance of riches comes together with an associated curse. It looks as if we had made a pact with the devil, gaining information wealth at the price of cognitive overload. The quantity and variety of information greatly exceeds our cognitive capacity for processing it. The issue of cognitive overload is not totally new (Wilson 1996; Bawden et al. 1999; Edmunds & Morris 2000; Eppler & Mengis 2004) but electronic networks have made it even worse. We are simply not built to process so much information in limited time, because our cognitive equipment (both biological and cultural) was developed long ago in simpler times.

For example if one looks for information on a disease, or in order to solve some technical problem, there are often thousands of web sites potentially relevant to the question. But how can one find and read so many web pages? And how can one be sure of the pertinence and reliability

of each document? In case serious decisions have to be made (e.g. the appropriate cure for a disease), this turns out to be a vexing problem of very real importance.

Of course, mankind has always looked for information and has had to ascertain the reliability of any information that could be found. Various methods have been devised in history to manage human knowledge, ranging from priestly castes to public libraries. And material techniques such as writing and printing have proven to be very important informational tools, with momentous social consequences (McLuhan 1962; Eisenstein 1983). But our main point here will be that knowledge management is ultimately performed by social institutions, because no technique as yet can reliably evaluate documents without human supervision.

But if the problem is not really new, its present extent is unprecedented: we have never had such easy access to so much information in human history. For this information to be profitable, one should be able to retrieve, filter, evaluate and explain relevant knowledge from among a flood of dubious information. We will argue that classical institutions must learn to cope with such abundance, but we will also describe how novel virtual institutions are emerging on the internet so as to deal with electronic information.

In this text, we will first analyze the problem of information overload caused by the internet, and envision possible means to deal with this problem. We will see how information has been managed historically by a variety of social institutions, and we will describe the diverse functions they fulfill. We will then show that new virtual institutions better adapted to the nature of electronic information are emerging on the internet, taking advantage of its speed and easy use, and we will discuss the characteristics of such institutions. This will allow us to suggest future directions.

2 Information overload

The internet has made available most of human culture literally at our fingertips. The architecture of the internet with its public, open protocols spans the whole globe and gives access to practically any computer anywhere (Tanenbaum 2003). The virtual structure of the web, built on top of the internet, includes millions of pages and documents touching on every aspect of human culture. E-mail can be used to communicate potentially with millions of people, and social media can track thousands of acquaintances at the same time (to cite only the most frequently used applications).

The very numbers of potential connections on the internet reveal at once a huge problem of information overload: how can we deal with thousands of documents on a given subject? How could we manage thousands of relationships? The answer is that we simply cannot, because we are just not built to perform such enormous work. Psychologists have shown that we can only deal with half a dozen chunks of information at the same time (Miller 1956), and most people cannot keep track of more than two or three hundred casual relationships. Performance can be improved with training and experience, but this takes time and effort. When information supply exceeds individual processing capacity, the overload leads to stress and poor decisions (Eppler & Mengis 2004). Information that might be relevant or even necessary gets ignored.

Not only should we be able to take in and understand so many sources of information so as to locate relevant knowledge, but it is also be crucial to evaluate what we have found, because bad information is worse than none. To be useful, knowledge must be understood and evaluated, a task that requires complex and slow cognitive processing. This cannot be done by an individual for more than a dozen documents at a time (say, within one workday). One may speed up with

training and experience, but the discrepancy between human cognition and the amount of information remains unmanageable and unbridgeable.

Before giving up on the problem, however, it would be appropriate to analyze more closely what might be needed to address the issue of cognitive overload. We will see that social institutions have traditionally found ways to manage information to some extent. More precisely, we need the following capabilities to deal with the present information flood:

- information retrieval

ability to find information pertinent to a given subject or theme

- information filtering

picking relevant information from a continuous flow of news

- information evaluation

appraising the truthfulness and reliability of documents and sources

- contextualization and explanation

putting information in context and explaining its significance

To give a practical example, an engineer working on a particular technical problem wants to be able to find or select information relevant to the problem from a range of possible sources. He must then make sure that the information is reliable, and last but not least contextualize it: is this an isolated research result, potentially interesting but not yet widely tested, or a well-established practice in the field? What are the costs and potential risks? A physician looking for the best treatment for an unfamiliar disease would have to go through a similar procedure.

By the way, this might the place to make a distinction between *information* and *knowledge*, a distinction we have so far glossed over. Information could be defined as mere data, unevaluated and out of context, whereas knowledge is contextualized, verified, organized and systematic. Other (and finer) distinctions could be made: one might describe a gradual process turning raw data into selected information, information into verified knowledge, and knowledge into plans of action. But we should distinguish at least between unprocessed information, and knowledge that has been cognitively integrated. What is really needed is reliable knowledge, not mere information.

Now, what resources are presently available to give us the necessary capabilities to deal with the information flood? For retrieval and filtering, there are technical tools in common usage (although their overall performance is limited). But evaluation and explanation are mostly beyond the reach of today's techniques, and are still done by humans. This situation might point to promising research directions, but the state of the art is quite restricted.

Information retrieval is the most advanced yet of the tools we need. This is a mature discipline (Baeza-Yates & Ribeiro-Neto 1999) and search engines have become indispensable. They allow untrained users to retrieve pertinent documents with just a few keywords. Still, the overall performance of search engines is not that good: in spite of constant improvements, silence (documents not found) and noise (irrelevant documents returned) remain very high (results can be roughly estimated to reach only about 50% of the desirable performance). Moreover, results depend on the right choice of keywords.

Information filtering is also well understood technically, using techniques similar to information retrieval (Adomavicius & Tuzhilin 2005). But the results are highly dependent on a proper characterization of the relevant subject with a few appropriate keywords or tags. To go beyond keywords would be computationally expensive and is still a research domain. But if one knows exactly what one is looking for, or from where, syndication threads and now social media work quite well, thus contributing to the information flow…

Information evaluation, however, is basically outside the reach of technology. Google's ranking algorithm uses the Web structure to recommend the best-cited sites (Brin & Page 1998), but this a clever measure of social reputation, not a real evaluation of the content of documents. There are surface indicators as to the reliability of a text: a well-structured document with specific subtitles and a bibliography is more trustworthy than a rambling text without any references. But a proper evaluation would require an understanding of the content of the text, and this remains too difficult and too expensive to perform automatically.

Contextualization and explanation are also far from being even considered as problems to address by computing techniques. Putting information in context requires encyclopedic knowledge, culture and judgment. Finding the right frame for a given piece of information is a open-ended problem, often depending on prior experience (context can be widened and generalized). Explanation is similarly open-ended: deeper and deeper explanations can be developed with some effort by changing the frame of reference. These cognitive skills present interesting research themes, but there are no corresponding techniques for the time being.

These considerations could be taken as arguments to justify the Semantic Web domain: using logical formalisms (such as RDF or OWL) to annotate web pages and documents allows automatic processing of document content, opening the door to deeper reasoning about texts (Antoniou & van Harmelen 2008). This is indeed a promising research field, but results are still tentative. Reasoning on semantic content is limited to logical annotations and falls short of real evaluation needs. And there remains the basic problem of annotating documents in the first place: this must either be done by competent humans, or be derived from complex (and uncertain) text analysis techniques.

In short, we do not yet know how to evaluate information and knowledge automatically so as to deal with the information flood on the internet. This task must still be done by humans, often working in groups. In fact, information evaluation is usually performed by organized groups, i.e. social institutions, rather than by isolated individuals. Evaluating information presupposes a recourse to some kind of norm, and discussing norms and their application is very much a social concern. Individuals working by themselves would not be very credible without the endorsement of some social organization. We will now analyze in more detail the roles played by social institutions, in particular when dealing with information and knowledge.

3 Social institutions

Social institutions are social organizations: groups of people working together to perform specific tasks and functions, according to explicit or implicit rules, procedures and norms (Berger & Luckmann 1966; Mintzberg 1979). Social institutions are durable, and they represent more than the collection of people they consist of: institutions are physically embodied (in people and buildings) but they also have an abstract structure, as every member could be replaced and the institution would still continue to exist.

Social institutions can be explicitly created by decree, but they usually emerge more or less spontaneously throughout history, as people tend to work together repeatedly to accomplish tasks of common interest. With the passage of time, procedures and rules take shape and institutions become established (Greif 2006). For example, informal open-air markets become regular trade-fairs, associations of scholars turn into scientific academies, private tutoring gives rise to public schooling… And particular institutions are often supported and regulated by local or national administrations, which are themselves large-scale social institutions.

There exist social institutions specifically devoted to dealing with information and knowledge: notably the mass media, publishing houses, universities and research centers. These institutions perform crucial functions: they help filter and evaluate information, they continuously rephrase and adapt knowledge to changing social needs, but they also play an important role in the production of knowledge in the first place.

Of course there are important differences between various informational institutions. Some are private (the written press, publishers), some are strongly regulated (radio and television), some are public or heavily subsidized (universities and research centers). Some are more superficial with a short reaction time (the mass media: daily press and television), others take more time to produce more thoughtful comments (universities and publishers). But together, they perform an extremely valuable social task by managing collective information and knowledge.

Let us examine for example the various roles played by traditional publishing houses and we will see that these roles are more complex than meets the eye. The stereotypical image of publishers is that of firms waiting for unsolicited manuscripts to arrive and choosing the best ones to publish. In fact, publishers often participate actively in the production of written material from the very beginning. They usually have a wide knowledge of literary or technical circles, and often keep authors among their staff. They encourage or elicit specific themes, and approach potential authors for special subjects. Writers are often subsidized with cash advances or a regular stipend, and then prodded to finish the work on time. Literary agents act as go-betweens to negotiate contracts and smooth relations.

But the task of good publishers doesn't stop there. They are partners in the writing process as well: publishing houses employ specialized editors (often writers themselves) who criticize texts, request rewrites, offer suggestions, edit promising work… Only at the end is a decision made to accept or reject a text, but rejection is then much less likely because of the co-operative nature of the whole process. In short, publishing is in fact a highly social, collaborative activity.

In a similar way, universities and research centers constantly produce and criticize knowledge by the joint labor of thousands of specialists in any given field. Pertinent research themes are selected and quality levels are enforced by peer pressure, research grants, and the filter of peer-reviewed journals. A competitive system of symbolic rewards and penalties determines the professional reputation of individual researchers, thus globally ensuring the remarkably efficient production, filtering and evaluation of reliable, verified scientific knowledge. Teaching then explains and disseminates the results to a wider public.

The academic model of knowledge production is particularly interesting because of its collaborative yet anarchic nature, resulting from a long history (Rüegg 1992). Although universities are highly subsidized and professors are employees, research is basically free in two senses: research subjects are not imposed by a hierarchy, and there is no direct financial incentive for doing good research. By a mixture of collaboration and competition with the looming sanction of peer-review, high-quality knowledge is "freely" produced, evaluated and published.

This is a very effective model, and we should already make a note of its collaborative, decentralized, voluntary aspects.

By being so closely involved in the production of knowledge, traditional institutions solve most of the problems associated with information overload. Because only socially relevant and verified knowledge is published, retrieval and filtering are much easier, and evaluation in context is part and parcel of the knowledge production process. Of course, the system is far from perfect. One may wonder about the number of futile books published every year, and there is clearly a surfeit of academic publications due more to career considerations than to intellectual importance. But on the whole, traditional institutions deliver fairly useful, reliable and verifiable knowledge.

We may now witness the emergence of social institutions online, which exhibit both traditional and more innovative aspects. In the same way that information processing in general has become more and more virtual (switching from paper to computers), social institutions dealing with information may themselves become virtual. Let's take the example of Wikipedia, as particularly successful and representative. We will describe it in some detail, and then compare it with other similar institutions to highlight common features.

4 The example of Wikipedia

Wikipedia is an online encyclopedia based on a very particular model (started in 2001). Articles (encyclopedic entries) are written and edited collaboratively by a large number of anonymous contributors working without any financial reward. A wiki architecture makes it easy to edit articles, and many entries are repeatedly modified till they settle with time.

A wiki is a collaborative technology (launched in 1995) allowing a group of online users to write a common document. It's basically a web site with facilities to write, enlarge and edit a shared text. Users may have to register beforehand, but this not always the case. To correct possible mistakes, the history of recent changes is kept so as to be able to restore previous versions if necessary.

The Wikipedia model has been a huge success: at this time (end 2011), about 3,816,000 articles have been written in English, 1,324,000 in German, 1,181,000 in French, and hundreds of thousands each in the world's main languages (Spanish, Polish, Italian, Russian, Japanese, Portuguese, Chinese…). Altogether, there are more than 82,000 contributors in the world, but we will now deal mostly with the English version as the most extensive so far.

In order to offer unbiased, reliable and verified knowledge, Wikipedia has gradually developed evaluation procedures, a social structure and a set of norms appropriate to enforcing the quality of articles. Because of the need to teach new contributors how to write and edit good articles (and to solve potential disputes), these rules and norms are clearly formulated and freely available from Wikipedia's web site (www.wikipedia.org). But we will also profit from our own personal experience as a long-standing and frequent user of this encyclopedia.

The collaborative rewriting, sheer number of contributors and the continuous update of entries ensures that incomplete, biased, poorly written or unverifiable information is usually corrected within days or weeks, and blatant vandalism is often rectified within hours. Various mechanisms have been devised such as tagging problematic entries or sections of text to elicit further corrections, and notice boards to report vandalism and discuss editing problems. In the course of time, entries will thus improve by and large and eventually stabilize.

Some articles change a lot with time, and frankly, later versions may sometimes be deemed inferior to earlier drafts. But as the history of changes is kept (often with corresponding discussions), no version is final and it is always possible to improve on the current version and to retrieve deleted passages if needed. On the whole, most entries evolve toward stable versions of very decent quality (although the writing style may be uneven).

Moreover, individual articles have become more organized and contextualized in the past few years. The web structure makes it easy to link entries with other entries in Wikipedia (or to corresponding entries in another language) as well as to outside web pages. Subjects and themes have been increasingly placed within structured tables, hierarchical categories and generally situated within wider domains. Portals organize knowledge about a general subject and serve as an entry point to more specific articles. Articles are now rated for quality as well, problematic entries are tagged and outstanding articles are flagged as such, giving the reader a good idea of the reliability of the information offered.

But there is also a noticeable social structure to the mass of contributors. They are organized in a hierarchy of participants, with increasing responsibilities and rights: editors, confirmed editors, administrators and bureaucrats (to quote the titles used in the community). Newcomers do not have to register, but influence and editing privileges are dependent on peer-review. Positions are mostly gained through experience, subject to community approval. An arbitration committee has the last say in case of protracted disputes that could not be resolved otherwise. This is probably unavoidable, but a far cry from the mythology of totally free-wheeling online communities.

Rules and norms have been made more and more explicit, even though they remain rather informal for the sake of flexibility (Wikipedia's "five pillars"). Writing principles and dispute resolution procedures are proposed to help editors write and rewrite articles, and to solve conflicts about content. Neutrality, objective argumentation, consensus-building, and mediation if necessary, are put forward as recommended behavior among contributors. Dispute resolution advice in particular is detailed, thoughtful and of general interest (this looks like the outcome of long experience!). For instance, consensus is strongly preferred to simple majority rule.

Wikipedia's collaborative online model also exhibits features unusual among encyclopedias. Contrary to a paper encyclopedia, there is practically no space limit to the number of articles, and coverage includes popular culture with detailed entries on pop music and movie stars. Whether this is an interesting feature is for the reader to decide, but it certainly makes for a large diversity of content. Continuous editing means that recent events are treated quickly and new knowledge is promptly incorporated when it becomes available. If necessary, articles are rewritten in the light of more recent knowledge. Yet it seems that, for English at least, the growth of Wikipedia is now slowing down, probably because most obvious subjects have already been covered.

The overall result is a very useful online encyclopedia. The range of available information is enormous, knowledge is up-to-date, article quality is usually quite good and sometimes remarkable, and the hyperlink structure allows easy navigation between entries and subjects. Some caution should be exercised with more recent articles, but in our experience, they tend to be incomplete or badly written rather than grossly mistaken. And Wikipedia is absolutely free if you have access to the internet.

In short, Wikipedia is a largely self-organized online community that has evolved in the span of ten years into a fully-fledged social institution, with its tools and procedures, social structure, rules and norms and even the informal equivalent of a legal system. We have been witnessing the emergence of an efficient institution able to offer good-quality, reliable and useful knowledge at

little cost. Because Wikipedia is an online institution, it exhibits specific novel characteristics that we believe to be of general interest.

A comparison could be made with the free software movement, which exhibits similar characteristics, so as to highlight the points common to virtual institutions. So-called free and open-source software is also produced in a very interesting manner, thanks to interactions on the internet (Feller et al. 2005).

Free software projects are self-organized and carried out by contributors working without pay (though they usually have a regular job in a related field). The software produced is generally free and publicly available. Work is done collaboratively by proposing suggestions, additions, criticism and rewrites on the internet. Because the code is reviewed and rewritten repeatedly by numerous contributors, the end-results are often of high quality, and very useful software has been produced in this way (such as the programs running on internet routers). And the free production model is a good answer to the perennial problem of software piracy.

Compared with Wikipedia, free software projects are closer to classical work teams: contributors are more specialized (writing good software requires highly technical expertise) and there is a stronger social hierarchy at play (software engineering requires organization). Projects are implicitly organized in concentric circles, with an inner core of expert developers, and wider circles of contributors, critics and end-users. But the large number of participants ensures systematic and rapid debugging, as well as an ongoing evaluation of the software's usefulness and ease of use.

Here again a distributed organization and free contributions bring good results. These models could also be seen as examples of a more general movement toward the collaborative production of content by end-users (Surowieki 2004; O'Reilly 2005). We should now try to see what various online institutions have in common.

5 Virtual institutions

The internet has thus given rise to a new kind of social institutions which could be called virtual institutions. In the same way as classical social institutions emerge from social interactions, virtual institutions emerge from previous virtual communities (Kollock & Smith 1999; Memmi 2006). But virtual institutions show innovative features. The speed and low cost of interactions online have made possible new models of social organization that have proven to be quite effective in performing important tasks. There is by now a variety of such institutions, with various goals and structures, but they exhibit common characteristics:

- virtual organization

the organization is essentially disembodied, without a physical base: there are usually no specific buildings, no face-to-face meetings, no traveling necessary…

- highly distributed

work is distributed among thousands of participants, who could be located anywhere in the world (as long as they have internet access).

- massive collaboration

tasks are performed in collaboration, and an individual task may involve tens or hundreds of contributors, who are largely self-organized to get the work done.

- agility and speed

reaction times can be very short on the internet, which makes distributed, tentative, incremental piecemeal work feasible, quick and reasonably efficient.

- work without pay

participants contribute time, effort and expertise without any financial reward. Their motivations are mostly social (working for the common good) and symbolic (acquiring prestige and expertise).

- low cost

the absence of buildings to maintain or salaries to pay, the low cost of electronic interactions make those institutions cheaper and easier to operate than classical institutions (though one might argue that the real costs are borne elsewhere, e.g. by internet providers).

- informal hierarchy and rules

virtual institutions are real institutions with social structure and norms, but structures and rules are kept rather informal, with flexibility for adaptation and negotiation.

- decentralized decision-making

decisions are made locally as much as possible, with a view to consensus-building. Higher echelons intervene only in case of unresolved disputes and in order to maintain overall project coherence.

To sum up, virtual institutions are cheap, agile and quick, thanks to their distributed and collaborative operation. They seem particularly appropriate to deal with information overload on the internet. But their virtual nature might also render them rather fragile if there is no basic consensus on common goals and norms, and if clear rules are not articulated somehow. Last but not least, for these institutions to prove durable, participants must have sufficient time and resources to keep contributing through time.

Of course, there is a lot of variation among virtual institutions. The above features are not always present together to the same degree. For instance, free software projects may involve fewer participants, and regular contributors are often employed in fact by classical institutions (universities and technology companies). By comparison, contributors to Wikipedia are much more numerous (for the encyclopedia as a whole, though not for individual articles), usually less specialized, and contributing to Wikipedia is not part of their regular job.

And no virtual institution can be entirely cost-free. There are inevitable management costs that must be paid for in one way or another. Wikipedia employs minimal staff directly (about one hundred) but must still maintain hundreds of servers. Hence the regular funding drives asking for donations to cover the encyclopedia's operating expense.

6 Are virtual institutions really useful?

We have seen that virtual institutions exhibit some very interesting qualities, but also some potential flaws. To determine whether these institutions provide a workable solution to the

problem of information overload, we have to evaluate their strengths and weaknesses more closely. To recapitulate what we have seen so far:

- advantages

Virtual institutions are very cheap to operate (compared with traditional brick-and-mortar organizations) and they are highly reactive. Their distributed character makes them fairly reliable even with amateur contributors, because mistakes can be spotted and corrected quickly. Their informal social structure and flexible norms allow for more adaptability than rule-bound bureaucracies, and they can evolve naturally without too much pain.

In short, their light footprint makes them very agile, a potentially useful quality in order to deal with the constant information flow we are faced with in modern life.

- drawbacks

The other side of the coin is that lightly-built structures may prove fragile in the long term. Virtual institutions are still too new to be sure that they will pass the test of time. Organizations depending on voluntary contributions are notoriously uncertain of their future without constant efforts to maintain their structure and resources. This may lead to a waste of resources when too much energy is sunk into the structure itself, and much work must be repeated over and over.

Therefore, it is not clear whether virtual institutions will be a long-term solution, and they may have to evolve in order to last.

In brief, it looks like virtual institutions might not be so much better than ordinary organizations after all. They experience similar difficulties in maintaining their stability in the course of time, and to do so, they need more or less the same ingredients: strong norms, clear rules, a stable social structure, conflict-resolution procedures, and sufficient resources.

Without good rules and an efficient organization, the quality of a collaborative model breaks down quickly. Voluntary contributions without adequate supervision may degenerate easily into a free-for-all of knee-jerk comments devoid of serious argumentation, which do not build up to a common discussion. For example, comments posted by readers on news sites are often inane, and messages sent on social media are usually futile. Being virtual does not ensure any particular quality of discourse, unless the discussion itself is well organized.

One may also wonder about the long-term effects of replacing more and more social activity by virtual transactions. As more time is devoted to impersonal relations, could ordinary social bonds be threatened by a growing scarcity of face-to-face contacts? But this phenomenon can be seen as a mere continuation of the general trend toward more abstract and functional relations that has been operating in modern society for a couple of centuries. As long as we have a real, non-virtual life before we turn to more impersonal virtual activities, we may hope to acquire the social skills necessary for both ordinary and virtual communities.

And traditional institutions are not exempt form serious problems either. They are not always durable, strife-free or effective. Some endure throughout the centuries, though usually not without violent schisms, some collapse quickly and suddenly, many plod on without performing any really useful service. Churches, empires, academies, and firms grow and decay, many disappear for ever. This has been the subject of much historical debate, and virtual institutions will now take their place alongside older organizations to face similar long-term problems.

Yet the low operating costs of virtual institutions make them more resistant to market fluctuations. A private firm that suffers a decrease in its market share or its profits will bleed

money all the while, because its fixed costs (wages notably) cannot be adjusted easily. Virtual organizations are much leaner, and hence less vulnerable in this respect.

It remains to be seen whether virtual institutions can last in their present form and retain the same agility in the long run. It is quite possible that some of them will become more rigid and adopt stronger rules because they could not survive otherwise. Time will tell. It should also be remarked that virtual institutions emerged not only to perform specific tasks, but simply because they had become technically possible. Whether they render a useful service is moot in this respect, and these organizations will last mostly if they can muster sufficient resources to maintain their overall structure.

This being said, virtual organizations represent at this time a remarkable alternative to classical institutions, an alternative that shows great promise to manage the informational flood more rapidly, at little cost and with more flexibility. But just like classical institutions, they may need deliberate reinforcement after emerging more or less spontaneously.

7 Supporting virtual institutions

If they are socially useful (as we believe them to be), virtual organizations should be encouraged and helped, so far as is socially possible. We will discuss supporting mechanisms that have already appeared, but also suggest possible developments. Assistance may come from a whole range of different measures:

- socially: the corresponding social roles should be acceptable and valued
- economically: voluntary contributors must be subsidized if possible
- technically: software tools facilitating interactions must be developed
- legally: the status of end products should be clarified and protected

In fact, there is already a strong interplay between virtual institutions and classical ones (notably with academia) that could be encouraged even more. In the case of free software for example, much of the work is contributed either by academia or by private firms. Researchers, software engineers and programmers donate a portion of their time to free software projects, usually with the assent of their employers. It is important to understand how such models work, so that they could be sustained and fostered.

Socially, voluntary contributions to virtual organizations should be socially regarded in the same way as contributions to charitable and non-profit organizations. For contributions to be rewarding, some social prestige must be attached to such an activity, beyond the personal pleasure of doing something interesting and useful. Fortunately, this seems to be case (writing free software is a respected activity among software engineers) but more could be done to popularize volunteer work. For instance, it could be officially recognized and sanctioned as part of job descriptions in academia and engineering.

Economically, volunteers must of course support themselves somehow. The free software model is particularly interesting to analyze. Contributors to free software often have a regular occupation, and devote a variable portion of their working time to non-proprietary software with the approval of their organization. The motivations of individual contributors (beside social

convictions) are probably to acquire reputation and expertise that will be useful to their career. The motivations of private firms are less obvious (free software cannot be sold), but have been strong enough so far to sustain the free production model. The technology companies that accept to subsidize free software (not all them do) gain expertise and prestige in strategic domains, and may recoup their costs by selling services (advice, documentation, maintenance…) associated with such software.

The free software model would deserve a more detailed discussion which is beyond the scope of this article. There are deeper economic reasons for the success of this model. Software is a type of informational goods, which cannot be strictly controlled since they are so easy to copy, and which it is not socially necessary to control since the cost of reproduction is practically nil (Foray 2004). In other words, software is comparable to public goods, which are notoriously difficult to furnish by using market mechanisms alone (investments are hard to recoup, and there are no social reasons to restrict usage). Free production of software is a reasonable answer to these problems, and a very good example of a viable non-commercial model. Still, this model might also be officially encouraged, e.g. with tax deductions.

Technically, virtual institutions have been made possible by the advent of the internet and easy-to-use collaborative tools such as e-mail, wikis, and mobile devices. The simplicity, speed and low cost of electronic interactions explain why virtual institutions have become socially and economically feasible. Yet communication tools have been developed more or less haphazardly, and they are not always compatible or coherent with one another. We are convinced that work remains to be done by realizing that interactions take place more and more not just between individuals, but also within structured organizations as such. It should be clear that communication is primarily a social task.

Legally, there is already some framework in place. It was recognized early on that free software, and later public documents such as Wikipedia had to be protected against misuse, and notably against commercial appropriation that would make the free model unsustainable. In 1989 was devised the GPL license for free software produced in the GNU project. There are now a variety of such licenses, but GPL remains the most widely used (e.g. for Linux) and a paradigmatic example of public licenses. Wikipedia used the GNU Free Documentation License (an adaptation of GPL for textual documents) till 2009, and has then switched to a Creative Commons license.

All those licenses possess very similar characteristics. They affirm the public character of the products they cover: software, texts, images, etc, can be freely copied and distributed. Works or portions thereof may be reused, added to or modified at will, provided their origin is publicly acknowledged and the derived products are themselves covered by the same license ("viral" nature). This does not usually exclude commercial use, but the material must remain publicly available. Still, the legal status of these licenses is not totally clear so far (though they have been derived from copyright laws). Very few court cases have explicitly confirmed the full validity of such licenses, but courts have usually refused to invalidate them either. It might be time to give them a more assured legal footing to dispel any remaining uncertainty.

In short, there are already various social mechanisms to support and assist virtual institutions. They have developed by and large to answer specific needs, without overall reflection and planning. We would like to suggest that virtual institutions are an important new social development that should be recognized as such and be provided with a systematic social and legal framework. In the course of history, it happened repeatedly that social developments were enabled at critical points by crucial changes. For example, limited-liability companies, patent and

copyright laws, public education and research have been powerful tools for economic development. In the same way, it is now time to support virtual institutions explicitly.

8 Conclusion

The internet has brought us quick and easy access to a wealth of information. Unfortunately, individuals cannot process so much information in limited time. Information must therefore be evaluated and knowledge must be verified before use. This has traditionally been performed by social institutions (such as academia), that filter and evaluate the information available. These institutions must now adapt to the information overload caused by electronic networks.

But the internet has also given rise to new virtual institutions (such as Wikipedia) that operate without a physical base by performing most of their interactions on the internet. These institutions seem better qualified to deal with information overload because of their speed and low cost. Virtual institutions use a distributed, collaborative and voluntary model that has proven to be very flexible and socially useful, and that should now be encouraged as far as possible.

However, the highly decentralized and informal character of virtual institutions might also make them rather fragile, and they have not passed the test of time yet. Our analysis strongly suggests that they need clear internal rules and norms in order to survive, as well as a supportive social environment. Such internal and external structures are being slowly put in place, but the issue should be explicitly recognized and systematically addressed, so as to offer virtual institutions the best possible environment in which to prosper and endure.

References

Adomavicius G. & Tuzhilin A. (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, *IEEE Transactions on Knowledge and Data Engineering* 17 (6): 734-749.

Antoniou G. & van Harmelen F. (2008) *A Semantic Web Primer*, MIT Press, Cambridge MA.

Baeza-Yates R. & Ribeiro-Neto B. (1999) *Modern Information Retrieval*, Addison Wesley, Boston MA.

Bawden D., Holtham C. & Courtney N.C. (1999) Perspectives on information overload, *ASLIB Proceedings* 51 (8): 249-255.

Berger P.L. & Luckmann T. (1966) *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*, Anchor Books, New York.

Brin S. & Page L. (1998) *The anatomy of a large-scale hypertextual Web search engine*, Computer Science Dept., Stanford University.

Edmunds A. & Morris A. (2000) The problem of information overload: a review on the literature, *International Journal of Information Management* 20, 17-28.

Eisenstein E. (1983) *The Printing Revolution in Early Modern Europe*, Cambridge University Press, Cambridge.

Eppler M.J. & Mengis J. (2004) The concept of information overload: a review of literature from organization science, accounting, marketing, MIS and related disciplines, *The Information Society* 20 (5): 325-344.

Feller J., Fitzgerald B., Hissam S.A. & Lakhani K.R. (eds) (2005) *Perspectives on Free and Open Source Software*, MIT Press, Cambridge MA.

Foray D. (2004) *The Economics of Knowledge*, MIT Press, Cambridge MA.

Greif A. (2006) *Institutions and the Path to the Modern Economy*, Cambridge University Press, Cambridge.
Kollock P. & Smith M. (eds) (1999) *Communities in Cyberspace*, Routledge Press, London.
McLuhan M. (1962) *The Gutenberg Galaxy*, University of Toronto Press, Toronto.
Memmi D. (2006) The nature of virtual communities, *AI and Society* 20 (3): 288-300.
Miller G.A. (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information, *Psychological Review* 63: 81-97.
Mintzberg H. (1979) *The Structuring of Organizations*, Prentice Hall, New Jersey.
O'Reilly T. (2005) What is Web 2.0. Design patterns and business models for the next generation of software, Retrieved from: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.
Rüegg W. (ed) (1992) *A History of the University in Europe*, Cambridge University Press, Cambridge.
Surowiecki J. (2004) *The Wisdom of Crowds*, Anchor Books, New York.
Tanenbaum A. (2003) *Computer Networks*, Prentice Hall, New Jersey.
Wilson P. (1996) Interdisciplinary research and information overload, *Library Trends* 45 (2): 192-203.