

Metajournals. A federalist proposal for scholarly communication and data aggregation

M.C. Pievatolo*

Abstract While the EU is building an open access¹ infrastructure of archives (e.g., Openaire²) and it is trying to implement it in the Horizon 2020 program, the gap between the tools and the human beings – researchers, citizen scientists, students, ordinary people – is still wide. The necessity to dictate open access publishing as a mandate for the EU funded research – ten years after the BOAI³ – is an obvious symptom of it: there is a chasm between the net and the public use of reason. To escalate the advancement and the reuse of research, we should federate the multitude of already existing open access journals in federal open overlay journals that receive their contents from the member journals and boost it with their aggregation power and their semantic web tools.

The article contains both the theoretical basis and the guidelines for a project whose goals are:

1. making open access journals visible, highly cited and powerful, by federating them into wide disciplinary overlay journals;
2. avoiding the traps of the “authors pay” open access business model, by exploiting one of the virtue of federalism: the federate journals can remain little and affordable, if they gain visibility from the power of the federal overlay journal aggregating them;
3. enriching the overlay journals both through semantic annotation tools and by means of open platforms dedicated to host ex post peer review and experts comments;
4. making the selection and evaluation processes and their resulting data as much as possible public and open, to avoid the pitfalls (e.g., the serials price crisis) experienced by the closed access publishing model.

It is about time to free academic publishing from its expensive walled gardens and to put to test the tools that can help us to transform it in one open forest, with one hundred flowers – and one hundred trailblazers.

* Università di Pisa, PI, Italy, pievatolo@dsp.unipi.it

¹ <http://legacy.earlham.edu/~peters/fos/overview.htm>

² <http://www.openaire.eu/>

³ <http://www.budapestopenaccessinitiative.org/boai-10-recommendations>

Keywords: open access, serial crisis, net neutrality, peer review, open scholarly communities, filter bubble, semantic web, data aggregation, academic publishing, academic freedom

1. Introduction: beyond digital humanities

While the EU is building an open access infrastructure of archives (e.g. Openaire) and it is trying to implement it in the Horizon 2020 program, the gap between the tools and the human beings – researchers, citizen scientists, students, ordinary people – is still wide. The necessity to dictate open access publishing as a man-date for the EU funded research is an obvious symptom of it: there is a chasm between the net and the public use of reason [Kant 1784: 37] as it is customarily understood by academics, between the information treasures made available by sites like Europeana and their actual understanding and usage, and, in some countries, also between a research assessment [Gillies 2011] that relies heavily on the bibliometrics of proprietary, closed access databases (Wok, Scopus)⁴ and the very concept of publication – in the etymological meaning of making something public [Guédon 2001; Molinié and Bodenhausen 2010]. Is it just a matter of time?

To answer such a question, we have to focus on an area wider than the small portion of the net – open or closed – that is dedicated to researchers. In fact, while the internet and its major service, the web, were conceived in an environment of researchers, its very success transformed it in the ground of diverse concerns: researchers themselves, pulled as they are – as authors – by political and business concerns are affected by them. Therefore, relying on the openness of the net and on the Mertonian science ideal [Merton 1979] might not be enough. In addition, while the physicists gave their openly accessible shape to the net – by the very invention of the web, and by their openly accessible ArXiv – human and social scientists are, generally, rather passive in their use of it. The very presence of the “digital” humanities [Marche 2012] seems to imply that the building of a knowledge network – of a re-public of science in which we can be citizens, [Crane 2011] not subjects – is not a concern for the humanities at large. The humanists that are not “digital”, in other words, seem to believe that they can keep on living comfortably in their academy, even if their publishing practice are becoming obsolete and expensive⁵, and the net is going elsewhere, without their saying and outside their control – even if their self-styled ivory tower is, to quote Kathleen Fitzpatrick, more and more similar to an academy of undead [Fitzpatrick 2011].

To remain alive, and to be living contributors to our cultural awareness, we have to rethink and to take back, as scholars, our communication processes. We do not need to create another bubble of expertise, another small group of “digital humanists” around some specialized tool. We need to promote a general change in our way of communicating and of assessing our work, so that we become able to extract a scholarly and social value from the sophisticated tools – digital libraries, semantic web and linked data applications – that are now spending their existence swimming in bright, but insulated, bubbles.

⁴ See, on the Italian idea of building a research assessment exercise on closed, proprietary data, [Galimberti 2011]

⁵ Sometimes, humanists are well aware of the insularity of their studies, even if they are not accustomed to work as experimental political philosophers. See for instance [Capra 2012]

As we do not want to create another walled garden, but trails in the open forest, we have (1) to consider the state of the forest or the web at large, outside the academic bubbles; (2) to consider the state of academic walled gardens; (3) to ask whether there is a way to open the gardens to the forest without making them wild, and giving the forest the opportunity to contaminate itself with a wide variety of carefully cultivated flora. While the first two points linger in the fields of applied and theoretical political philosophy, the last one will suggest a social software project related to a federal overlay journal – a journal of journals – whose first layer is application-neutral, but whose second layer should be enriched by a panoply of semantic web tools and API, to make the overlay journal a thick, but walkable forest.

According to a Clay Shirky's well-known speech, social software is the experimental wing of political philosophy. If it is so, it is about time that political philosophers take on their responsibilities and start experimenting.

2. The state of the web

2.1. Tinkering with philosophy

Stating that philosophers and social scientists do not study the internet would be unfair. It is, for instance, a favorite topic of philosophers who see the social reality as documentality [Ferraris 2008: 110-123], or, on the contrary, as a more ethereal infosphere inhabited by inforgs [Floridi 2006].

Such theories, however interesting, share a major flaw: they all have a contemplative stance. While philosophy, as theory, is necessarily contemplative, there is, however, a topic in which it is unavoidably active, even without awareness: the field of the communication of itself. Plato did not need to foresee the internet when, in the dialog *Phaedrus*, meditated on the opportunities and the limitations of writing. Socrates does not discuss with a communication technician, like *Phaedrus*, to scrutinize a contemplative question – whether the soul exists or not: that is only the matter of a myth – but to understand the relation between the media technologies and the environment they help to create (*Phaedrus*, 274b-275c). Is it sociology of communication? Certainly. But it is also philosophy as criticism and theoretical practice, as a reflection that discusses its tools while shaping them, rather than crystallizing them in a kind of metaphysical contemplation that accepts them as a fait accompli.

Dealing with digitization in a contemplative way is not just pointless: it may be dangerous, if we fail to understand that, on the net, a philosophy is – without our say - constantly under construction around us.

We make up extensions to your being, like remote eyes and ears (web-cams and mobile phones) and expanded memory (the world of details you can search for online). These become the structures by which you connect to the world and other people. These structures in turn can change how you conceive of yourself and the world. We tinker with your philosophy by direct manipulation of your cognitive experience, not indirectly, through argument. It takes only a tiny group of engineers to create technology that can shape the entire future of human experience with

incredible speed. Therefore, crucial arguments about the human relationship with technology should take place between developers and users before such direct manipulations are designed [Lanier 2010: ch. I].

Software is not a fact: it is a construction, a worldview, which affects us – as users – before any reasoning. Just like the poetry that Plato criticized in the Republic (392c ff.), it is constrained by tradition or, more specifically, by the so-called lock-in, which transforms thoughts into facts.

Humanists are well acquainted with such a phenomenon: it is the sedimentation of ideas, mores and institutions that makes human cultures so powerful and constraining – it is Hegel’s *Sittlichkeit*, it is Burke’s convention and prescription, it is the system of guardianship that makes so difficult to get out of tutelage in Kant’s *Beantwortung der Frage: Was ist Aufklärung?* (35-36). However, when it applies to software, the very same phenomenon leaves humanists helpless and inclined to an ancillary contemplation of the world as it is, unless they are “digital humanists”. It is good to be skeptical on the tools we use, but the difference between skepticism and technophobia lies in a technical knowledge that seldom concerns the contemplatives. The very concept of digital humanities – as distinct from the humanities in general – seems to assume that the tools and the interfaces of knowledge do not concern the human science in general, but only a very exclusive, specialized field. In a networked republic of science, if code is law [Lessig 2000], such a belief is likely to educate passive subjects rather than active, well-informed citizens.

2.2. Living in the bubble

The web was invented in a research environment, for the Cern community; its architecture was minimal and open. It earned a spontaneous cooperation and participation just because it was a good idea. To use the web 1.0 as authors, people had to learn HTML and plan personally the interface through which they introduced themselves to the world. Proprietary social networks, while making life easier for HTML illiterates, encourage people to create standardized presences, to fit them more easily in their databases. But accepting to subordinate ourselves to their reductive patterns and to their restraining of people in bubbles [Pariser 2011] – LinkedIn, Facebook, but also Academia.edu or Citeulike or Mendeley – means to accept to reduce ourselves to stereotypes designed by others, for purposes that are different and sometimes opposite to ours. Indeed, if we follow the money, we can see that, on the internet, it is spent mainly in advertising: from the perspective of the money, the internet is a giant manipulation machine based on a carefully cultivated illiteracy, an enclosure of private and public data and a blind self-complacency [Lanier 2010: ch. III].

Can an internet made of proprietary bubbles and of a proprietary control on the data we are giving away so happily think for us, just by means of the amount of its data, connections and calculation power? Certainly, if we let it do it, by reducing ourselves to useless appendages of the machine. While believing in an individual consciousness might appear as idly metaphysical as believing in the universal consciousness of the noosphere, there is practical difference between these two worldviews. Believing in a noosphere towering upon us produces a contemplative stance; believing in the autonomy of reason awakes a critical attitude: Kant himself was well aware that free thinking cannot come from an algorithm, but only from a very personal, uncomputable vocation [Kant, 1784: 36]. In such a perspective, the noosphere is, prosaically, just the world of culture, with all its pat-terns, its sedimentation, its lock-ins, its open or closed data, its networks and its bias.

Thinking it as a single metaphysical entity means treating it as a fact rather than as a construction. Even the noosphere can be collectively stupid, biased and crystallized – just like everyone of us. If it a universal, it is and should remain a universal without totality [Lévy 2001].

2.3. Digital feudalism

While the battle on copyright is raging in courtrooms and parliament chambers, as well as in books, its underlying question is philosophically deeper and wider than the oligopolies that keep on defending and enlarging it. How is it possible to acknowledge and compensate, in our societies, the work of intellect, without using some kind of monopoly [Boldrin and Levine 2008]?

In the age of manuscript, the work of intellect used to be a privilege that compensated itself, and it was usually assumed that its product was common. Authors who could not afford to keep themselves had to resign themselves to patronage. The invention of printing did not make them independent: it gave power to new mediators, the publishers. In the early modern age, the printing privilege held by the publishers – a companion of government censorship – was perpetual, while authors' only gain was connected to the sale of their manuscripts to them. After the Statute of Anne (1710) and the French Revolution laws on the *droit d'auteur* [Le Chapelier, 1791; Lakanal, 1793], the copyright was attributed to authors, making their interplay with publishers a little more balanced for a while [Rose 1998]. Nowadays, the copyright term in the EU is 70 years after the author's death, and its scope has been widely broadened, disconnecting it again from the individuality of authors and breaking the XVIIIth century balance. The hive mind and its exploitation of the work of individuals is not an invention of the digital age: the commons of knowledge and the attempts to exploit them to earn a private profit or to manipulate the people are as ancient as the human culture. The Web 2.0, its lords of the cloud and their digital serfs, who present them with stereotyped bits of them-selves, are just its most recent manifestation. It is very interesting to notice, by the way, that the very same criticism to the web 2.0 is made both by Jaron Lanier, who believes in the pedagogical value of the market [Lanier 2010: ch. IV], and by Dmytri Kleiner, “telekommunist” [Kleiner 2010].

2.4. Application neutrality

Should the network operators be allowed to differentiate the traffic that goes through their infrastructure, altering, therefore, a native characteristic of the internet [Basso et al. 2011] ? On a lower layer, this question – the question of the “network neutrality” and its economic, technical and political facets – has already received clear definitions and answers. For instance, the 2003 paper [Searls and Weinberger 2003] by Doc Searls and David Weinberger explained why “stupidity” is Internet greatest wealth: “The Internet only knows one thing: this bunch of bits needs to move from one end of the Net to another. There are technical reasons why stupidity is a good design. Stupid is sturdy. If a router fails, packets route around it, meaning that the Net stays up. Thanks to its stupidity, the Net welcomes new devices and people, so it grows quickly and in all directions.”

In a well-known 2010 paper [Berners-Lee 2010] Tim Berners-Lee criticized the proprietary social networks because they betrayed the very principle of the web as he originally conceived it: the principle that any person can share information with anyone else, anywhere. In Facebook, for instance, connections among data exists only within its site: of you want to see them, you have to

join Facebook and start entering your data as well. “Your social-networking site becomes a central platform - a closed silo of content, and one that does not give you full control over your information in it”. Facebook, with its social graph, is smarter than the Web, but to enjoy it you have to enter in a closed bubble: and a Web made of closed bubbles, whose lords have the power of shaping your experience of the net and to exploit your data, is not application-neutral any longer. This kind of architecture fragments the Web, suppressing its major quality: being a single, universal information space.

Taking back the web is not an easy enterprise: we can certainly design the distributed and free Diaspora against Facebook, or Identi.ca against Twitter, but, in a web of bubbles, they are just two more bubbles. Furthermore, while Facebook meets very basic social needs, to know and join Diaspora or Identi.ca a user should be fairly educated and sophisticated, so that the struggle of Diaspora against Facebook shall remain the uneven contest of a dwarf bubble against a giant one. If the value is in the aggregation, the biggest aggregators will remain the incumbent ones and will keep on balkanizing the Web.

3. Learning from experience: academic publishing and its crisis

3.1. The serial price crisis and the open access movement

Living in proprietary bubbles exploiting data given away for free, in exchange of some self-styled added value, is not unfamiliar to scholars. Well before the invention of the Web 2.0 we have already experienced an oligopoly based on the alienation of our work to an oligarchy of scientific publishers. They succeeded in making us believe that the worth of our texts did not depend on their content, but on their distributors, the so-called core journals, and their bibliometric aggregation, i.e. the impact factor calculated on the ISI database⁶. Such a system bears some similarities with the Web 2.0: it monetizes what is given away for free, to the advantage of very few lairds of learning and it encloses data in proprietary bubbles. The only difference is just an improvement: if data and their aggregation are the actual value, it is opportune to let people navigate “freely” within our closed systems, so that we are able to pay ourselves with their data.

The so-called serial price crisis as a result of the oligopolistic market of scholarly publishing is already well-known. Even the Harvard Library, in an April 17 2012 memorandum, communicated that it was not able to sustain any longer the prices of the major periodicals subscriptions and invited its scholars to “move prestige to open access”. The very circumstance that such an invitation should come from librarians is the symptom of a bitter fact: unlike Plato – and unlike Timothy Gowers – the majority of researchers is still keeping on believing the ways in which they do – or do not do – a public use of reason is not part of their research, but only of their career.

⁶ See for instance, in addition to the chapter 7 in [Guédon 2001], the Cern conference of L. Lessig, The architecture of access to scientific knowledge: just how badly we have messed this up, “CERN Colloquium and Library Science Talk”, 2011. As it can be seen in Fig. 1, the subscription costs increase of the scholarly journals is completely out of proportion if compared with the cost increase of otherwise similar products, like the monographs

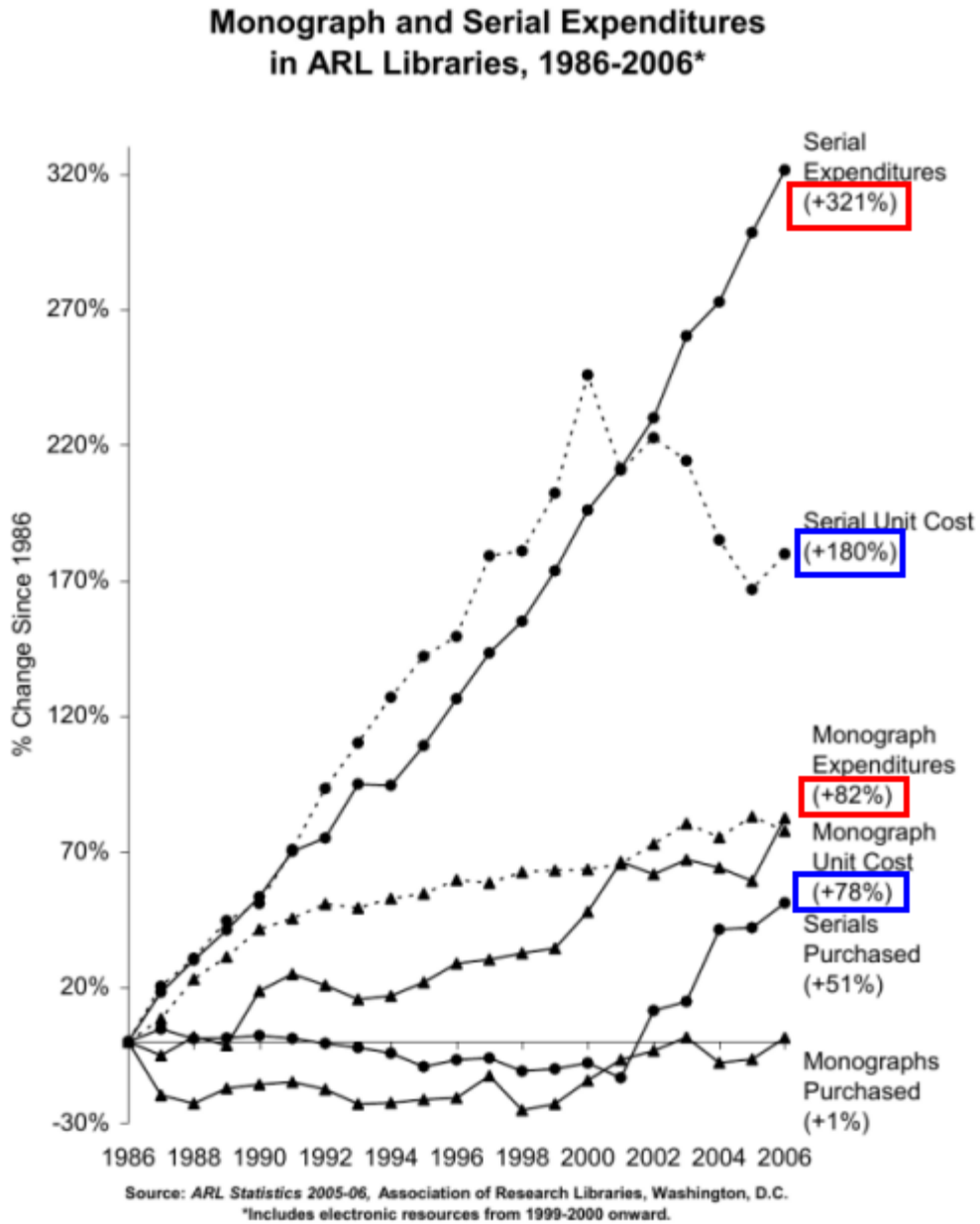


Fig. 1. [ARL 1986-2006] Monograph and Serial Expenditures in ARL Libraries

The open access movement countered the oligopoly in a surprising way: scholars endorsing it did not ask publishers to be paid in conformity with the profits they earn from the exploitation of their work. In their opinion, the value of the work of intellect is not monetary and, generally, money is not the ultimate measure of all values. There are cultures that are free from such a bias. The very patrons that in the early Modern age funded the science revolution lived in a world in which money,

rather than intellectual activity, should justify itself, by devoting itself to further values [David 2007; De Meulemeester 2012]. How can we learn to reaffirm the gratuitous meaning – the grace – of our work? How can we help society to recognize it by means of compensations different from monopolies, like crowd-funding or suitable tax policies?

It would be interesting to put to the test a two-pronged solution:

1. From an engineering point of view, we should go back, from the user generated content, to the user generated interface. Containers – from publishers to social networks – are not irrelevant, above all when they shape a world-view for us. Researchers, also in the field of humanities, should take back the code and ask developers to help them to get out of tutelage. Contents cannot be actually free if their containers are not free as well.
2. From a philosophical perspective, we need to link the question of the public use of reason with the degree of control everyone should have in their relationship with the public. Culture – with its patterns and prejudice – is under our collective responsibility. The proprietary social networks, just like the copyright-based monopolies, cannot replace the knowledge communities without taking away their freedom – the freedom of arts and science⁷. Such an idea should become as commonplace as the principle of free speech.

3.2. An instructive failure: Hyperjournal

Hyperjournal [Barbera and Di Donato 2005] was an open source web application for the administration of academic journals. To encourage a fair evaluation, the submitters were completely anonymous: the submission took place while the submitting users were logged off, and the attribution of the paper to an author was made, by means of an associate password, only after the conclusion of the peer review process. It provided also a dynamic contextualization engine enabling users reading an article to see not only a link to the articles cited but also a link to the articles citing it, within a single journal or within a contextualization federation. If more than one journal running on Hyperjournal had joined the federation, the resulting citation data would have been the basis of an open, not proprietary, expansible bibliometric.

Hyperjournal was praised by everyone, but it failed to reach the critical mass of users needed for the survival of a free software project. Its development, therefore, was stopped in 2007.

Why did the HyperJournal project fail? Like Diaspora, it is free software; like Diaspora, its goal had ethical undertones: making peer review fairer, opening up and democratizing bibliometric data. But to use Hyperjournal – or Diaspora – users should have very sophisticated needs and should be willing to learn a workwith new, unusual tools⁸. For this reason, HyperJournal created a community that was simply too little to survive, against the competition of established, proprietary bubbles meeting the very basic scholarly need of advancing in the career without asking too many questions.

⁷ According to Anil Dash, “privately-owned public spaces aren’t real public spaces. They don’t allow for the play and the chaos and the creativity and brilliance that only arise in spaces that don’t exist purely to generate profit. And they’re susceptible to being gradually gaslighted by the companies that own them” [Dash 2012]

⁸ On the other hand, the free software world was already offering a well-known, established tool like Open Journal Systems to the few scholars interested in treading the way of open access publishing. Choosing a new tool, without a critical mass of users, would have been a leap in the dark

There is a lesson to learn: if you do not like the bubbles, never compete with incumbent bubbles by creating further bubbles. Insularity is a weakness we do not need to imitate [Dash 2012].

4. A federalist solution: don't hate the aggregator, become the aggregator

While it is easy to detail the principle of openness and application-neutrality as the points of a program, it is not easy to implement them, because of the mentioned cultural lock-ins. We may well theorize – with Jean-Claude Guédon – that the current model of science publishing generates oligopolies, oligarchies and colonialism. The databases of Thomson-Reuters and Elsevier are designed according to the interests of the incumbent countries and of the incumbent publishers. If they are used in the domestic research evaluation, a researcher from a peripheral country trying to conquer the mainstream stage will be inclined to study topics that might be far from the concerns of his people and his culture: such a trend, aggregated, determines – in the EU like in the rest of the world – a kind of reverse Robin-Hoodism, draining intellectual power from the poor to the rich [Guédon 2008].

If we value the biodiversity of research, we should endeavor to reach a global power law distribution of the journals impact that has not only a very long tail, but it is also, as much as possible, flat. A “the winner takes it all”, competitive model sacrifices the diversity of sciences and cultures, and exposes our societies to the danger of helplessness in any “black swan” event. However, again, while theorizing is easy, we do need to ask how we can put theory into practice⁹.

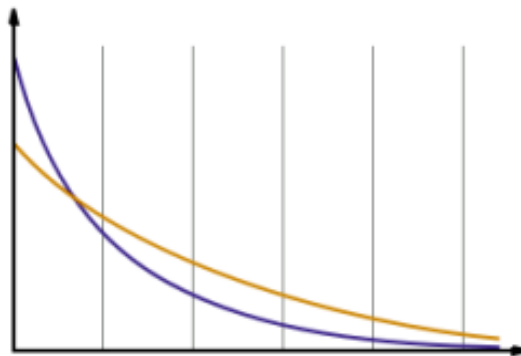


Fig. 2. Long and flat tails

⁹ [Brembs and Munafò 2013] suggest to abandon journals altogether, because of the many unintended consequences of their ranking, in favor of a library-based scholarly communication system. In a conservative environment, however, we have to overcome so many bias that it could be easier – or less difficult – to redesign journals while preserving their names, as this project is trying to propose

The net – it is true – has made publishing cheap and easy. However, it does not eliminate any incumbent position. The role of the core journals in scientific publishing is now played by aggregators like search engines and proprietary social networks, with their ability to orient our attention.

The open access publishing removes price barriers and permission barriers by exploiting the new opportunities offered by the internet. More than one half of the open access journals listed in the Doaj are free both to readers and to authors. They can afford to be free because they are little journals, whose workload can be a side activity of small groups of researchers. But if a little journal began to receive hundreds or thousands of submissions every month, it would need a more expensive management, which is paid by readers, in closed access publishing, or by authors, in open access publishing¹⁰. The “authors pay” model, while granting everyone the freedom to read, is exposed to the danger of putting some restrictions in the freedom to write.

If we reject the “authors pay” model, being little, obscure and poor seems to be the price of freedom. It would be a price worth paying, if it were only for the sake of freedom. However, a journal making such a choice might be damaged by the most witless research evaluation exercises and eclipsed by old and new aggregators. We have to face a dilemma: if we remain in the long tail of a power law distribution, we are free, but obscure and easily exploitable; if we become famous, we gain attention, but lose ourselves. Is there a way out?

A solution worth exploring is suggested by the experience of the so-called mega-journals, like “Plos One”¹¹. It aims to publish any article that met the test of scientific rigor, without taking into account any measure of importance or impact in its editorial and peer review process; it can afford to publish thousands articles every year, because it employs a huge quantity of academic editors in automated, scalable workflows.

If all the Doaj journals federate in wide disciplinary open access journals, we would get open access mega-journals that would be able to publish a huge amount of articles. Their federal nature would allow every title of the federation to preserve its internal organization, its editorial staff, its interests and workflows, while the federation as a whole would produce an amount of articles and citations too big to be ignored. *E pluribus unum*.

Such a federalism might be feasible if the open access journals conceive their activities as cooperative rather than rival. When can we describe a research project as successful? When it wins the bibliometric championship, or when it helps society to understand a portion of reality? The spirit of cooperation would be rewarded, by the way, by the sharing of the impact of the federal journal as a whole.

It would be a cultural revolution - one open garden, hundred flowers – that could help EU human and social sciences to become truly federalist i.e. both diverse and interconnected, by adopting a bottom-up approach.

Its goals, in a nutshell, would be:

1. making open access journals visible, highly cited and powerful, by federating them into wide disciplinary overlay journals;

¹⁰ See for instance the Plos publication fee policies

¹¹ The Ploh steering committee, whose aim is to create a Plos style open access journal for humanities, is considering a similar solution: see Academic Steering and Advocacy Committee Agenda #1 (25th February 2013)

2. avoiding the traps of the “authors pay” open access business model, by exploiting one of the virtue of federalism: the federate journals can remain little and affordable, if they gain visibility from the power of the federal overlay journal aggregating them;
3. enriching the overlay journals both through semantic annotation tools and by means of open platforms dedicated to host ex post peer review and experts comments;
4. making the selection and evaluation processes and their resulting data as much as possible public and open, to avoid the pitfalls experienced by the closed access publishing model (see paragraph 3.1.).

4.1. The challenge: thinking a federalist model for scholarly communication and data aggregation

The federal meta-journal would be new publication environment made of four interconnected layers.

4.1.1. Layer 1: the federated journals

The first layer should be a network of federated open access journals. According to the application neutrality principle, no strict technological requirement should be needed: each journal might be based on a different platforms including Open Journal System (OJS), generic content management systems (as Wordpress, Joomla, Drupal, etc.), scholarship targeted platforms as Omeka, or even specific ad-hoc web applications. An initial federation could be constituted by selected open access journals. However, such a federation will be open and possibly grow as the project goes on, the only requirement being that of adhering to the Open Access principles.

4.1.2. Layer 2: the semantic overlay journal

The second layer is the semantic overlay journal, whose core is an aggregator of articles from Open Access journals and archives. Articles will be aggregated in different ways.

1. **Automatic submission via metadata harvesting.** First, it should be possible to automatically populate the overlay journal relying on metadata harvesting protocols, as OAI PMH¹², which is the de facto standard in this context, and supporting the most used metadata schemas. While some platforms (as OJS, Wordpress¹³, Omeka¹⁴) already provides solutions for exposing metadata in compliance with the OAI-PMH standard, other platforms should possibly implement specific plugins to support the protocol. However, a research strand of the project will be that of designing a metadata harvesting protocol based on the emerging Linked Data paradigm. As ongoing work within the Europeana initiative (e.g. see

¹² The Open Archives Initiative Protocol for Metadata Harvesting
<http://www.openarchives.org/OAI/openarchivesprotocol.html>

¹³ DOAJ Export Wordpress plugin, <http://wordpress.org/extend/plugins/doaj-export/>

¹⁴ Omeka OaipmhHarvester, <http://omeka.org/codex/Plugins/OaipmhHarvester>

the Europeana Linked Data prototype or the DM2E project¹⁵), the Linked Data approach allows more flexible and rich semantics to be expressed and higher level of contextualization of cultural objects to be supported. Other protocols to automatically collect articles from the federated network should be simple and widely supported approaches such as RSS and Atom. The editorial board of the overlay journal will then be able to periodically choose from the freshly harvested articles those that will be included in the overlay journal, in order to guarantee a qualitative standard.

2. **Bottom up submissions from end-users.** To go beyond the top down approach followed by the traditional journals, scholars and end users should be involved in the process of building an interdisciplinary quality archive of open access publications. Like in a traditional environment, scholars should possibly submit original articles directly to the overlay journal, but they should be able to submit already published works as well. To such purpose, the overlay journal should offer a lightweight web clipping tool (e.g. a browser plugin or a bookmarklet) similar to the one used by popular systems like Mendeley or Pinterest. Using this simple tool, scholars should be able to quickly submit any article they come across browsing the web (even outside of the federated network) and to specify basic metadata such as author, relevant tags and comments. Before being published online, submission to the overlay journal will go through a peer review work-flow. Such filter will ensure that the submissions are relevant and will mostly act as an “anti-spam” filter, while the core review process will be constituted by an open peer review (see Layer 4).

4.1.3. Layer 3: semantic tools and content curation/enrichment

The semantic overlay journal should be not only a collection of papers, but also a working environment, where scholars can easily find and discover interesting articles as well as perform actual research on them. The aggregator should become more lively and social by offering people the research trails that cannot be found in open access archives. For this reason we need (a) curation tools to add human choices and touch (such as a blog) and (b) semantic annotation tools which allow metadata extraction as well as the possibility to recombine users annotations in order to create new scholarly contributions.

1. **Automatic metadata extraction.** All the articles published in the overlay journal should be processed to extract as much semantics as possible. Available terminology extraction tools (e.g. DataTXT) should be used to automatically tag and contextualize articles, linking them to the Web of Data. Such a semantic tagging procedure attributes labels to articles and links them to an interconnected semantic network (DBpedia), where additional data can be used to provide novel and powerful recommendations to users.
2. **Semantic annotation and enrichment.** Users of the overlay journal should not only be able to explore the archive and read articles: they should have at disposal built-in tools to enhance their research work by enabling them to annotate, categorize, put in relation and augment contents. To that purpose, the Pundit semantic annotation tool should be integrated into the journal and further developed to meet the specific requirements

¹⁵ Digitised Manuscripts to Europeana, <http://dm2e.eu/>

individuated in a first phase of the project. With Pundit scholars should, on the one hand, create private or public annotations (possibly in a collaborative fashion) that help them in understanding and studying the content, and, on the other hand, contextually create new structured data and metadata that the system can use, behind the scenes, to incrementally improve browsing and recommendation capabilities. While Pundit should be built-in into the semantic overlay journal, annotation capabilities should not be confined to the article published in the journal, as this would lead to yet another closed system. On the contrary, the same annotation tool should be available as a bookmarklet that users could easily install into their browsers and that should allow them to annotate generic web pages on the web (e.g. other journals but also blogs, newspapers, etc.), as well as to directly link (via semantic relations) journal articles with other web resources relevant to their research activity (e.g. primary sources, Wikipedia entries, etc.). Once the creation of such a personal, web scale, knowledge is made possible, it can be used to support a number of scholarly activities.

- a. **Creating interactive e-learning objects.** The web, and in particular the hypertext, is a powerful mean for scholars to create interactive learning materials where critical texts can be linked to online primary or secondary sources that are matter of study and where key concepts can be linked to encyclopedia entries or to other related web resources to facilitate the understanding of text by students. Examples in this contexts are given, for example, by the “Bollettino telematico di filosofia politica” where a set of hypertexts are used as learning material in teaching philosophy. Such learning objects, however, are created “by hand”, by writing down XML pages which are then converted into XHTML. A tool based on semantic annotations could allow an easier creation of such learning objects. Scholars will be able to mark specific text excerpts (from journal articles as well as from other “external” online resources) and add commentaries and semantic links pointing to online entities and other web resources (e.g. linking a cited text, or a mentioned concept). Then they will be able to semi-automatically generate hypertexts out of such annotations and publish them on the web as learning objects to be consumed by students.
- b. **Supporting new contributions authoring.** New papers and contributions (especially in the Humanities) result from the study and the review of existing literature. Scholars usually take notes aside of books or papers and then integrate such notes to into new contributions. One of the goal of the project is to implement such pattern in a web environment. Annotations created with Pundit could generate a new editable document containing all the commentaries collected by a scholar, along with contextual references to the annotated texts in order to facilitate the authoring of a new paper.

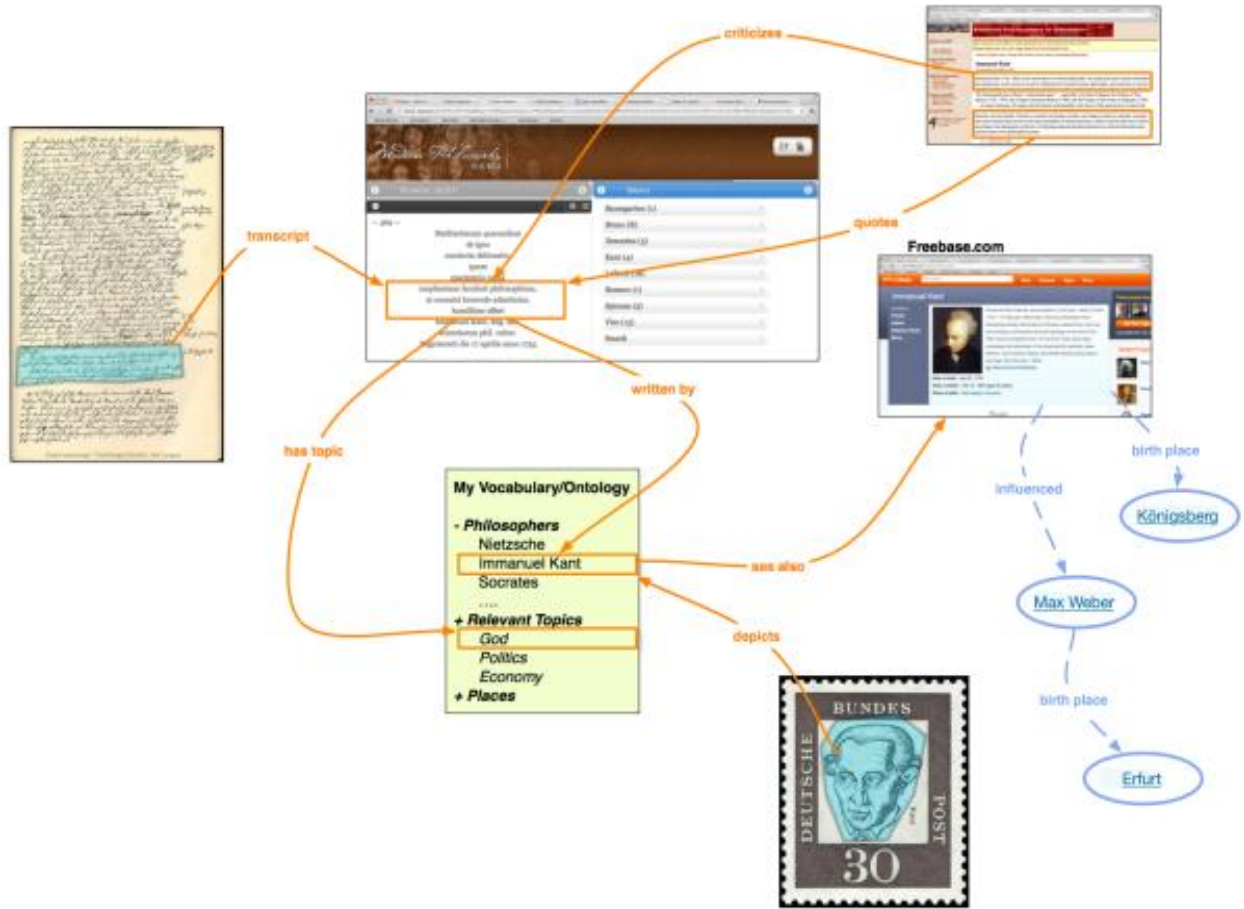


Fig. 3. A semantic graph resulting from annotations of web content made with Pundit, possibly by different users. Text excerpts as well as images (and part of them) are put in relations with concepts from domain vocabularies or with entities from the web of data (e.g. Freebase)

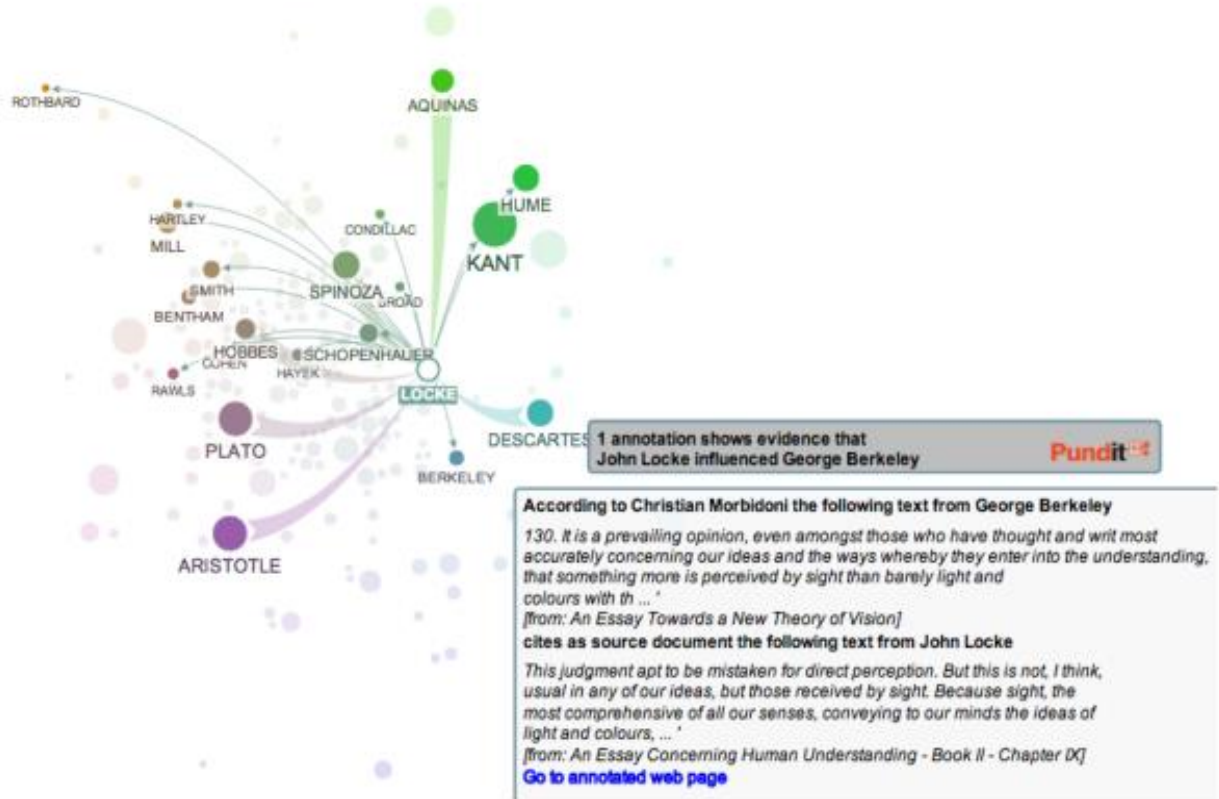


Fig. 4. A graph showing influences paths among philosophers. Edges in the graphs are automatically created from annotations that link two works using a "cites" relation. Exploring the graph a student can make sense of the influence relation by looking at the annotations and finally at the original works of the authors involved. This demonstrative application is powered by Edgemaps and by Wikisource.org

4.1.4. Layer 4: investigating novel open review methods

To avoid the pitfalls of the closed access model and its lairs of learning, as described in section 3, we have to make the selection and evaluation processes and their resulting data as much as possible public and open¹⁶. They should be public, because the eyes of everyone are the best counterbalance to the secret power of private aggregators; they should be open, because, in the field of humanities, the evaluation of a paper can be used to build semantic trails.

¹⁶ [Priem 2013] foresees a system in which journal and article will be superseded by algorithms that filter, rate and disseminate scholarship "as it happens". Such an evolution threatens to create new oligopolies, if the data involved in the collective evaluation are not open

5. Methodology, in a nutshell

In a social software project, the methodology overlaps with the philosophy of the design choices put to the test. It can be summed up in seven points.

1. Interdisciplinary. If we share, after the serials crisis experience, the Platonic assumption [Cerri 2007] that the way of publishing our research is part of the research itself, we need to cooperate with computer scientists in re-designing existing applications and taking back the web as an open environment of discussion and sharing. We should stop being simple generators of content which is shaped by containers designed and owned by others: we should also have a say on the containers themselves.
2. Layered participation. A federal journal should have a layered structure that is very simple at its bottom and very complex at its top. In this way users can be encouraged to participate starting from what they already know – e.g. their browser, or their CMS of choice – and can grow in wisdom and grace as far as they find the further functions of the system useful in their everyday research work. No one should be forced to experience painful installation and configuration procedures for the sake of ideals requiring uncommon levels of web literacy and awareness.
3. Application neutrality. We would like to avoid to balkanize the web by generating further bubbles. The design of a layered structure is not only for the sake of simplicity, but also for the additional goal of application neutrality. No one should be forced to leave the open web and to enter in any walled garden. Building yet another aspirant Facebook for scholars is philosophically different from contributing to the open web that was invented by a researcher, for researchers.
4. Federalism. A federalist approach is the best way to honor the application neutrality principle, as far as regards the containers, and to respect the traditional pluralism of human and social sciences, as far as regards the contents. Politically federalism was devised as a way to make peoples cooperate with-out building incumbent – imperial – positions. As social software is the experimental wing of political philosophy, we might well put political ideas to the test, in the very design of it.
5. Open access/open data. It is not only an EU mandate. It is not only the condition of possibility for the public use of reason. It is also and above all the only way to access to data and their metadata so that can be mined and elaborated by our semantic web tools. As, in the field of humanities and, partially, in the field of social sciences, the data are the same as the texts, we cannot open them without making the texts accessible.
6. Put Semantic Web on the top layer. While semantic web tools are usually designed to separate users from raw data, keeping them in a kind of tutelage made of predigested answers and recommendations, our design choice put semantics on the top of a stack of open layers. Semantics can be employed to make research easier, but everyone remains free to explore the lower layers to check how semantics works – or does not work.
7. Curation. To discover and make visible the singularities that eludes the aggregation, semantic web tools should support - but not replace – the human choice. Therefore, an overlay journal, however sophisticated, should harbor also curation tools to help human beings in selecting and recommending contents – with a human touch.

Someone might object that publishing is not a research topic, because it concerns just publishers, booksellers, librarians or their successors. Moreover, how could we hope to persuade the established scholars to leave their closed access core journals and their oligarchic systems, where they spent their very souls in making themselves acceptable in such an exclusive environment? They might well repeat the words of Protagoras, in Plato's dialogue:

Socrates, he said, I have undertaken in my time many contests of speech, and if I were to do what you demand, and argue just in the way that my opponent demanded, I should not be held superior to anyone nor would Protagoras have made a name among the Greeks [Protagoras: 335a].

Such a Protagorean objection, however, is misplaced. The project does not aim to win the approval of the academic stars. On the contrary, we would like to create, by a bottom-up, federalist approach, an environment so rich and diverse that we could afford to leave them to twinkle alone in their walled gardens, if they do not want to participate in an open discussion¹⁷. Socrates himself, in the Protagoras, is willing to leave the conversation and to go elsewhere. Believing that the current scholarly system is the only thinkable environment in which research can survive and thrive is as biased as believing that people can have a meaningful philosophical discussion only in the house of Callias, son of Hipponicus.

As regards as the opinion that the humble publishing has nothing to do with the noble research, and it is – even in times of media revolution – a menial task, it is worth quoting again Plato. In Gorgias 509a, Socrates says that his thesis are fastened “with reasons of steel and adamant” – but only until someone will be able to refute them. However, if he had enclosed his thesis in a barbed-wire enclosure, a refutation would have become more and more unlikely, both because of the smaller number of people allowed to access them, and for an unwanted - but unavoidable – consequence of exclusion. Closed access publishing and the traditional, anonymous peer review¹⁸ contribute to create and preserve closed oligarchic groups¹⁹ whose inner solidarity hinders or censors any possible refuter or innovator, far from the eyes of the public. Believing that our “published” or – better – privatized papers fall outside our responsibilities means forgetting the communicative and social nature of scholarship and the awareness of its reflexivity, to reduce ourselves to unresponsive academic bots. There are times in which leaving the house of Callias, son of Hipponicus, and helping others to do the same, is the very essence of research.

¹⁷ The change is already happening [Burgelman et al. 2010]: the scholars have only to choose whether to steer it or to suffer it

¹⁸ We should never forget that censorship lies among the historical roots of peer review [Biagioli 2002]. And, as regards as the future, we should dare to recognize that [Ginsparg 2004]'s question: “If we were not burdened with the legacy print system and associated methodology, what system would we design for our scholarly communications infrastructure?” is a major research question

¹⁹ As wrote [Guédon and Siemens 2001], challenging the established system of scholarly publishing and evaluation “changes the means by which distinction is imparted and imparting distinction is a sure sign of power. In other words, those who now hold that privilege are afraid of losing it («gate keepers») and they will every possible argument to protect it without, if possible, ever mentioning it”

References

- ARL, Monograph and Serial Expenditures in ARL Libraries, 1986-2006 <http://www.arl.org/bmdoc/monser06.pdf>.
- S. Basso, A. Servetti, J.C. De Martin (2011), The Hitchhiker's Guide to the Network Neutrality Bot Test Methodology, Congresso Aica, Torino, <http://nexa.polito.it/publications/neubot/2011-aica>.
- T. Berners-Lee (2010), Long Live the Web: A Call for Continued Open Standards and Neutrality, "Scientific American", <http://www.scientificamerican.com/article.cfm?id=long-live-the-web>.
- M. Biagioli (2002), From Book Censorship to Academic Peer review. *Emergences*, 12(1).
- M. Boldrin, D.K. Levine (2008), *Against Intellectual Monopoly*, Cambridge, Cambridge UP <http://levine.sscnet.ucla.edu/general/intellectual/against.htm>.
- B. Brembs, M. Munafò (2013), Deep Impact: Unintended Consequences of Journal Rank, <http://arxiv.org/abs/1301.3748>.
- J.C. Burgelman, D. Osimo, M. Bogdanowicz(2010), Science 2.0 (change will happen...) "First Monday", <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2961/2573>.
- A. Capra (2012), 'Why Plato Wrote': The Insularity of Platonic Studies, *CHS Research Bulletin* <http://wp.chs.harvard.edu/chs-fellows/2012/02/14/why-plato-wrote-the-insularity-of-platonic-studies-2/>.
- G. Cerri (2007), *La poetica di Platone: una teoria della comunicazione*, Lecce, Argo.
- G. Crane(2011), "From Subjects to Citizens in a Global Republic of Letters". In *Going Digital. Evolutionary and Revolutionary Aspects of Digitization*. Ed. by K. Grandin. Nobel Symposium 147. The Nobel Foundation, pp. 251-254 <http://www.center.kva.se/svenska/forskning/NS147Abstracts/KVAGoingDigitalwebb.pdf>.
- A. Dash (2012), *Rebuilding the Web We Lost*, 2012 <http://dashes.com/nil/2012/12/rebuilding-the-web-we-lost.html>.
- P.A. David (2007), The Historical Origins of "Open Science", Stanford, <http://www-siepr.stanford.edu/papers/pdf/06-38.pdf>.
- J.L. De Meulemeester (2012) Quels modèles d'université pour quel type de motivation des acteurs? Une vue évolutionniste, "Pyramides" (21), <http://pyramides.revues.org/804>.

M. Ferraris (2008), "Science of Recording", in H. Hrachovec and A. Pichler (eds.), Philosophy of the Information Society, Proceedings of the 30th International Ludwig Wittgenstein-Symposium in Kirchberg 2007, Frankfurt/a.M., Ontos Verlag,.

K. Fitzpatrick (2011), Planned Obsolescence. Publishing, Technology, and the Future of the Academy. New York. NYU Press.
<http://mediacommons.futureofthebook.org/mcpress/plannedobsolescence/Overview>.

L. Floridi (2006) Peering into the Future of the Infosphere. "TidBits", 1–13
<http://tidbits.com/article/8686>.

P. Galimberti (2011), I dati sulla ricerca. Un problema aperto, "Roars"
<http://www.roars.it/online/i-dati-sulla-ricerca-un-problema-aperto/>.

D. Gillies (2011), The Fundamental Flaw in Research Assessment Systems, "Roars"
<http://www.roars.it/online/the-fundamental-flaw-in-research-assessment-systems/>.

P. Ginsparg (2004), Can Peer Review be better Focused? "Science & Technology Libraries", 22 (3-4), pp- 5–17 <http://people.ccmr.cornell.edu/~ginsparg/blurb/pg02pr.html>.

J.-C. Guédon (2001), In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing, "Arl Proceedings",
<http://www.arl.org/resources/pubs/mmproceedings/138guedon.shtml>.

J.C. Guédon, R. Siemens (2001) The Credibility of Electronic Publishing,
<http://web.viu.ca/hssfc/Final/PeerReview.htm>.

J.C. Guédon (2008), Open Access and the divide between "mainstream" and "peripheral" science,
<http://eprints.rclis.org/handle/10760/10778>.

B. Haslhofer, E. Momeni, M. Gay, R. Simon (2010), Augmenting Europeana Content with Linked Data Resources, 6th International Conference on Semantic Systems (I-Semantics)..

S. Hirschauer, Die Innenwelt des Peer Review. Qualitätszuschreibung und informelle Wissenschaftskommunikation in Fachzeitschriften. Retrieved from
<http://www.sciencepolicystudies.de/dok/expertise-hirschauer.pdf>.

I. Kant, Beantwortung der Frage: Was ist Aufklärung?, Ak VIII
<http://korpora.org/Kant/aa08/033.html>.

D. Kleiner (2010), The Telekommunist Manifesto. Network Notebooks 03, Institute of Network Cultures, Amsterdam, <http://telekommunisten.net/the-telekommunist-manifesto/>.

J. Lanier (2010), You Are Not a Gadget, New York: A.A. Knopf.

L. Lessig (2000), Code is Law, "Harvard Magazine", January-February, <http://harvardmagazine.com/2000/01/code-is-law.html>.

P. Lévy, *Cyberculture*, Minneapolis, Univ. of Minnesota Press, 2001.

M. Luczak-Rösch, R. Heese, A. Paschke (2010), Future Content Authoring, "Nodilities – The Magazine of the Semantic Web", Issue 11.

S. Marche (2012), Literature is not Data: Against Digital Humanities, "Los Angeles Review of Books", October 28th <http://lareviewofbooks.org/print.php?id=1040>.

R. K. Merton (1979), "The Normative Structure of Science" (1942) in id., *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, University of Chicago Press.

A. Molinié, G. Bodenhausen (2010), Bibliometrics as Weapons of Mass Citation. La bibliométrie comme arme de citation massive, "CHIMIA", 64 (1/2), <http://www.chimie.ens.fr/Resonance/bibliometrics1.pdf>.

M. Nucci, M. Grassi, C. Morbidoni, F. Piazza (2012), Enriching Digital Libraries Contents with SemLib Semantic Annotation System, Digital Humanities conference, Hamburg.

E. Pariser (2011), *The Filter Bubble*, New York, The Penguin Press, <http://www.thefilterbubble.com/>.

J. Priem (2013), Beyond the paper, "Nature", 495 (28 March), pp. 437–440 <http://www.nature.com/nature/journal/v495/n7442/full/495437a.html>.

M. Rose (1988), The Author as Proprietor: Donaldson vs. Becket and the Genealogy of Modern Authorship, "Representations" 23, pp. 51–85.

D. Searls, D. Weinberger(2003), *World of Ends. What the Internet Is and How to Stop Mistaking It for Something Else*, <http://www.worldofends.com/>.

V. Uren, P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, E. Motta, and F. Ciravegna (2006), "Semantic annotation for knowledge management: Requirements and a survey of the state of the art", *Web Semantics: Science, Services and Agents on the World Wide Web*, Volume 4 Issue 1.