Date : 17/10/2006

Open access, institutional repositories and the digitization of cultures.

Jean-Claude Guédon
Université de Montréal (Canada)

Meeting:	**155 Information Technology with National Libraries with Academic and Research Libraries and Knowledge Management**
Simultaneous Interpretation:	Yes

WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL
20-24 August 2006, Seoul, Korea
http://www.ifla.org/IV/ifla72/index.htm

Abstract

This paper deals with open access, but it attempts to move beyond the usual issues presently surrounding this particular debate. Its main point, stimulated by older insights coming from J. C. R. Licklider and more recent ones by Clifford Lynch, is that present conceptions of open access are not sufficient. Present discussions around OA tend to assume that the forms of reading associated with printed documents will remain essentially untouched; they also tend to neglect the possibilities of "open computing" despite the fact that Google and other similar developments point us in that direction.

This means that, in their design, repositories must incorporate the possibilities of algorithmic processing, most of which do not yet exist, as well as the possibilities of strong interactions with various kinds of researcher networks.

I. Introduction

Open access, most people agree, could never have been imagined without electronic publishing and the existence of networks such as the Internet. The reason is well known and regularly repeated: the presence of both a digital environment and a global network allows the possibility of duplicating and sending documents at a marginal cost that approximates zero. In the digital environment, we can calculate the full cost of publishing on the basis of putting the first copy on a server. Duplication and transmission costs are insignificant.

Saying that open access and institutional repositories depend on both digitization and networks is not very original. What is more important is the question of the role open access and repositories should play within a fully digital culture. In other words, given that we now clearly see the rush toward digitization of all things cultural, one may begin to wonder what is really needed to bring about “the optimal” form of digital culture.

Digitization in itself does not ensure any particularly good or bad outcome; it simply opens up a large range of new possibilities. Which one of these will become tomorrow's reality? No one knows yet, but many possible digital worlds can emerge, some much better than others. Open access does promise some improvements in the way scientific and scholarly literatures are being used to create new knowledge, or in the ways in which a wide variety of publics can use this information. However, one issue is often left aside: could our present conception of open access impede future digital developments? In short, thinking about what role open access and repositories can play in this emerging digital world is the theme of this small presentation.

We must also be aware that the presence of open access and repositories will loop back upon the way digital cultures function, just as libraries have done for the Gutenberg era. And the comparison with libraries also bears important consequences because librarians harbour fundamental social values: it provides a reference point that should help future generations protect precious human rights, exactly as libraries nowadays offer an anchoring point to defend precious forms of individual freedom.

II. What does “going digital” mean?

As digitization proceeds and expands, the sense that something fundamental is changing is growing. But this sense remains impressionistic. Its extent is unknown, as is the time scale that encompasses this transformation. It feels big, but how big? How revolutionary? Some degree of anxiety also accompanies the process, mixed with some hopes.

In contemplating the fate of our cultures as they move through the digital threshold, we should never forget to look closely at what happened when a similar transition took place after Gutenberg's innovations took hold in Europe. The simple fact that the meaning of this mutation is still the object

of heated debates nowadays shows how complex a transformation it was. So is the present digital threshold.

Print was not a just a simple technical way to solve the problem of multiplying texts more easily. In particular, it displaced the centres that controlled the production and the diffusion of texts and it threatened well-established forms of power. Also, it unfolded in fits where quick accelerations followed periods of relative stagnation, and even a few crises. It was a large process which, once launched, never ceased to evolve: in fact, the print world is still evolving.

This simple observation should give us pause. In particular, it warns us against thinking that we are going to jump into the digital world with one hop and then all will be different, but all will be well. On the contrary, we can expect protracted battles to shape the future digital world according to the wills and desires of each one of us. The issue for most of us is to know how to act to be heard and influence events to make them move in the direction we desire.

One cannot mention Gutenberg as an example without issuing another word of caution: while the example of print helps us think about salient issues and assists us in raising potentially important questions, it should not lead us into believing that events will repeat themselves. Issues, if framed at a sufficiently high conceptual level, may recur, but not events. Building the conceptual framework is really the issue here.

Let me give a small example of what I mean. I often find myself saying that we are presently producing the incunables of the digital era. The image is pretty obvious and others, for example G. Crane, have independently reached similar formulations. However, in saying this, I do more than just creating an amusing metaphor. In this image, the past acts more as a cautionary tale than as a way to predict the unfolding of digital time and the phases structuring the digital era. I confess I often smile when looking at a pdf file, sometimes correctly labelled as “electronic paper”. Why do I smile? Because I think about early printed books that strove to look as much like manuscripts as they could. I smile and I hope to make people smile too. But I also hope they peer a little deeper into the issues raised by digitization by virtue of the critical distance that the metaphor is trying to establish. While it is not impossible that future historians will see the first few decades of the digital era as similar to the incunable period and while I would even go as far as to say that there is some probability in this outcome, the essential lesson is elsewhere: when we look at a digital object nowadays, please let us not be taken in by its apparent stability or fixity. Actually, it is nothing more than a snapshot of a moving target. The pdf format is just that – a format with a beginning and an end – which must be placed within not one but several generational strings of formats. Pdf, let us remember, is nothing more than the heir of Postscript, a language aimed at stabilizing the relationship between a computer and a printer. Pdf belongs entirely in the print world. Pdf has something to do with the “page”, not with digitization. Whatever importance pdf may have, it is tied to its relationship to the page – an object which merits study[1].

XML is another digital file format but it and pdf do not belong to the same historical “thread”. It aims at document structure rather than appearance. Which one reflects the past and which one is a

1 See for example: *The Future of the Page*, Peter Stoicheff and Andrew Taylor, eds. (Toronto, University of Toronto Press, 2004).

prototype of the future is precisely the kind of discussion we must carry on within our heads. Which one is cheap and easy and which one is complex and hard to implement is the kind of pragmatic debate that often blinds us and diverts us from the conceptual essence of a situation.

III. J. C. R. Licklider's insights

Some people saw the emerging picture of digitization rather quickly, amazingly quickly, in fact. One of the very best ever was J. C. R. Licklider. His life intersects the history of the Internet when he was involved with ARPAnet; less known perhaps is the fact that his career also intersected that of libraries[2]. At the invitation of the Council on Library Ressources, he undertook a study on the future of libraries in late 1961 which finally became a report in early 1964 and a book published by MIT Press the following year. Appropriately enough, the book kept the theme of the project as its title: "Libraries of the Future" is one of those books that retains an extraordinary freshness and provides incisive insights even though it was penned over forty years ago. Quite fittingly, it is dedicated to Vannevar Bush, of "Memex" fame[3].

Quoting the first two sentences gives a feeling for Licklider's slender volume and its tone:

> The "libraries" of the phrase, "libraries of the future," may not be very much like present-day libraries, and the terme "library" rooted in "book", is not truly appropriate to the kind of system on which the study is focused. We delimited the scope of the study, almost at the outset, to functions, classes of information, and domains of knowledge in which the items of basic interest are not the print or paper, and not the words and sentences themselves – but the facts, concepts, principles and ideas that lie behind the visible and tangible aspects of documents.[4]

To study the libraries of the future, Licklider immediately shifts ground while proceeding to provide some limits to his projects: only "transformable information" will hold his attention. Works of art and works of literature (to a lesser extent) are excluded, but Licklider's boundaries still take in most human intellectual productions. He also sets a time frame for his study: the year 2000. Because of this leap into the future – essentially forty years -, a more conceptual approach becomes both necessary and possible.

Licklider's analysis is strikingly bold. He invites us to look at libraries as systems, books as sub-systems and pages as components, and this leads him to a series of important conclusions:

1. The printed page is superb for the display of information;
2. Binding pages together to make a book destroys or decreases many of the qualities of the page;

---

2 See M. Mitchell Waldrop, *The Dream Machine: J.C.R. Licklider and the Revolution That Made Computing Personal* (Penguin, 2002).

3 See Vannevar Bush, "As we May Think", http://www.theatlantic.com/doc/194507/bush . Licklider admits having read the famous Atlantic Monthly paper only after having completed his report – an anecdote that can stand as an ironic commentary on libraries...

4 J. C. R. Licklider, *Libraries of the Future* (Cambridge, Mass., MIT Press, 1965), pp. 1-2.

3. If books are unsatisfactory “for the storage, organization, retrieval and display of information”, then libraries of books cannot do any better.

The strength of Licklider's strategy lies in his ability to free himself (and us) from the mesmerizing effect that books as objects induce, especially when one thinks about a library. Instead, he nudges us to think about a library without books and, to do so, he seizes upon an entirely different question: how do human beings interact with information. Licklider can do this because, for him, the library is not essentially a collection of books; rather, it is an information system. More precisely, it is a device designed to facilitate, amplify and improve the interactive process between humans and information sources. Alas, libraries of books are strongly limited on that score. They run into a fundamental obstacle – namely what Licklider calls the “passiveness” of the pages and of the books. But as we are also extraordinarily “inured to this passiveness”, he says, all we can ask is an apparently absurd question: “Do you suggest that the document read its own print?”[5]

Basically, Licklider's reaction has been to take on the apparently absurd question and find an answer to it. This is how he phrases it (and remember once more that it was in the early sixties...):

> We need to substitute for the book a device that will make it easy to transmit information without transporting material, and will not only present information to people but also process it for them, following procedures they specify, apply, monitor and, if necessary, revise and reapply. To provide those services, a meld of library and computer is evidently required.[6]

Concepts can achieve this kind of result: after refusing the pragmatist's judgment of “absurdity”, he can go further by taking advantage of the results of an apparently foolish question. If a library is a device to facilitate human interaction with knowledge, then rebuilding the library around something like a computer begins to make a lot of sense. This looks obvious with the 20-20 vision of hindsight, but it was wonderfully visionary ca. 1962-3.

As I am not lecturing on Licklider, I will stop here and do a fast-forward move to the present. Unhappily, this will force us to skip over Bob Kahn and Vint Cerf. They are better known as the creators of the TCP/IP protocols, but they also wrote a technical report on digital libraries and developed the notion of “knowbots”[7] back in 1988. But doing so will allow us to listen to one of the more visionary thinkers about libraries: Clifford Lynch.

IV. What is missing in open access? Clifford Lynch's viewpoint

In a recent paper[8], Clifford Lynch echoes Licklider's vision while putting it in the form of a question:

---

5 *Ibid*., pp. 5-6.
6 *Ibid*., p. 6.
7 *The Digital Library Project. Volume 1: The World of Knowbots* (draft). http://hdl.handle.net/4263537/2091 .
8 Clifford Lynch, “Open Computation: Beyond Human Reader-Centric Views of Scholarly Literatures,” in *Open Access: Key Strategic, Technical and Economic Aspects*, Neil Jacobs, ed. (Oxford, Chandos Publishing, 2006), 185-193.

> As the scholarly literature moves to digital form, what is actually needed to move beyond a system that just replicates all of our assumptions that this literature is only read, and read only by human beings, one article at a time?

If others beside humans (that is machines) “read” documents, we are not very far from documents reading themselves...

A great deal of Lynch's analysis relies on a thought experiment: what would happen, he asks, if copyright issues regarding scholarly and scientific publications were suddenly limited to respecting attribution? In working out this thought experiment, he basically demonstrates two points:

1. Yes, it would provide great improvements over the present,
2. No, it would still not be enough.

In other words, while he demonstrates that open access brings us closer to the desired objective of an optimal communication system for scientists and scholars, he also stresses that, as presently conceived, open access may not yet be enough. That is very important because most of the present discussions about open access focus on purifying the definition to make it more rigorous, rather than preparing it for future evolution. It also points to a sense of rigidity that is yet another bad sign for the future.

What are the limitations of present-day open access according to Lynch? First of all, copyright, even if limited to attribution, would still raise problems for results derived from the algorithmic processing of large corpora of texts. Given the rules on derivative work stemming from our property laws, the results could fall into the “derivative work” category. If so, even with works placed in open access, an absurd consequence could emerge through the attribution rule: all the thousands upon thousands of authors included in the processed large-scale corpus might have to be named as co-authors of the algorithmic results. To some extent, Lynch continues, Creative Commons helps clarify this situation, but it still does not go far enough. Work presently done around Science Commons only begins to grapple with these issues more precisely. In any case, both activities only confirm his suspicions that simple open access will simply not do.

Clifford Lynch is drawing our attention on the requirements needed to perform what he calls “open computation”. Enormous bodies of data are being stored every day. Large quantities of texts discuss and interpret these data in the form of journal articles. Both categories of information are now parts of huge sets of materials and when they become digital, they also become primary sources for a whole new category of procedures that we are just beginning to identify as a category and which evidently will grow almost infinitely in the future. These procedures are the forms of reading done by others than humans. The possibility of “reading” computers amounts to discovering an entirely new knowledge continent. If we want to make the best use of this continent, we will need “open computation” and we must therefore examine the conditions under which it can freely take place and develop.

To make ideas more concrete, let us carry out a thought experiment of our own. Imagine that all the laws of a given country were digitized (incidentally, they often are nowadays), accompanied with all the court decisions ever made in the same country. Some juridictions have begun to move in this direction, for example Canada with its Supreme Court decisions, but this is only the tip of the

iceberg. Tom Bruce, from Cornell University who has been studying this issue for years knows the full complexity of the work. Nonetheless, let us imagine it is done, open, and available to various individuals, companies (Google, for example), etc. This means that in a few seconds, the most isolated lawyer in the smallest community could check the jurisprudence of his/her country in ways that not even the largest legal firms of the United States can approach nowadays. This also means that access to the legal wisdom of a country would be much more evenly distributed. It would change the position of the law in society: not only would it apply equally to all – the present situation,more or less – but it would also be accessible to all – a situation from which we are very far, and which strongly contributes to reinforcing the power of the rich against the poor.

Now think of the conditions needed to reach this objective. Even when the law is in the public domain, its physical expression is often proprietary. A recent debate in Britain about the consolidated law of Parliament demonstrates this point all too well. Consider the following Crown Copyright declaration, only visible to the few that had actually access to the consolidated law database:

> "The Statute Law Database and the material on the SLD website are subject to Crown copyright protection. The Crown copyright waiver that applies to published legislation generally does not apply to SLD because it is a value-added product. Any reuse of material from SLD will be the subject of separate and specific licensing arrangements. No such arrangements have yet been entered into. Users should not therefore reproduce or reuse any material from SLD until further guidance is issued."[9]

Nothing is more in open access, so it seems, than the public domain. Yet, in the case of the law, gaining access to the document so as to transform it into an “open access” document is not necessarily a trivial operation. If, furthermore, the materialization is such that performing a number of tasks such as retrieval, comparisons, ordering according to various criteria, etc. becomes impossible or very difficult, then the potential benefits of open access will not be realized.

The example above illustrates Clifford Lynch's concerns. Even now, documents are not just read – what does that mean anyway -, and they are being “read” by others beside humans. This is exactly Licklider's point forty-five or so years ago. Lynch reiterates Licklider's point and (rightly) admonishes us to think about the present conceptions of open access so as to prevent turning them into obstacles to open computation. Present visions based on harvestable, inter-operable inventories are wonderful, but they remain “incunabular”, to use Grave's adjective. On the other hand, Cliff Lynch continues, if “open computation” could be grafted upon the ever growing digital collections of documents, then

> The opportunities are truly stunning. They point towards entirely new ways to think about the scholarly literature (and the underlying evidence that supports scholarship) as an active, computationally-enabled representation knowledge that lives, grows and interacts with its contributors rather than as a passive archive or record.

The same refusal of documents as fixed, passive content guides J. C. R. Licklider's and Cliff Lynch's thinking. It is interesting to reflect a moment about a conceptual echo that spans more than forty years. Licklider's capacity to peer into possible futures never ceases to amaze, but what is perhaps just as extraordinary is that, nearly half a century later, Clifford Lynch still surprises us and pulls us out of our Gutenberg-based complacency. Obviously, the forty-year time lag between Licklider and Lynch demonstrates that moving out of the Gutenberg era and associated mentality is

9 See Peter Suber's blog on August 18th (10.03.20 AM) , http://www.earlham.edu/~peters/fos/fosblog.html .

a bit more difficult and a little slower than generally realized. As McLuhan was fond of saying, when we try to look at the future, we tend to use our rear-view mirrors. This is what most of us have been doing while trying to peer into the digital future.

And yet, let us not forfeit the rear-view mirrors; they may be useful to identify what was prefiguring the future back then, and compare it to our best visionary thinkers nowadays. Rear-view mirrors are quite useful when they point to elements of the present past that will survive in the future..

V. The scientific revolution: bootstrapping distributed intelligence.

So far, the following points have been made:

1. We have been reminded that open access is dependent upon digitization. Without it, OA simply remains inconceivable;
2. We have been warned that open access, in its present form, does not go far enough to cover some of the future developments associated with digitization;
3. We have been pointed in one general, important direction, which is that the advent of large digital corpora means that a whole new layer of computational activity has to be factored in. In effect, this is the lesson we can already and directly draw from Google.
4. In trying to chart a good road toward the future, we should look for the deeper continuities which, at the conceptual level, underpin the superficially variegated landscape of evolving technologies.

In looking for some stable conceptual level of analysis, Licklider chose to focus on human-document interaction. This led him to consider new kinds of interactions, such as situations where the document, in effect, “reads itself”. We need to look for some equivalent conceptual shift needed to peer into the future. And although the present and rapid emergence of large digital corpora is very impressive and “mass digitization” is fast becoming a buzz expression, we should not be mesmerized by the corpus-in-itself, the object, however large or impressive. The human element is obviously central in the information and knowledge equations and we should therefore reject any basis for analysis that would neglect this crucial element. It is true that “open computation” amounts to new forms of “reading”, but the meaning to be assigned to these new forms of reading cannot be extracted from the machines doing these kinds of “reading”.

Even though documents are no longer “read” only by humans, and even though documents do not exist only to be read, ultimately they can live, survive and resonate only if they remain in symbiosis with human beings. Like Licklider, we need to focus on the nature of this symbiosis or interaction and not on the artefactual side of documents. But we must also move Licklider,s vision which was largely grounded in the psychology of the individual. Instead of simple, atomistically conceived, individuals, we may have to deal first and foremost with various kinds of human groupings: epistemic communities, communities of practice, collaboratories, networks, etc. Not only do documents need humans to survive and thrive, but they also need structured networks of humans[10]. Conversely, censorship is never entirely effective if it limits itself to destroying documents, or

10 An interesting development in this regard is the Biomedical research portal, BioWizard. See http://www.biowizard.com/.

burning books; full success also requires persecuting and dismantling the social networks that stand behind these documents.

As often, the history of science provides some interesting lessons when working on these difficult issues. For example, Loet Leydersdorff, rightly argues that:

> " With hindsight, the scientific revolution of the seventeenth century can be recognized as first and foremost a communication revolution. The printing press made it possible to change the dynamics of information storage, archiving, and retrieval."[11]

In effect, with the advent of print, the possibility of "intellectual action at a distance" was greatly amplified. Larger groups of people could compete in the wider intellectual arenas provided by print documentation. Journals and articles emerged and quickly multiplied from the new context provided by the Gutenberg innovation. Natural philosophers became, *de facto*, the reproductive organs of these journals and articles; conversely, without the presence of these same journals and articles, natural philosophers would also have found it more difficult to reproduce ... as natural philosophers. A new symbiotic relationship between humans and documents was thus established by print. In some ways, it carried forward some of the more positive traits of the manuscript era, but it also introduced many new characteristics. For example, the multiplication of copies made any given document almost indestructible, so that natural philosophy found itself endowed with an unshakable permanent memory. Science had unwittingly gained a form of scholarly jurisprudence.

Print helped generate another consequence: in the manuscript era, most scholars spemnt most of their time to copying. By contrast, in the print world, scholars were freed from the concern to preserve through copying and this amounts to liberating a great deal of time for other tasks. Instead of copying, new and more interesting tasks emerged: collating, comparing, critiquing, cutting and pasting, and generally building upon the past. Climbing onto the shoulders of giants[12], although a medieval image, came into full force only after print. From preserving at all costs the lessons from the past, a whole civilisation moved to the desire to innovate, to be original beyond the memory produced by the print archive. The battle between the Ancients and the Moderns really does not mean anything else. Ultimately, this situation reinforced the trend toward individualism. Descartes really ought to have said is: "I am printed, therefore I am".

Something else also began to grow with the new relationship to knowledge, its production, its preservation and its diffusion: the authoritarian landscape of great teachers gradually gave way to a different social structure which has often been labelled as the Republic of science. Participants in this new intellectual game began to contest, to argue and to build on the basis of observations, experimentations and ... refutations. Article by article, journal by journal, school by school, a gigantic, multi-layered, century-long intellectual battle began to unfold and grow at exponential rates. Mediated by published articles in journals, with years of delays due to the publishing process and the inadequacies of library resources, this large-scale debate dwarfed all the medieval disputations that ever existed. This enormous intellectual conversation/debate also began to lumber forward. It did as if it were in a creaking, not very well designed vehicle, but progress it did. Of

11 Loet Leydesdorff, "Can networks of Journal-Journal Citations be Used as Indicators of Change in the Social Sciences?" *Journal of Documentation*, vol. 59, No 1 (2003), 84. Ironically, Leydesdorff is a scientometrist. Luckily, credentials and competencies do not always coincide...

12 The origin of this image is generally attributed to Bernard of Chartres (fl. 1117). Newton's recycling of the metaphor is more widely known.

course, attempts at restoring hierarchies and lines of authorities multiplied for the idea of a Republic of science did not agree with everybody, but, ultimately, peer review won and, sooner or later, community evaluation came to assist peer review whenever it went astray. From Mr. Blondlot's N rays to the recent scandal about stemm cells, the same mechanisms have kept the Republic of science fundamentally honest, independently of the morality of its participants.

The picture of science as it was just adumbrated should elicit comparisons with more recent events. This large conversation around a unified corpus of documents that slowly evolves, corrects itself, reworks itself and moves on, forcefully evokes a recent phenomenon -that of Wikipedia. Of course, print-based science is both creakier and more assured than present-day Wikipedia. In terms of efficiency, print-based science may appear like transmitting TCP/IP packets over smoke signals or drums: it has been done but the bandwidth is terrible. In terms of validation of knowledge, present-day science remains far ahead of Wikipedia. But this is for the moment and this moment appears increasingly as if it is going to be very, very short.

Once again, let us not be diverted by surface movements. What is perhaps most important in scinece is the ability of establishing what counts as true and real distinct from political or religious authorities. But saying so is to admit that scientific truth and reality have become contingent products of a particular community's activity. Somehow, and there lies the whole complexity of the Scientific Revolution, communities of natural philosophers and collections of documents somehow validated came together and allowed for the **bootstrapping**[13] of distributed intelligence. The scientific revolution actually witnessed the priming of intellectual production in a new mode[14].

Wikipedia is recapitulating this process all over again. For the moment, it is not yet focused on producing new knowledge, but it is well on its way to building a new distributed intelligence mechanism to aggregate existing, validated knowledge. It is not based on entirely the same communities as those involved in the scientific and scholarly efforts, but neither were all natural philosophers members of the clergy. Part of the resistance stems from this shift in communities because it challenges existing credential situations that have been fairly well stabilized in the last couple of centuries.

Science, therefore, is the result of the relentless activity of a distributed intelligence system. The mail system between European courts ignited it. Print gave it enough supplementary impetus to allow for self-emergence or auto-poiesis, but it was far from providing the optimal conditions for its full blossoming. As Licklider has reminded us, print also developed a fantastically good interface for humans: the page. However, books (or journals) are only fair when it comes to storage, and poor when it comes to retrieving information. Further progress was obviously possible, and Licklider saw it forty years ago.

13 See the Wikipedia article on this topic, http://en.wikipedia.org/wiki/Bootstrap. A conceptual extension of this metaphor has been carried out by Thierry Bardini. See his *Bootstrapping. Douglas Engelbart, Coevolution, and the Origins of Personal Computing* (Palo Alto, Stanford University Press, 2000).

14 This theme is masterfully explored by Yochai Benkler in his recent book, *The Wealth of Networks. How Social Production Transforms Markets and Freedoms*. (New Haven, Conn.: Yale University Press, 2006). A free, Creative Common, version is available at the following address: http://www.benkler.org.

VI. Doing science in 2040-50.

I will now try a little bit of Licklider's strategy myself. Simply allow me another forty years: it opens enough room for one's imagination to roam, which is fun; it also allows to maintain some contact with reality – a drag, of course, but quite a necessary one.

Circa 2040-50, the Republic of science should still be a republic. That is to say: the basic process of producing observations, experimental results and theoretical interpretations of these “facts” should remain very much alive. The scale of the worldwide disputation, will have probably grown enormously if only because 80% of humanity's population presently disenfranchised from the distributed intelligence system of science should begin to have access to the kind of education that allows for the mass production of scientists. I do not know what technology will be available forty years from now, but I can safely predict that the best of our present (ca. 2000) communication and storage technologies will be widely available by then, even among poor populations. This means imagining a world where one-megabit-connections are ubiquitous, and perhaps even in free access, exactly as most of our roads are nowadays. Of course, and as usual, the high tech of 2040, whatever it may be, will be in the hands of only a few. But this does not remove the important point that, forty years from now, the worldwide republic of science will be much bigger in terms of its demographics, and it will be completely digital.

Let us judiciously use our the McLuhan rear-view mirror at this stage of our reasoning. The difficulty here is to know where to look. We do know that the future is already being prototyped somewhere today. But where?

I would contend – and again, there is not much originality in saying this - that one of the good places where to look is among software developers, and particularly those who are involved in free software. Software is a peculiar form of document production that does not rely on the page or the book. However, what the disciples of Richard Stallman and Linus Torvalds have discovered is that when projects excite people, when people understand how symbolic rewards ultimately translate into monetary and position gains, then armies of contributors spontaneously emerge. These armies fluctuate rapidly in their composition, but the army remains and the continuity of the action is not threatened by this pulsating mode of involvement. After all, in the seventeenth century, science began like this too and the interest in science has never flagged even though the Republic of science has undergone some significant mutations in its history.

The free software world is itself a subset of the Republic of science. In my opinion, it is also showing the way of the future. And it is growing at phenomenal rates. More than a year ago, Eben Moglen, of the Free Software Foundation, in a talk at the Wizard of OS meeting in Berlin, was explaining how the free software movement already had the programming capacity of several Microsofts, and was growing considerably faster. Moglen's important measure of the phenomenon must be complemented by another remark: many of the people contributing code would probably never be hired by Microsoft; yet, despite the absence of classical credentials, they are quite capable of contributing and, let us not forget, they are legions. The free software culture has learned how to harness this neglected, yet phenomenal, source of distributed intelligence.

More generally, the whole software kingdom shows the way of the future in the way computer scientists, understandably enough, were quick to seize their own instruments to enhance their ways of working. The positive feedback on knowledge production that ARPAnet injected into the communities of computer scientists was spectacular almost from the outset, as the rate of growth of this and other networks demonstrates. The Internet is but the direct consequence of these trends, but unlike previous networks, its aim is to link up everybody. It also provided for a great amount of flexibility in the ways in which anyone can contribute. In the software world, no one tries to write the equivalent of an article or, worse, a book; yet, many do much more than that. But the ultimate unit of contribution is one line of code, no more, or perhaps simply a corrected line of code. A fully mature distributed intelligence system is also very tolerant with regard to the intensity of the contribution, which means many more people can contribute.

Another way to say similar things is to pick up on Jim Neal's remark in the December 9 issue of the *Chronicle of Higher Education*. As a good member of the Republic of science and a staunch believer in the virtues of distributed intelligence, - and I would be winking at Jim now, if I had the privilege of being with you - Jim quoted someone else. He used Clayton Christensen, the author of the *The Innovator's Dilemma* and in order to recycle the concept of disruptive technology and make use of it. This allowed Jim to remind his readers that, by disruptive technology, Christensen meant technologies that "enable a larger populations of less-skilled people to do things that historically only an expert could do." This is precisely what the free software movement and Wikipedia are doing. This is exactly what a fully matured distributed intelligence system fosters. As Thomas Friedman would exclaim in yet another semantic register, the world is indeed being flattened.[15]

And this is what digitization and the Internet are doing to almost everything human, and in particular to science. Of course, Wikipedia and other social-cognitive experiments are experiencing some teething pains here and there, but then so did science at the beginning. The ultimately stable social recipe may yet change in the future, and probably will, but if one considers what has been achieved worldwide in less than five years, one can only be astonished.

It is also necessary to conclude that the expression "mass media" has been turned upon its head: it used to mean mass audiences receiving messages from a few; it now means masses of producers depositing pretty much what they want in a kind of immense virtual and digital library. Like scientists, the producers in this case are also the users, so they turn around and seek what they are interested in in the same vast, virtual library, aided and abetted by Google and other similar tools. And this is just the beginning. If we do not mess it up by unexpectedly creating obstacles to this brilliant future. Like an ill-conceived and rigid form of open access.

VII. Return to open access and institutional repositories.

Caught between enormous corpora of digitized texts and seeing enormous networked knowledge communities in the offing, one may ask: what is more important? My answer is that this is the wrong question. Both are important, equally important in fact. And the real question is to ask: how

15 Thomas L. Friedman, *The World is Flat. A Brief History of the Twenty-First Century. Updated and Expanded* (New York, N.Y.: Farrar, Straus and Giroux, 2006).

will these huge masses of digital documents of all kinds interact with these huge armies of people involved in the knowledge production centres? Once again, we are facing an interfacing issue. Once again, we will discover that documents allow communities to maintain themselves and even expand, while the same documents could not be produced without the stewardship of these communities.

With the advent of the new communication tools, not surprisingly, new kinds of communities are fast emerging. Many will live only brief lives, but others will survive. And older types of communities will have to learn living in a vastly changed knowledge eco-system (and probably a different ego-system, to use Jan Velterop's amusing pun). Collaboratories are sprouting all over the planet in an increasing number of fields, ranging from astronomy to medicine. The debates about publishing data and making it available to anyone interested is on and was widely advertised by the race to publish (or not publish) the human genome. Gradually, researchers are becoming aware that sharing facts is a win-win situation, just like programmers have known for twenty years that sharing code benefited everybody.[16] In short, traditional scholarship and science as specific forms of practice are quickly discovering new neighbours in their vicinity. Open Access and institutional repositories relate to these new practices.

Institutional repositories are a key element in this general movement. They are developing faster and faster. In a very recent posting to American-Scientist, Alma Swan was proposing a rate of growth presently standing at about one new repository per day. However, as we all know, they do not fill up as quickly as one could hope.[17]

To accelerate the rate at which they fill, repositories must start again with the importance of the human-interaction interaction brought to light by Licklider and Lynch, as well the network dimension attached to this interaction. Depositories cannot expect to thrive if they are only sitting there: adding their own passivity to that of the traditional page will not work. Waiting simply to be harvested is not enough. Even if Google Scholar managed to solve the retrieval problem fully, it still would not solve the computational level identified by Clifford Lynch, especially in view of the fact that any performing form of harvesting will, left to its own devices, overwhelm any human readers with enormous quantties of documents to peruse, if not read. In effect, depositories must take a page (so to speak) from Wikipedia's experience and begin to evolve in such a way as to become a living carrier of the voices of some subset of the distributed intelligence system. Journals, it must be remembered once more, really reflect the values, choices and perceptions of some network of individuals. If repositoriees are to become vibrant elements of the globalized scientific dialogue, they must lean on strong, well-defined networked communities. In science, these are called disciplines, specialties or invisible colleges. Depositories are devices to facilitate the interaction between documents and communities, just like libraries were devices to enhance the interaction between documents and individuals. Depositories have to become libraries for networks rather than simply libraries for individuals.

16 See the important recent book by Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven, CT: Yale University Press, 2006).

17 Seeking the mandating of self-archiving has become one the holy grails of the Open Access movement, but other strategies can be advanced that will help move closer to the objective full open access objective – namely access to the peer reviewed research results from anywhere.

This means that, starting now, institutional repositories should orient themselves in two complementary directions: on the one hand, they should adapt their activities to specific local communities as much as possible; then repositories should begin to network in ways which agree with the history of their institutions and the networks of their individual faculty members. In doing so, they will begin to develop incentives to make open access grow as researchers want it to grow. Rather than being acceptable or reluctantly acceptable, mandated self-archiving would become an essential tool for the creation of profitable career and knowledge strategies and would not have to be mandated at all any longer. Repositories would then begin to ensure the transition from the traditional journal to a scientific version of Wikipedia by providing all the needed visibility, authority and prestige that authors seek through journal labels. And such repositories would naturally evolve in such a way as to facilitate and even promote open computation: too much symbolic capital would be wasted if it were otherwise.

VIII. Conclusion.

The present debates around open access have been extremely useful to place the issue in full light and make it advance. In the process, the idea of open access has been temporarily frozen, as one freezes code for a release version, but no one should take this definition as the definitive truth. The digital world is still so very young and we know so little of what will result because of it that claiming to know what open access is for all times and places is a little arrogant. I believe this is the important message of Clifford Lynch who has already succeeded in identifying some regrettable problems stemming from our present notion of open access thanks to his deep understanding of computational activities placed on top of large digitized corpora. We now know that “open computation” must accompany open access and the list may extend further.

Institutional repositories must take this result into account to improve their relevance and importance. They must also recentre their activities not to become a passive digital repository of “stuff”, but rather to become a visible, relevant tool that all faculty members, specialty after specialty, are eager to use. This will require repositories to work closely with networks of researchers, and to mesh with networks of repositories. The function of a repository is not to be a digital box that just sits there with the university's intellectual production; rather, it should be a tool allowing individuals, teams and institutions to reach ever further and weave new connections ever more densely. The idea is to transform repositories into places that feed into active intellectual debates, into prestige issues, into fame and authority, just as journals do nowadays. The idea is also to be a disruptive technology such that even a mediovre researcher still finds opportunities to verify calculations, measurements, etc.

Open access, redefined as above, and ably supported by institutional repositories (as well as by open access journals, but I have left this issue aside here to focus on the repositories) will then provide an adequate terrain for the full development of the digital potential. Ideas will flow more easily from brain to brain; evidence will be culled more easily as well. As Science Commons is beginning to demonstrate, we actually own bits of knowledge that do not yet look like knowledge because they have not yet managed to be scripted within some meaningful narrative. However, with the right kinds of computational devices, we may bring these factual segments into sufficient proximity to trigger the narrative supplement in one or several individuals who are themselves networked. Likewise, we find that dynamic strategies of preservation such as Lockss work relatively easily and very well when the documents are in open access; on the other hand, when proprietary rights are present, the same task becomes hellishly difficult and costly.

In short, it is absolutely true that open access and institutional repositories cannot exist without digitizations. And is it just as true that no optimal state of digital cultures can ever be achieved if open access, defined in a flexible, should I say “open”, way is not present. Open access supported by active networks of thousands of institutional repositories and even greater numbers of human networks is the sine qua non condition for digital cultures to reach new summits and science to flourish everywhere.