Designing a Semantic Web Path to e-Science[1]

Francesca Di Donato
didonato@sp.unipi.it

Dipartimento di scienze della politica, Facoltà di scienze politiche,
Università di Pisa, Italy (http://sp.unipi.it/hp/didonato)

Abstract. This paper aims at designing a possible path of convergence between the Open Access and the Semantic Web communities. In section 1, it focuses on the problems that the current Web has to face to become a fully effective research means, with particular regard to the question of selection according to subjective quality criteria. Section 2 exposes the main principles and standards which lie behind the Open Access movement, and tries to demonstrate that the Open Access community is a fertile ground where to experiment Semantic Web technologies. Finally, section 3 sketches a number of practical strategies and suggests the combination of existing tools for e-Science, in order to create a real Semantic Web of scientific knowledge.

Introduction

Nowadays, the benefits provided by the spreading of the Internet and the Web to the scientific research practice are evident even in the Humanities, so that a discussion on these benefits may appear foregone and inessential. As most of the participants in this conference know better than I do, the Web was originated as a universal documentation system, and was conceived in order to connect different pieces of information in a unique global network.

Some of its characteristics encourage philosophers to think of the analogies between the Web and the structure of Science. The fact that the protocols on which the Web is based are open and the Web "lightness" (e.g. being founded on few basic principles: URI, HTTP, HTML) make it a universal system, a "system with common rules that would be acceptable to everyone. This meant as close as possible to no rules at all" [5]. In practice, this particular feature allowed the Web to grow and become the most wide-spread information system ever conceived. Berners-Lee [5] expressly stated that one of the major innovations introduced by the Web, the possibility to link everything to all, is exactly what scientists have done for centuries: "Tables of

---

1 This paper has been accepted at SWAP2005 <www.swap2005.org>. It is available online at the CEUR Online Workshop Proceedings website <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-166/44.pdf>. This work is part of a more general research described in [10]. I am grateful to Maria Chiara Pievatolo for having encouraged turning my efforts both to philosophy and computer science and exploring the interconnections between the two. I thank the core group of the Hyperjournal project for making this work possible by putting in practice what it is possible in theory, especially Michele Barbera, with whom I discussed the different parts of this paper as well as the general idea on which this work is based. Thanks also to Laura Cignoni, for the English version proofreading.

contents, indexes, bibliographies, and reference sections are hypertext links". Tim Berners-Lee anticipated the potential effect of links would have, while foreseeing that "suddenly scientists [could] escape from the sequential organization of each paper and bibliography, to pick and choose a path of references that served their own interest". The practical application of the philosophical principles that identify Science as an inter-connected and freely-accessible knowledge system, caused a rapid development of two fundamental elements of Science: the collaborative work and the degree of connections among elements of knowledge via the usage of citations.

However, there are also important differences between the Web and the structure of Science. Specifically, [4] posed the problem of information selection in accordance to quality criteria. The inventor of the Web identified in selection one of the main problems of the present Web, which has to be faced and solved. He made a distinction between the latter problem and the problem of "access". In fact, if "it is understood that a collection of works, such as a set of technical reports or a library, only includes articles reaching a certain standard, and some early dial-up information services similarly amassed information according to some quality criterion", and "some people miss that with the Web", nonetheless, it is important that "the Web itself doesn't try to promote a single notion of quality", while fostering "to carry … beauty and ugliness, honesty and lie". The fact that the Web holds many different kinds of information, and in particular that the information published on the Web has not been certified by the scientific community, raises a wall for the evolution of e-Science. Nonetheless, Tim Berners-Lee advocated this characteristic as one of the main principles on which the Web is built: being free from any form of centralized control able to prevent people from accessing the Web or adding to it. He clarified that the problem of quality should be solved without breaking an "architectonic"[2] principle of the Web: its universality. A central authority that can judge on the quality of information and its intended audience can be very dangerous; furthermore, we must never forget that the "unimportant notes of today could be at the basis of revolutionary ideas in the future" [4]. Therefore, selection must be performed ex-post; the problem of selection according to quality criteria can be reformulated in terms of how to give the researcher the subjective perception of higher quality, while maintaining an open Web for people whose criteria are different.

1 Discussing Quality

It is worth noticing that, in this way, Berners-Lee brings into question the concept of selection commonly accepted by scientists, an idea and a practice that is deep-rooted in the scientific research process, and that has been in use since the Print claimed it as the research community medium, between the end of XVII and the beginning of XVIII century, as explained and

[2] The use of the term "architectural" in this contest (see [3]) can be compared with those of Immanuel Kant in [20]. For a discussion of the latter, see [9].

discussed in [18]. The main point of Berners-Lee is that access has to precede selection. On the contrary, in the traditional publishing model of Science, selection comes before access. This point is not of little importance. The entire organization of Science, far from being a republic, is constituted and organized according to a consolidated system in which the selection of scientific results has a fundamental role (suffice it to think of the importance of publications for accessing positions in the Academy, career management, fund raising and so on). Therefore, Tim Berners-Lee's vision, if implemented, would be a real Copernican revolution in the domain of Science.

The problem of selecting information by quality criteria is linked to another problem faced by the Web users, especially scientists: on the Web it is difficult to find exactly what we are looking for. The benefits of being able to use powerful search engines such as Google are unquestionable. However, they do not help to find relevant information (in a strict meaning). Just adding more and more scientific content into the mass of searchable items, does not help either. Why is it so difficult to find relevant information? How can we face the problem?

As you know, the Semantic Web answers both questions, and provides the practical solutions to solve them. in brief, the problem is that the data on the Web continue to be non machine-readable; this problem can be faced by making a further distinction between "documents" and "data", the former being devoted to the human use, the latter to machine treatment. Before this problem can be solved, a pre-condition must be met: "For an international hypertext to be worthwhile, of course, many people would have to post information. The physicist would not find much on quarks, nor the art student on van Gogh, if many people and organizations did not make their information available in the first place" [5]. In the following section I will show the great improvements produced in this direction by the Open Access movement [36], whose tools are based on the Web and the Internet as well as on the free/open-source software philosophy and methodology. Furthermore, we will see that a solution to this point can lead, in a short time, to challenge the causal-relation direction between selection and access in the e-Science community, thus making a Semantic Web of scientific knowledge possible.

2 Open Access to Scientific Knowledge

In the process of Science, collaboration seems to prevail on tendencies that aim at restricting access to knowledge. Guédon [15] describes the coexistence of this tension in scientists with a metaphor that compares the researcher with the schizophrenic personality of Stevenson's characters, Dr Jekyll and Mr. Hyde. A researcher as an author can be seen as a "Dr. Jekyll"; as a reader, he becomes "Mr. Hyde". The latter is mostly interested in communicating with her/his colleagues: (s)he wants to collect all useful information for her/his research, including informal messages and news; the former, who is involved in the production of scientific results and their publication, becomes one of the great paladins of property rights, and therefore (s)he is mainly concerned with citing the most relevant sources. However, the republic of Science, which nowadays includes the free-software as well as the Open Access movement, shows a common principle, known as “distributed intelligence”. Moreover, thanks to the World Wide Web, the communication and publication phases, which had been separated by print, have been reconnected; and so the schizophrenic aspects of scientists can be conciliated.

Specifically, the e-Science movement was originated in 1991 and since then it has developed widely. This movement gave rise to a sheer amount of initiatives as well as to a set of standards subscribed by the scientific community and its stakeholders. All initiatives, standards and official declarations are now included under the umbrella of “Open Access”.

2.1 OAI

Since the creation of ArXiv, the high-energy physics repository installed [1] by Paul Ginsparg in 1991, which has now become an irreplaceable tool for physicists, has gone a very long way. In order to understand the significance of the innovation that took place in Los Alamos (arXiv is now run by Cornell), it is useful to briefly outline the history of Open Access. This will lead to a definition of Open Access. This history is of interest to my discourse because the spread of the Open Archives network and the Open Access movement makes it possible to climb the Semantic Web “layer cake” stair and therefore to experiment tools which, building on the lower layers, will make effective use of inference and trust. I will also discuss the recent technical evolutions that came to light during the OAI4 workshop [28] held in Geneva in October 2005. The goal of this discussion will be to put forward a possible path of convergence between Open Access and the Semantic Web communities.

I will briefly summarize the history between 1991-1999, claiming that the Los Alamos archive did not go unnoticed. On the contrary, similar initiatives in several institutions and in different scientific domains sprang up and proliferated. The rise of experiences like the one of Paul Ginsparg entangled researchers as well as librarians and computer scientists. The adoption of a interdisciplinary perspective led to an expansion of the movement, the need for a federative action being the result of the quick growth of open archives. The 1999 meeting in Santa Fe, known as the “Santa Fe Convention” [35] drove to design a technical and organizational framework in order to simplify the discovery of information archived in the distributed network of e-print repositories. The spirit that presided over the Santa Fe initiative is reminiscent of the spirit that conducted to the birth of the Web: optimize the ratio between results and technical constraints while laboring to keep the constraints as low as possible. The objective was to obtain some elements of interoperability that would allow any registered archive to be easily harvested through a common search instrument, without creating a new information system. This goal has been achieved. Librarians introduced the use of elements compatible with the OCLC ontology Dublin Core Metadata Set, and the results have been conveyed in the Open Archives Initiative (OAI)[30]. The purpose of OAI is well described in its mission statement: “The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. [...] The fundamental technological framework and standards that are developing to support this work are, however, independent of both the type of content offered and the economic mechanisms surrounding that content, and promise to have much broader relevance in opening up access to a range of digital materials. As a result, the Open Archives Initiative is currently an organization and an effort explicitly in transition, and is committed to exploring and enabling this new and broader range of applications. As we gain greater knowledge of the scope of applicability of the underlying technology and standards being developed, and begin to understand the structure and culture of the various adopter communities, we expect that we will have to make continued evolutionary changes to both the mission and organization of the Open Archives Initiative” [32].

The birth of the protocol for metadata harvesting (OAI-PMH) [31] in 2001, which is today adopted by a wide federation of archives and journals, turned the early intentions into a well-established reality. The OAI-PMH protocol defines a metadata-negotiation interface between a data provider and a service provider. The OAI-PMH protocol allows to transfer metadata over HTTP. The protocol has a low-entry approach: it defines a minimal metadata set to be conveyed by a provider so that the provider can be OAI-PMH compliant; the minimum set of metadata is defined by the Unqualified Dublin Core Element Set. The message is encoded in XML. Once the compatibility is met, the protocol encourages data providers to expose metadata in more expressive and complete schemas than the Unqualified Dublin Core. This latter characteristic makes the protocol extensible.

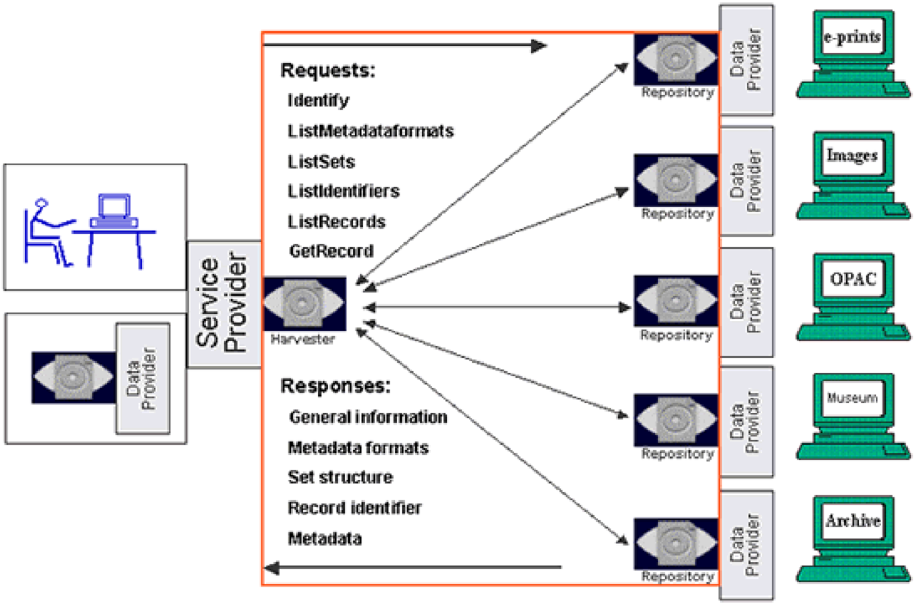

Fig. Sequenza Fig.. OAI-PMH structure-model

In this paper, I will focus on the nature of the transported data rather than on the low-level details of the protocol. The OAI-PMH has already demonstrated a high degree of reliability and scalability.

2.2 The spread of Open Access

Since 2002, the OAI-PMH is now at its 2.0 version, which has become a stable interoperability framework [22]. Moreover, the critical mass of content deposited in the open repositories during the last four years has allowed Open Access to take great strides. I will outline the most important Open Access definition, which has been signed by many academic institutions. I will then expose what is brewing up from the technical point of view.

The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities [2] (2003) is based on the fact that "new possibilities of knowledge dissemination not only through the classical form but also and increasingly through the open access paradigm via the Internet have to be supported". Open Access is there defined as "a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community". It is the open space of e-Science. Moreover, the declaration states that "the future Web has to be sustainable, interactive, and transparent, in order to realize the vision of a global and accessible representation of knowledge". Therefore, both content and software tools must be openly accessible and compatible. This statement implies the adoption of adequate licensing policies. In fact, law acts as a regulator as well as a software code while permitting or denying access to scientific resources

[23]. Therefore, Open Source Software and Open content need to be released under open licenses (such as GNU/GPL, BSD or Creative Commons). This is a necessary condition to have a critical mass of documents, and to treat them. The Berlin Declaration also provides a definition of Open Access Contribution, which includes “original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material”. An Open Access contribution must satisfy two conditions: the first concerns copyright issues; the second states that “a complete version of the work and all supplemental materials, including a copy of the permission [read: license], in an appropriate standard electronic format is deposited (and thus published) in at least one online repository using suitable technical standards (such as the Open Archive definitions [read: OAI-PMH]) that is supported and maintained by an academic institution, scholarly society, government agency, or other well established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving.” Note that open archiving grants access to documents. Here access goes before selection, the latter being to the responsibility of scientific journals, which perform peer reviewing.

In November 2005, the Berlin Declaration was signed by several institutions (including 63 out of the 75 Italian universities). These institutions officially pledged themselves to achieve Open Access to Scientific Knowledge. There is a growing interest towards alternatives to the traditional publishing paradigm. In fact, Open access provides a solution to a well-known problem for university Administrators and Librarians: the “serial price crises”[3], which is outside the scope of this work. However, the adoption of copyleft licenses is growing and these strategies are beginning to bear fruit. On September 2005, UNESCO adopted Amendments to the Draft Programme and Budget for 2006-2007 that have the effect of endorsing Open Access [39]. Coherently with its principles, all information and sources concerning Open Access declarations, resolutions and initiatives are freely accessible on the Web.

Nowadays [28], research groups located in Los Alamos and Cornell which are leading the development of standards for interoperability affirm that, on this basis, it is possible to develop new tools for e-research and e-publishing: “Pre-print repositories have seen a significant increase in use over the past fifteen years across multiple research domains. Researchers are beginning to develop applications capable of using these repositories to assist the scientific community above and beyond the pure dissemination of information.” [34]

3 Designing a Semantic Web path to e-Science

There are several experiences devoted to the construction of tools using OAI-PMH compliant repositories to assist the scientific community above and beyond the pure dissemination of information. In this part I will summarize

3 On the “serial price crisis”, see [15].

and discuss some tools based on OAI, which have been proposed by Peter Murray-Rust, Marko Rodriguez, Johan Bollen and Herbert Van de Sompel. These tools can serve as a source of inspiration for the OAI and the Semantic Web communities in order to implement a Semantic Web of Science.

3.1 Extraction of Hidden Semantics from Scientific Journal Articles

Let me remind that the communication model adopted by the scientific community is strongly focused on publications. On-line digital publication is becoming more and more common; nevertheless, the assessment of scientific production depends on quantitative criteria relying on the concepts of "journal" and of "article". As a matter of fact, the article is the core object on which the whole system in based (specifically for fund raising and career management). Therefore, e-publishing models are still, more or less, a mere reproduction of the traditional publishing paradigm.

In some scientific domains (especially in STM) raw data are firstly obtained from experiments, then processed by researchers and, finally, turned into articles, which are published in journals. Raw data, which are normally at the basis of the results published in article form, are not published and remain in laboratory hard-disks. But primary sources are a necessary complement to literature circulating under the form of textual articles, and the limited availability of raw data is an obstacle to Science development. In fact, it makes the basic practice of Science (i.e. starting from the existing results in order to produce new scientific one) really arduous and complex. The scientific progress has its roots in knowledge reuse. The raw data (or primary sources), which can be used by machines, are the foundations on which researchers can extract and derive new, unanticipated information and knowledge. Thus, limitations in the re-usability of data (such as the ones deriving from semantic depriving which takes place during the traditional publication process), bounds the evolutionary power of knowledge. This argument aims at showing the inappropriateness of the document (intended as human-readable information) in order to convey information. The problem is that during the process, which starts from the experiment and ends in the article, a lot of information is lost. This happens both in the traditional and in the digital publishing process.

There are two possible ways to overcome the semantics loss, due to the use of documents as the unique form of conveying scientific information: the first is to publish documents which are both human- and machine-readable; the second consists in extracting hidden semantics contained in human-readable documents. While the former strategy eradicates the problem at its root, the latter is a mere workaround. In any case, the latter strategy does not only permit to subvert completely the basis of the current scientific communication system. Thus, the extraction of hidden semantics can be considered as a first move towards the evolution of a communication system that was conceived in XVIII century. That system did not change for centuries, but has now become unable to respond to the changed needs of the scientific community.

Murray-Rust [25] shows how to use the extraction of hidden semantics inside the Chemistry community. NesC [26] and Oscar [33] are text analyzers applying heuristics in order to identify text portions whose semantics can be coupled to a domain ontology. Such ontology, although not formalized, is widely subscribed inside the Chemistry community (molecule names, chemical reactions, and their properties). Once the text-portions have been identified and made unambiguous, they are assigned a unique identifier [37]. Furthermore, Oscar e NesC are able to apply reasoning rules and to perform actions such as either correcting errors present in the document which have not been found by the (human) reviewer, or using identifiers to collect more metadata from different sources on the Web and visualize them to the reader. Unfortunately, both tools lack a function, allowing to serialize metadata in RDF and then to re-publish the extracted semantics. We are now not far from Bush's predictions while anticipating the birth of a “new profession of trail blazers, those who find delight in the task of establishing useful trails through the enormous mass of the common record” [8]. According to Bush's vision, the inheritance from the master would become the entire scaffolding by which they were erected. In Bush's example, “the chemist, struggling with the synthesis of an organic compound, has all the chemical literature before him in his laboratory, with trails following the analogies of compounds, and side trails to their physical and chemical behavior”. As a matter of fact, as it is possible to demonstrate starting from Murray-Rust’s experience, nowadays we are able to go beyond the possibilities anticipated by Bush’s vision. While having at disposal a function, which makes the data extracted with NesC and Oscar expressible in RDF, such tools could allow to embed explicit machine-readable semantics inside human-readable documents. This can be done even after the document has been published, using the RDF text-encoding model exposed by Tummarello and Morbidoni [38].

Similar systems can be initially implemented for all those disciplines where domain ontologies can be formalized easily. Then, we could start implementing similar tools for the Humanities[4], a research domain where the fuzzy nature of the semantics and a missing agreement on ontologies make this enterprise much more arduous.

3.2 On Quality, again. Suggesting a Convergent Path between Semantic Web and Open Access Communities

---

[4] In digital Humanities, primary sources are represented by reproduction of manuscripts, works of art or craft. The secondary sources represented by translations, interpretations, editions and similar works play a twofold role within the Humanities: they are, as usual, the preferred way of disseminating knowledge but they also often take the place of primary sources (e.g. The critique of a philosophical text is written on the basis of a translation instead of on the basis of the original work). As much of the research in the Humanities originally lacks connections with its original primary sources, the ability of identifying links a posteriori becomes even more important and can highlight mistakes in the formulated theories as well as lead to new unanticipated discoveries.

The core issue of the works discussed in this paper is selection, i.e. evaluation of Science. Rodriguez [34] sketches a deconstructed publishing model in which peer-reviewing is mediated by a OAI-PMH service. A social network is used to select potential reviewers to a manuscript and to weight the influence on each reviewer participating in the evaluation. In [34], peer reviewing is performed after a paper has been archived in an open repository. The assessment system that is based on the concept of Impact Factor, in particular, is more frequently recognized as an obstacle to the universality of the scientific communication paradigm. IF is one of the possible impact measures, not necessary the most representative [12; 16; 24; 27]. In fact, IF is calculated on an arbitrary set of scientific journals (named "core journals") which do not represent the global scientific production [14, 17]. Building on these premises, Bollen [7] suggests to consider an impact factor which a) is to be calculated from both usage data and citation data (not only from the latter); b) would be calculated by metrics based on the topological structure of the existing journal network (of citations and usage, i.e. downloading). These metrics, known as "social network analysis", allow to obtain more significant impact measures, which consider the topological position of a journal inside the network instead of being based only on the number of citations received. After having noticed the stakeholders' interest in these new models, and once these models have been experimented inside a close organization such as Los Alamos National Laboratories (LANL), it will be extremely interesting to apply those metrics in a wider context whose boundaries are not defined a priori. According to this objective, Semantic Web tools such as RDF and OWL seem to be the most natural solutions. It is outside the scope of this work to discuss the nature of the connections between social networks, the Semantic Web and trust, these issues being fully exposed by Golbeck [13]. It suffices to say that a combination of these elements in a unique methodology can lead to a complete revolution of the existing communication process of scientific publication. It can pave the way to a semantic Web of Science, able to provide scientists with high-quality contents without enforcing a single quality notion.

The experiences and proposals outlined here lead to identify the intents of the Open Access community as a specific sub-set of those of the Semantic Web. Unfortunately, although there is compatibility between the architectonic models proposed by the two communities, there are also technological incompatibilities, which could be easily solved by increasing the degree of interaction between the two communities. Projects such as World Wide Molecular Matrix [40] and HyperJournal [18] could create a bridge between the two communities, thus demonstrating in practice the synergies deriving from the convergence of Semantic web and Open Access communities. In parenthesis, it is appropriate to add that Tim Berners-Lee's attention on Open Access in the last months has let us hope in a possible convergence of both communities towards a common strategy.

HyperJournal is today the only open source journal publishing application, which is based on RDF storage and using RDF encoded metadata. Besides being compliant with the OAI-PMH specifications, HyperJournal also offers different and semantically richer means for expressing metadata. These extended semantics annotations are obtained using RDF to encode metadata and using ontologies (expressed in OWL) such as FOAF, DC and SKOS. The employment of RDF instead of plain XML helps to overcome at least one order of problems: firstly, XML and its Schemas only enable to express loose constraints; this is the reason why service providers are often forced to treat the harvested data in a “data provider specific” fashion. Secondly, a conceptual encoding tree can correspond to many XML serializations; this means that, independent of the degree of detail given in a Schema, XML does not discourage ambiguous semantic description. Although the efforts of the OAI community to increase interoperability between digital archives have been extremely fruitful, they presently face the risk of remaining bound into a “close” system. If they do not join the Semantic Web, the potential of applying trust metrics and inference rules on a virtually unlimited dataset (the whole Web), will be lost. Hence, even if the usage of non OAI complaint interfaces is not a good practice, experimenting different approaches could help to identify the limitations of the current OAI-PMH, especially those deriving from the usage of XML to encode metadata. For these reasons, the future steps of HyperJournal development will be geared towards the inclusion of tools similar to those presented in this section.

Conclusions

The Open Access community has made it possible to invert the direction of the causal-relation between selection and access, which has been a characteristic of the scientific communication model for three centuries. The achievement of this fundamental result now permits to rethink the concept of selection in itself. The practical experimentation of new models for selection (e.g. the development of tools such as those described before) paves the way of a convergent path between the Open Access and the Semantic Web communities. OAI already provides a sheer amount of data, and in e-Science the demand of functions typical of the Semantic Web, such as reasoning and trust, is high. Therefore, Open Access and in a broader sense the whole Scholar community can be an excellent environment in which to begin the development of a real Semantic Web of Science. The ability of the user to combine different selection perspectives, and the possibility to apply criteria of others (either individuals or groups), may then cause a user-driven revolution, a Copernican revolution for e-Science.

References

1. ArXiv.org e-Print archive. <http://arxiv.org/>
2. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.

<http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html> (2003)

3. Berners-Lee, T.: Design Issues. Architectural and philosophical points. <http://www.w3.org/DesignIssues/>(1990-2005)
4. Berners-Lee, T.: The World Wide Web - Past, Present and Future. Japan Prize 2002, Commemorative lecture. <http://www.w3.org/2002/04/Japan/Lecture.html> (2002)
5. Berners-Lee, T.: Weaving the Web. The original design and ultimate destiny of the World Wide Web by its inventor, Harper, San Francisco (1999)
6. Tim Berners-Lee on Open Access. See: <http://www.earlham.edu/~peters/fos/2005_08_28_fosblogarchive.html> (2005)
7. Bollen, J., Van de Sompel, H., Smith, J., Luce, R.: Towards alternative metrics of journal impact: a comparison of dowload and citation data. Information Processing & Management. Preprint at <http://arxiv.org/abs/cs.DL/0503007> (2005).
8. Bush, V.: As we may think. The Atlantic Monthly. 176 1 (1945) 101-108. <http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush.shtml>
9. Di Donato, F.: Conoscenza e pubblicità del sapere. Le condizioni della repubblica scientifica a partire dall'Architettonica della ragion pura di Kant, Bollettino telematico di filosofia politica. On-line journal of political philosophy. <http://bfp.sp.unipi.it/dida/arch/>(2005)
10. Di Donato, F.: Conoscenza e scienza. Le condizioni della repubblica scientifica. Una pista di ricerca in evoluzione. <http://bfp.sp.unipi.it/~didonato/ovre/>
11. Dublin Core Metadata Element Set (2005). <http://dublincore.org/documents/dces/>
12. Egghe, L.: Mathematical relations between impact factors and average number of citations. Information Processing and Management. 24-5 (1998) 567–576
13. Golbeck, J.: Computing and Applying Trust in Web-Based Social Networks, Ph.D. Dissertation from the University of Maryland, College Park. <http://trust.mindswap.org/papers/GolbeckDissertation.pdf> (2005)
14. Groote, S. L. D., & Dorsch, J. L.: Online journals: impact on print journal usage. Bul l. Med. Libr. Assoc., 89-4 (2001)
15. Guédon, J-C: In Oldenburg Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing, ARL Proceedings. <http://www.arl.org/arl/proceedings/138/guedon.html> (2001)
16. Harnad, S., Carr, L., Brody, T., & Oppenheim, C.: Mandated online RAE CVs linked to university eprint archives: Enhancing UK research impact and assessment. Ariadne 35 (2003)
17. Harter, S. P., & Nisonger, T. E.: ISI's impact factor as misnomer: a proposed new measure to assess journal impact. Journal of the American Society for Information Science, 48-12 (1997) 1146-1148.
18. HyperJournal. Open Publishing Made Easy! <http://www.hjournal.org>

19. Johns, A.: The Nature of the Book: Print and Knowledge in the Making. University of Chicago Press. Chicago (1998)
20. Kant I. : Architektonic der reinen Vernunft, in Kritik der reinen Vernunft. <http://www.gutenberg.org/etext/6342>
21. Lagoze, C., Van de Sompel, H: The Open Archives Initiative: Building a low-barrier interoperability framework (2001). Draft of accepted submission to JCDL2001.
22. Lagoze, C., Van de Sompel, H., Nelson, M., Warner, S.: The Open Archives Initiative Protocol for Metadata Harvesting, v.2.0 (2002) <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
23. Lessig, L.: Code and other laws of cyberspace. <http://www.code-is-law.org/> (1999-2005)
24. Lewison, G.: Researchers and users perceptions of the relative standing of biomedical papers in different journals. Scientometrics 2 (2002) 229–240
25. Murray-Rust, P.: Open Access and the Chemical Semantic Web, ACS Spring Mtg. <http://wwmm.ch.cam.ac.uk/presentations/acs2005> (2005)
26. Name enchancement system for Chemicals (NesC). <http://wwmm.ch.cam.ac.uk/nesc2004/webintro/authors.php>
27. Nederhof, A. J., Luwel, M., Moed, H. F.: Assessing the quality of scholarly journals in linguistics: An alternative to citation-based journal impact factors. Scientometrics, 51 1 (2001) 241 – 265.
28. OAI4: Innovations in Scholarly Communication. Conference Web site. <http://indico.cern.ch/conferenceDisplay.py?confId=0514>
29. OECD, Working Party on the information Economy - digital Broadband Content: Scientific Publishing. <http://www.lessig.org/blog/archives/oecd_sci.pdf> (2005)
30. Open Archives Initiative.<http://www.openarchives.org/>
31. Open Archives Initiative – PMH. <http://www.openarchives.org/OAI/openarchivesprotocol.html>
32. Open Archives Initiative - Mission Statement. <http://www.openarchives.org/organization/index.html>
33. Oscar text extraction. <http://www.w3.org/2004/10/swls/Murray-Rust/communal/oscar.html>
34. Rodriguez, M., Bollen, J., Van de Sompel, H.: The convergence of Digital Libraries and the Peer-Review Process. Preprint at <http://arxiv.org/abs/cs.DL/0504084>. Accepted submission to the Journal of Information Science (in press). (2005)
35. The Santa Fe Convention. <http://www.openarchives.org/sfc/sfc_entry.htm>
36. Suber, P.: Open Access Overview <http://www.earlham.edu/~peters/fos/overview.htm>
37. Stein, S.E., Heller, S.R., Tchekhovskoi, D.V.: "Toward the development of a standard chemical identifier". Abstracts of Papers, 222nd ACS National Meeting, Chicago, IL, United States, August 26-30, 2001, CINF-005.(2001)

38. Tummarello, G., Morbidoni, C., Pierazzo, E.: Toward textual encoding based on RDF. ICCC 9th International Conference on Electronic Publishing. Katholieke Universiteit Leuven, Belgium, 8-10 June 2005.<http://giovanni.ea.unian.it/semanticweb/submissions/ELPUB2005/rdftef.pdf>.(2005)

39. UNESCO: Amendments to the Draft Programme and Budget for 2006-2007.
<http://unesdoc.unesco.org/images/0014/001408/140864e.pdf>

40. World Wide Molecular Matrix (WWMM). <http://wwmm.ch.cam.ac.uk/wwmm.html>