Web 2.0 and Scientific Publishing: A Survey*

Diego Ponte

Department of Computer and Management Sciences

University of Trento

Judith Simon

Institut Jean Nicod

Ecole Normale Supérieure

Paris

Abstract

This paper presents the first results of a survey which was created to gauge the potential acceptance of a collaborative and Web 2.0 inspired production and dissemination of scientific publications by different scientific communities. The rationale of the survey lies in the fact that while both technological opportunities (e.g. Web 2.0) and new publishing proposals (e.g. open access) have given birth to a multitude of innovative initiatives in the sector, there is still little empirical evidence of the acceptance and use of such new initiatives by researchers. We received 349 completed questionnaires from researchers of many different disciplines. The results of the survey show that there is a strong positive attitude towards Web 2.0 and open publishing approaches. Respondents use a variety of Web 2.0-inspired tools in their work and are willing to even increase this usage. Moreover, there is a strong dissatisfaction with current standard copyright policies in scientific publishing, which seem to demand for novel, more permissive copyright solutions. In particular, the vast majority of respondent does a) not want to transfer copyrights to publishers and b) make their content freely available. However, the major challenge still resides in combining free dissemination of results with robust and reliable quality control mechanisms. Further interesting insights concerned the potential acceptance of novel ways of attributing authorship as well as the perceived pros and cons of different types of peer review.

* The authors acknowledge the financial support of the EU-funded project Liquidpub – Liquid Publications: Scientific Publications meet the Web (http://project.liquidpub.org; FET-Open grant number: 213360) and the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission. The authors gratefully thank the Liquidpub team for their help and support in conducting the survey.

The authors can be contacted at: diego.ponte@unitn.it and judith.simon@univie.ac.at

INTRODUCTION

Over the last two decades, the field of scientific publishing has undergone a period of change thanks to both technological (Internet and Web 2.0) as well as governance reasons (new publishing forms such as the Open Access). From the first point of view, the sector is experiencing the birth of a burgeoning number of innovative initiatives and tools that should support the collaborative and open creation and dissemination of scientific knowledge. From the second point of view, actors such as the Open Access movement call for a different and more open configuration of the sector while criticizing restrictive access rights to scientific knowledge. In this paper we show the results of a survey which aimed at exploring the habits of researchers and their beliefs regarding the many innovations happening.

METHOD AND DISSEMINATION

The survey consists of six sections. The first section covered general information regarding the respondent's professional background, while the other five sections are directly related to main topics and areas of the publishing industry. The second section covered the usage of Web 2.0 tools and authorship issues, while the third section addressed credit attribution and reputation assessment. The forth section explored the beliefs of researchers about different forms of peer review. The fifth section investigated about publishing models, such as the Gold and Green Open Access. The final section explored copyright and dissemination issues.

The dissemination strategy followed several directions. The first consisted in advertising the survey on the European R&D Syndicated Newsroom which is a newsletter bringing together news related to European science, research policies and activities published on external websites[1]. The newsletter gets published widely across disciplines. Moreover, the survey was advertized in an article that our research team prepared for "ICT Results", which was also widely distributed[2]. The research team further advertised the survey on the project website, as well as via our blog and Twitter account. Finally and most importantly, the survey was

[1] http://cordis.europa.eu/newsfeeds/syndicated-newsroom.cfm#

[2] Please find the note here: http://cordis.europa.eu/ictresults/index.cfm?section=news&tpl=article&BrowsingType=Features&ID=91404

announced on a number of open academic mailing-lists covering a diversity of disciplines[3]. The survey was open for participation from May to August 2010. During this time a total number of 689 people took part in our survey, 349 surveys (~50%) were completely filled-out. Even if a large number of the incomplete questionnaires were only lacking few questions, we base our analysis for this paper on the fully completed questionnaires only.

ANALYSIS OF THE RESULTS

In this section we briefly portray the findings of the survey per each of the section composing it. The **first section** of the survey asks for general questions about the respondents. Most of the respondents were Senior Researchers/Professors (41.2%) followed by PhD Students (25.8%) and Junior Researchers (23.2%) (Figure 1).

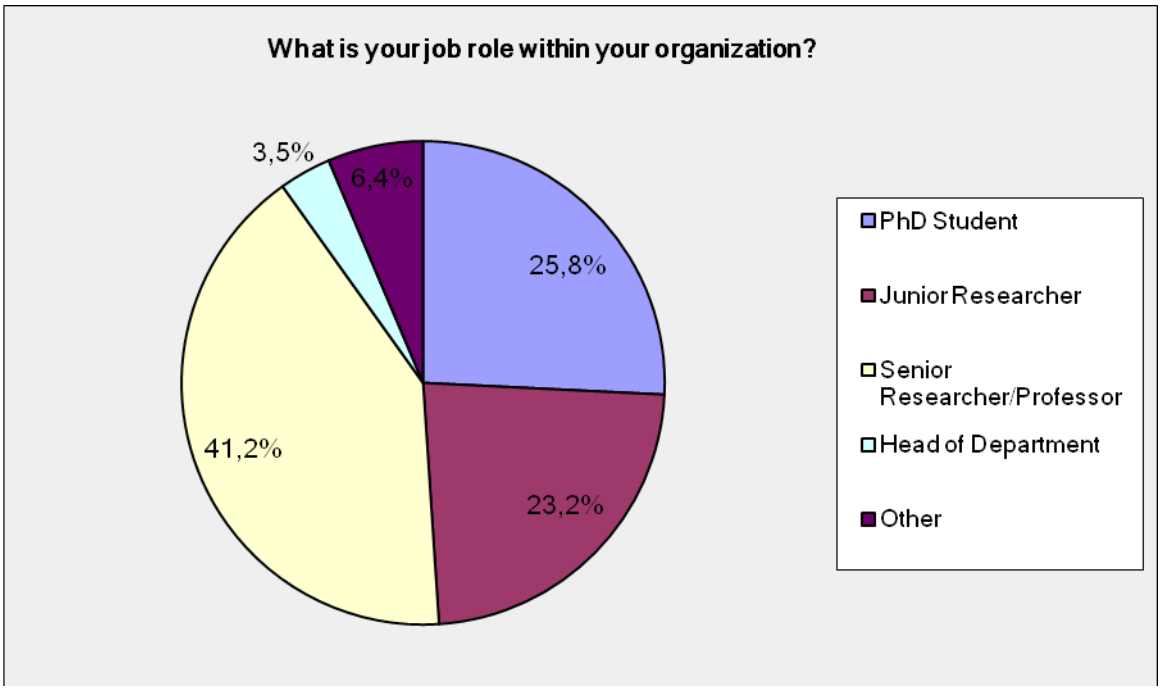

Figure 1: Job position of respondents

When focusing on the age of the respondents, the majority of them is younger than 41 years old (Figure 2).

[3] Inetbib, Liblicense, Communia, Philosop, Philos-L, Simsoc, Nta, Wiss-Org, Iacap, Dbworld, Air-L.

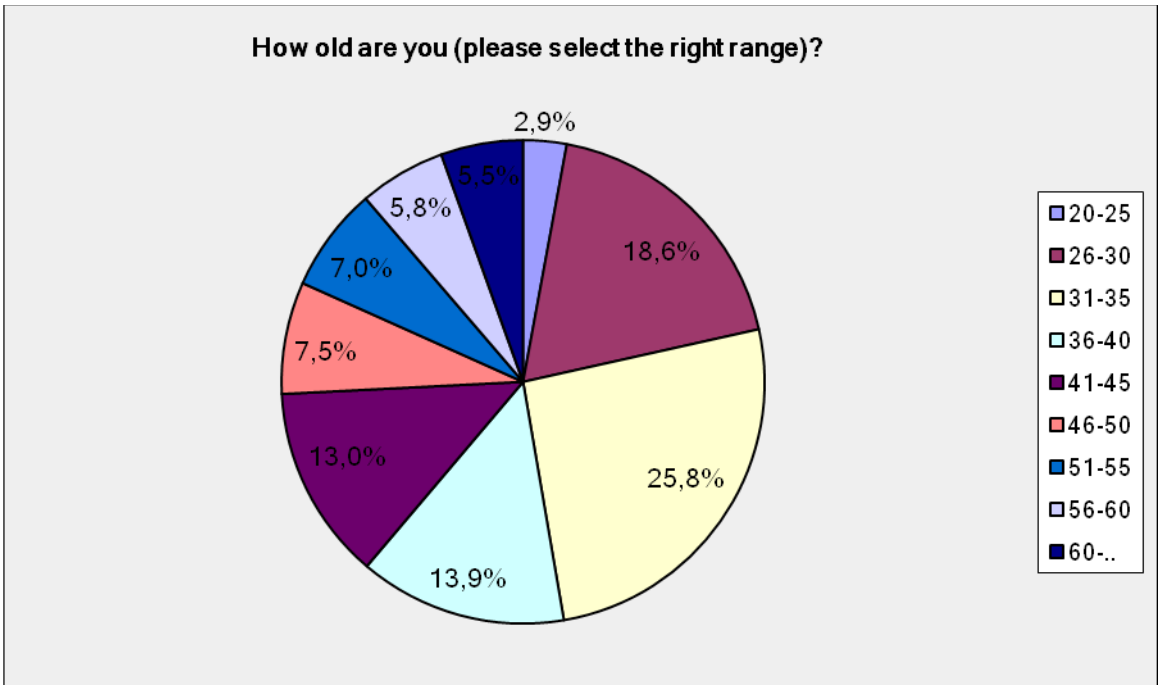

Figure 2: Age range of respondents

Fortunately and in all probability due to the diversity of the mailing lists we used for dissemination, we were able to attract respondents from a heterogeneous set of disciplines. People were asked to select one or more representative research area(s). The most selected areas were social sciences (44.6%) and computer science (47.5%) (Figure 3).

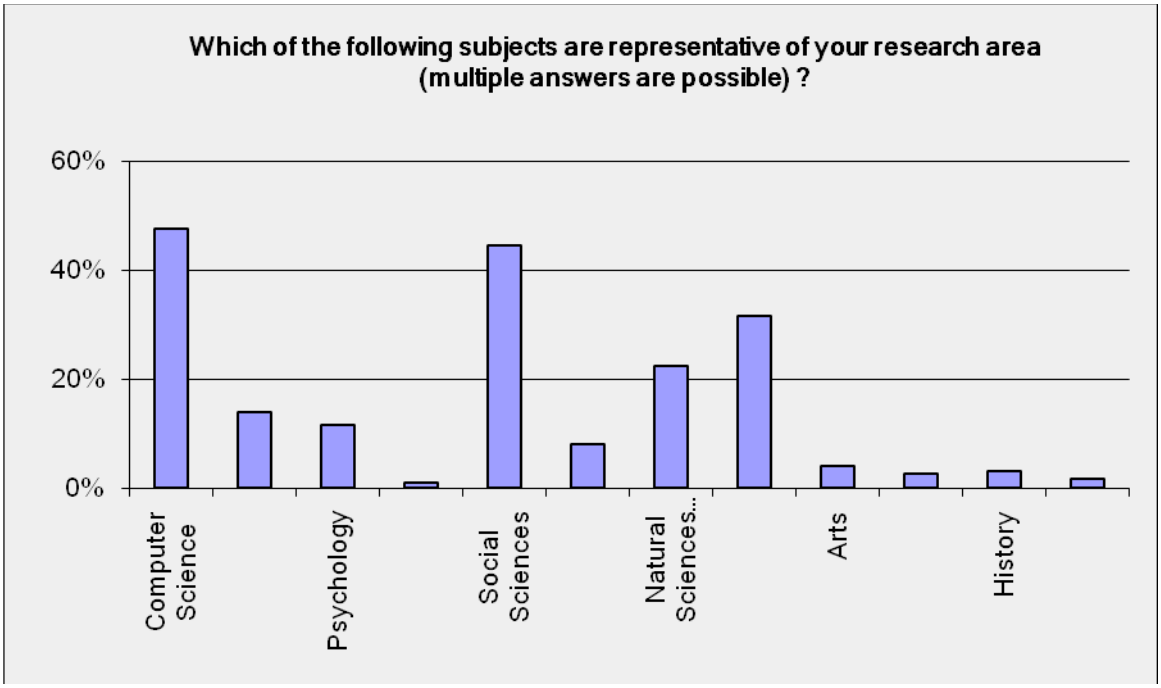

Figure 3: Research fields of respondents

In the **second section** we started by asking about the use of Web 2.0 tools and the collaborative creation of scientific knowledge. Starting from the recognition that web applications which aim at facilitating collaborative knowledge creation and sharing (usually referred as Web 2.0) have been growing fast in the last years; the first question asked whether researchers use such tools for research purposes. Responses show that almost all researchers (99.7%) use search engines (e.g. Google Scholar) and more than half of them (56.5%) also uses citation indexing initiatives (eg. Web of Science). From the Web 2.0 inspired tools, wikis (e.g. ScienceWikia) (42%), blogs (e.g. Science Blog) (38.6%), and social networks (e.g. Nature Network) (34.8%) are used by more than a third of all respondents. Social bookmarking (e.g. CiteULike) (25.8%) and micro-blogging (e.g. Twitter) (17.7%) are used to a lesser degree (Figure 4).

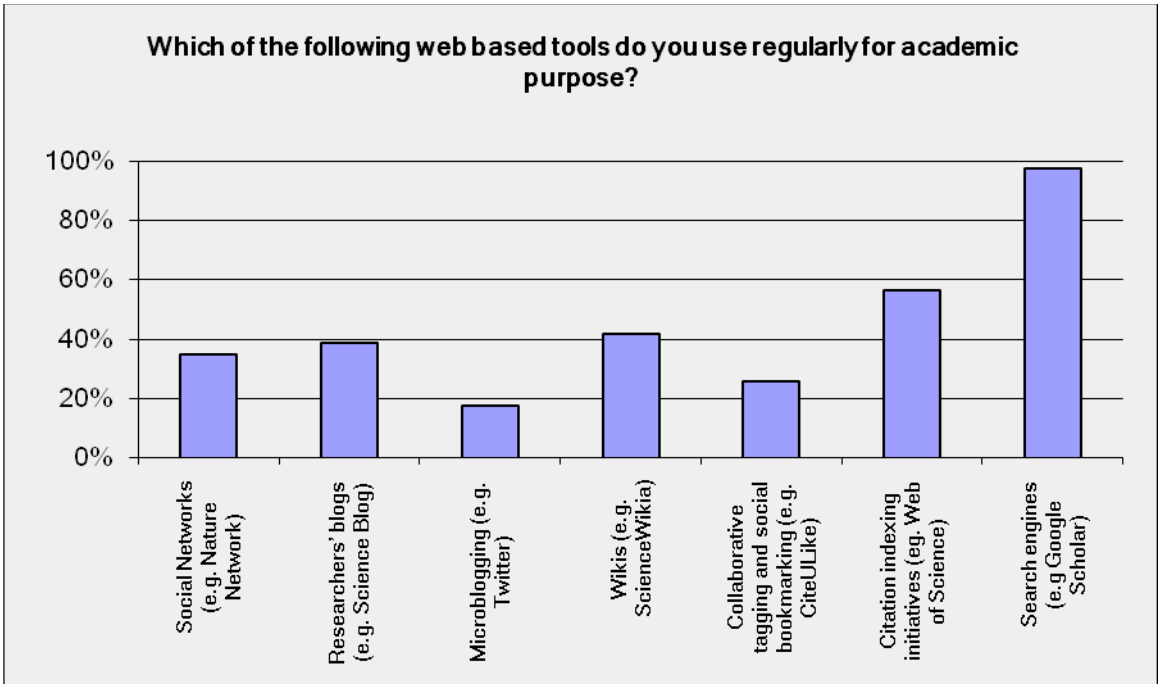

Figure 4: Percentage of usage of web based tools

Furthermore, Web 2.0-based applications have been used to varying extents depending on the *type* of scientific object (e.g. reference material or articles) as well as the *phases* of the scientific knowledge production process (Figure 5). Since multiple answers were possible, the absolute numbers are provided. The results show that concerning the objects, Web 2.0 is least used for monographs and handbooks while more experiences exist with scientific articles and educational material.

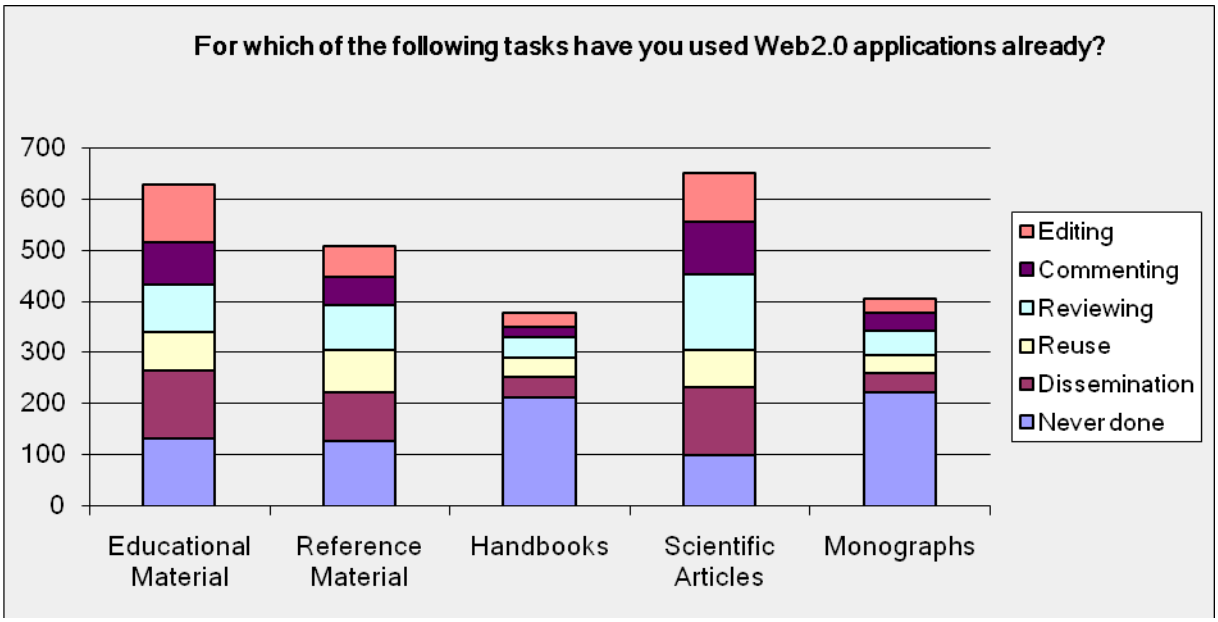

Figure 5: Ever used Web Tools and for what

Since we also asked whether they would like to use Web 2.0 tools in the future we can compare the actual use with the desired or planned usage in the future. First of all the results show that researchers want to use Web 2.0-based applications more extensively in the future for almost all types of knowledge objects as well as for all phases of the production process. Web 2.0-based applications appear particularly attractive for reviewing and dissemination phases (Figure 6).

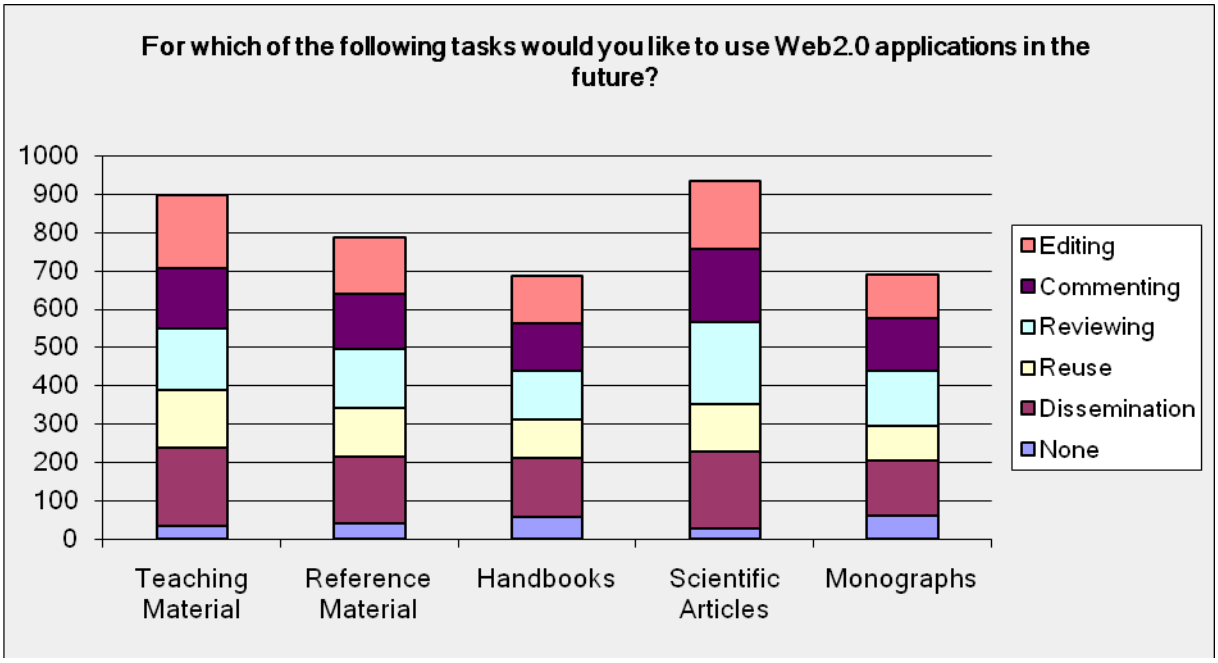

Figure 6: Possible use of web 2.0 tools

Another important topic we explored concerns novel models for the attribution of authorship. When asked to what extent researchers would be willing to extend authorship to people who have contributed to the article with comments, by editing, improving readability or providing data. The results indicate that respondents are most inclined to include those who provided data into the list of authors (62.6%). For the absolute numbers please confer to Figure 7.

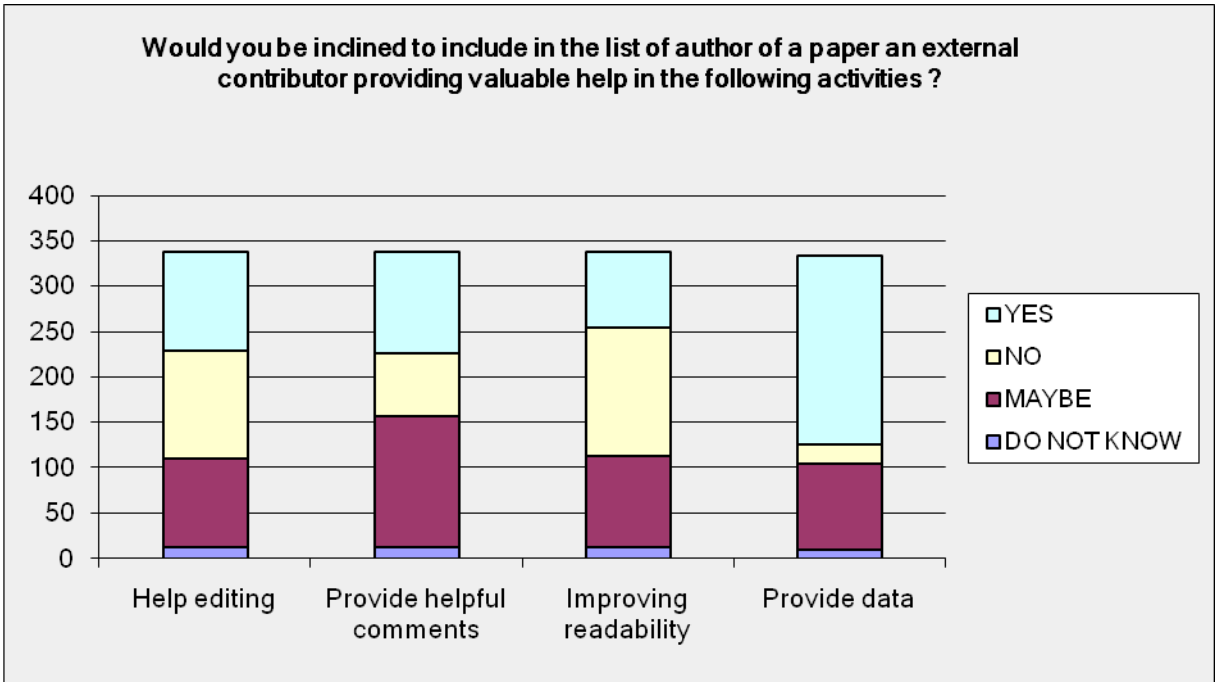

Figure 7: Willingness to add new authors according to their contribution

Another possibility concerns the possibility to attribute authors to subsections of papers (Casati et al. forthcoming). The responses concerning the merits of such more modular forms of credit attribution were divergent. While 42.3% would welcome such an option, 42.6% cannot conceive making use of such a possibility and 15.2% are undecided (Figure 8). Another proposal made by (Casati et al. forthcoming), concerns the introduction of a "production box" on top of articles, in which roles other than authors can be indicated, e.g. lab director, idea giver, etc. The majority of respondents would like to make use of such an option to disentangle credit attribution from authorship order (Figure 9).

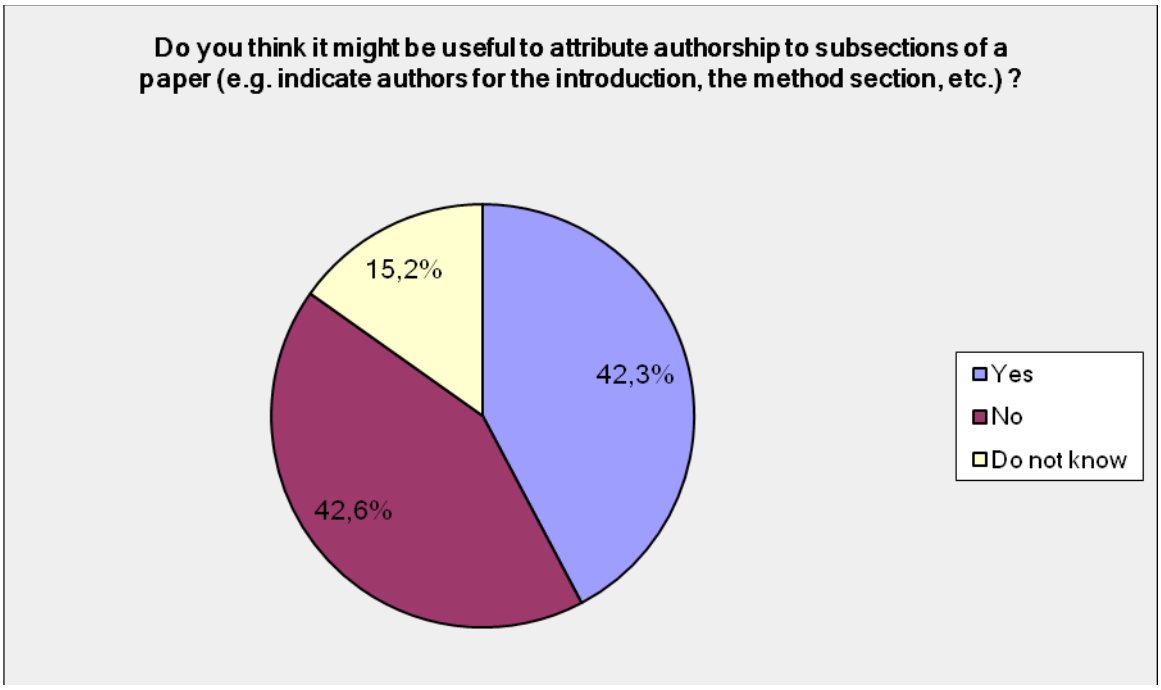

Figure 8: Willingness to attribute authorship to subsections of a paper

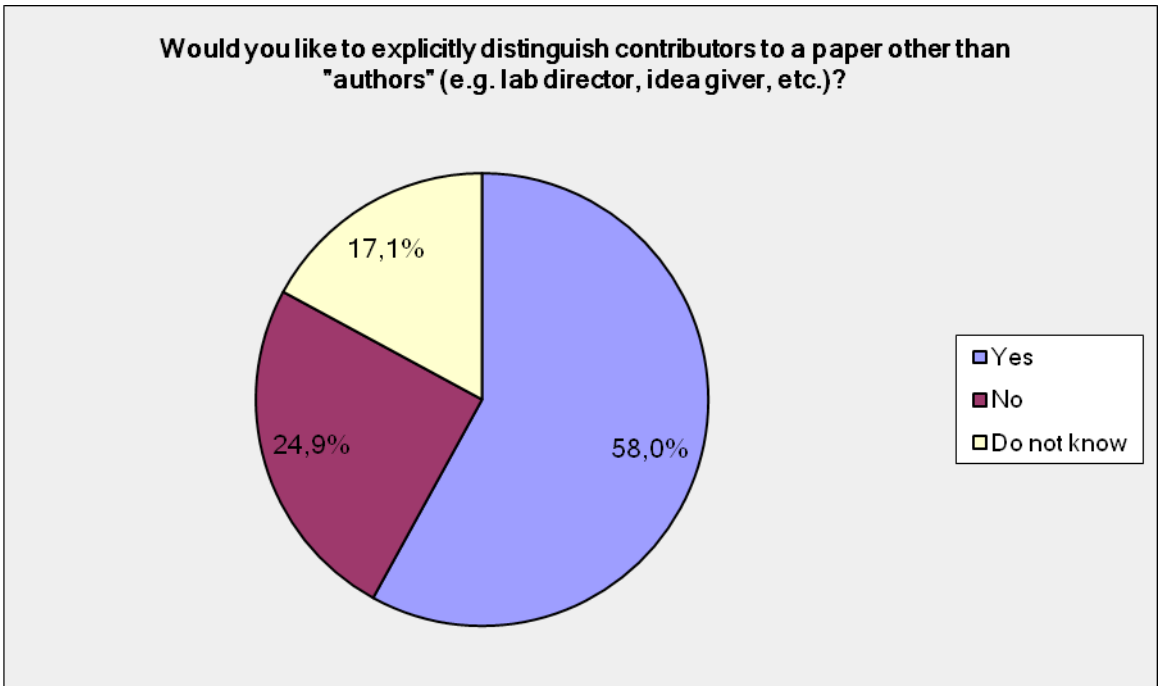

Figure 9: Willingness to distinguish contributors to a paper other than "authors" (e.g. lab director, idea giver, etc.)

The **third section** asked about reputation assessment. Respondents were asked to what extent a set of criteria were currently considered relevant for the evaluation of a researcher in their respective discipline (cf. Figure 10). Among these options, the number of citations and the number and quality of publications were still considered most relevant. On the opposite side, four criteria (Personal Webpages, Personal blogs, Membership in professional social networks, Presence in user-generated tagging services) were considered least relevant.

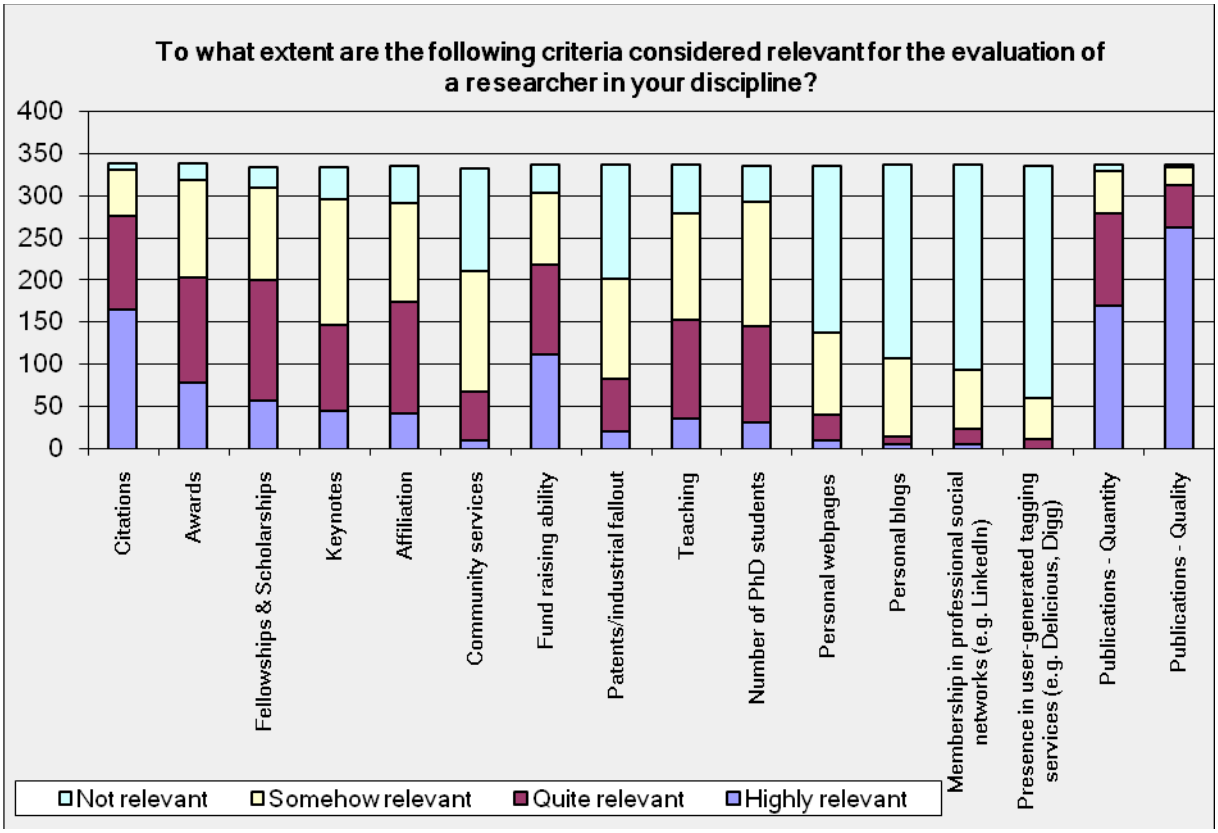

Figure 10: Relevance of different criteria used to evaluate researchers

The section further asked respondents what criteria influence the researcher's evaluation of a scientific contribution. Here, also traditional criteria still are considered most important (i.e. presence in citation indexes, reputation of publishing venue). However, publications in open access journals or archives and the number of occurrences in search engines are also considered relevant, albeit to a lesser degree. Finally, personal blogs or websites as well as presence in social bookmarking systems are not considered very relevant for the assessment of the quality of scientific content (Figure 11).

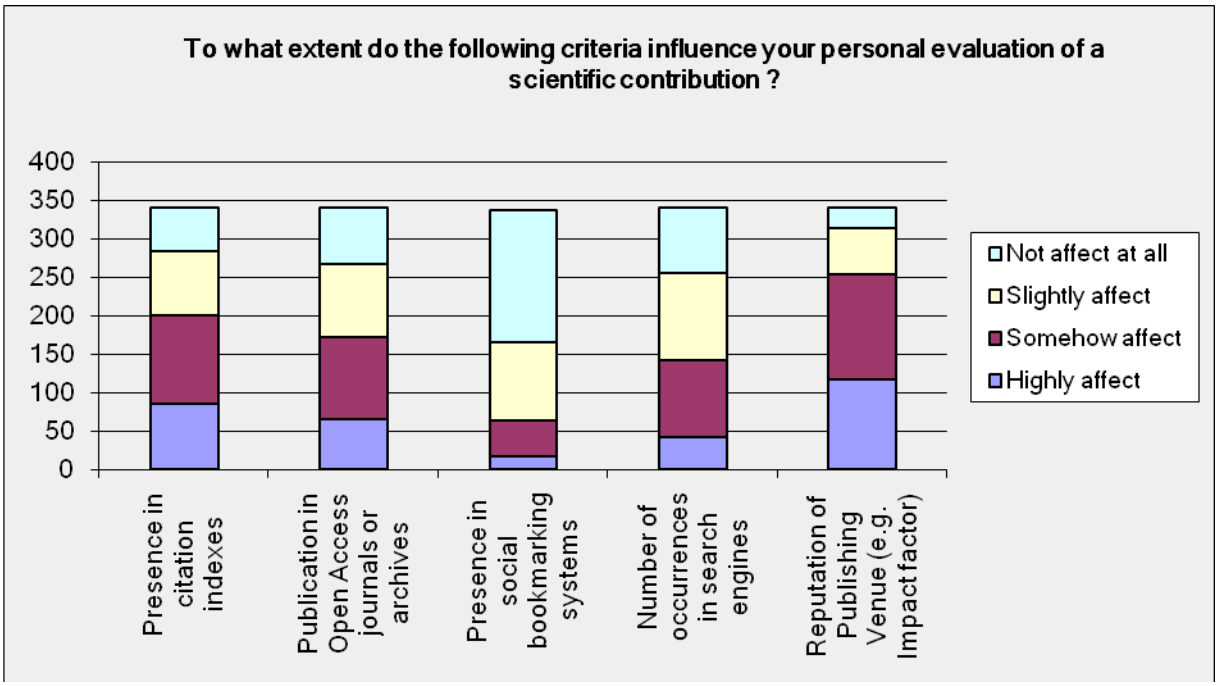

Figure 11: Relevance of criteria used to evaluate researchers

Section four addressed the topic of peer review. Respondents were first asked whether peer review is sufficiently able to achieve a set of chosen objectives. Peer review is considered quite effective as a filter to select the best manuscripts for a journal (strongly agree + somehow agree: 72.5%), to improve the readability of the published papers (73.6%) and to detect error (74.0%). Yet, it is perceived less effective in improving the quality of research by making suggestions for improvement and future research (49.6.2%) and in detecting plagiarism and fraud (44.1%) (Figure 12).

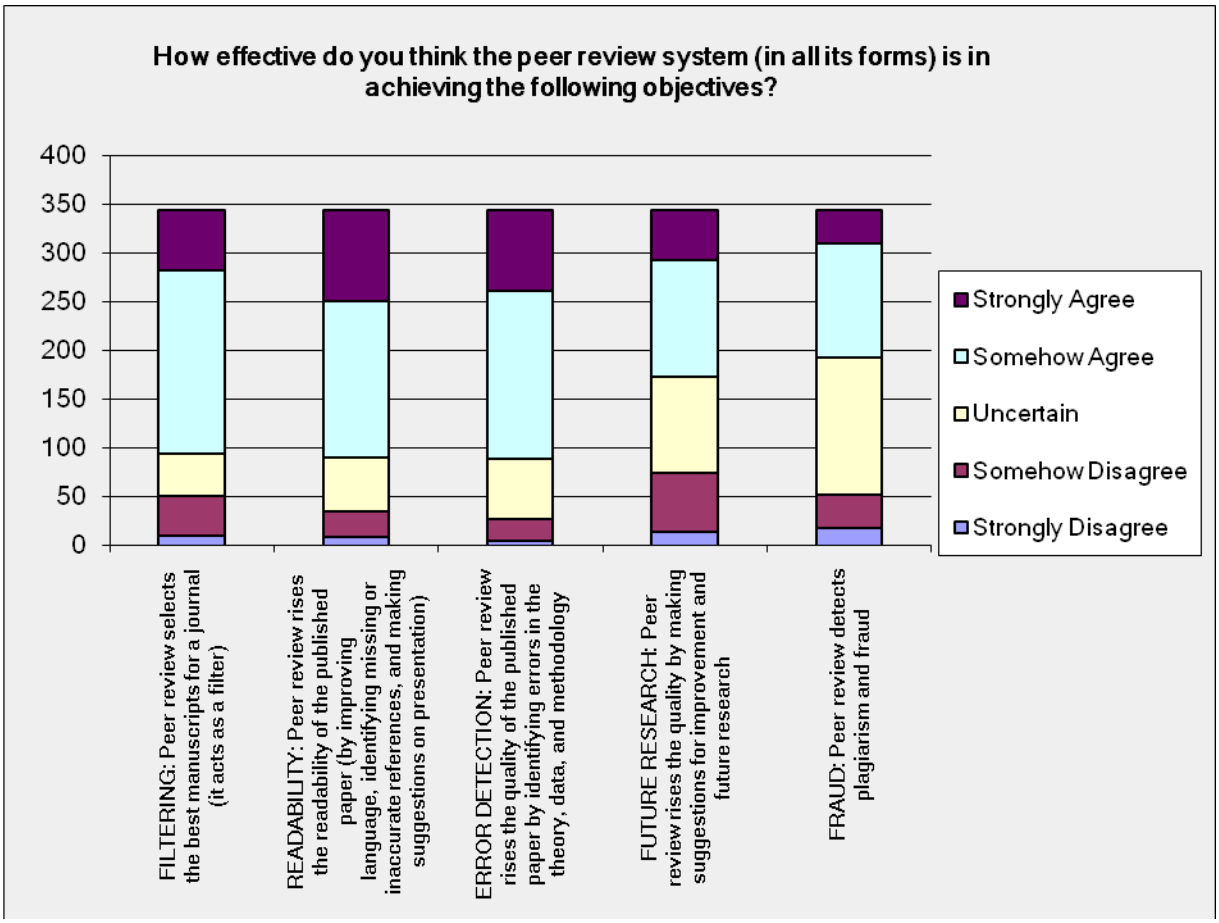

Figure 12: Beliefs of effectiveness of the peer review system

When focusing on specific types of peer review, respondents still consider double blind review (neither author's nor reviewers' names are known to each other) as the most effective type in assuring a good quality of results (strongly agree + somehow agree: 78.9%). Other types are not considered as reliable and effective: only 38.2% think that single-blind review (author's name is known to reviewers, but not vice versa) ensures quality, 40% think the same for open peer review (author's and reviewers' names are known to each other). Indeed, new models of community-based informal reviews are considered to be much more reliable than single-blind and open review processes: 61.7% agree that post-publication review (paper is commented and/or rated by readers following publication) assures quality and 59.1% state the same for pre-publication review (pre-print of paper undergoes a public informal review before publication) (Figure 13). This implies that while community-based reviews are preferred to open and single-blind, the *time* of such informal review (before or after publication) is not considered very important.

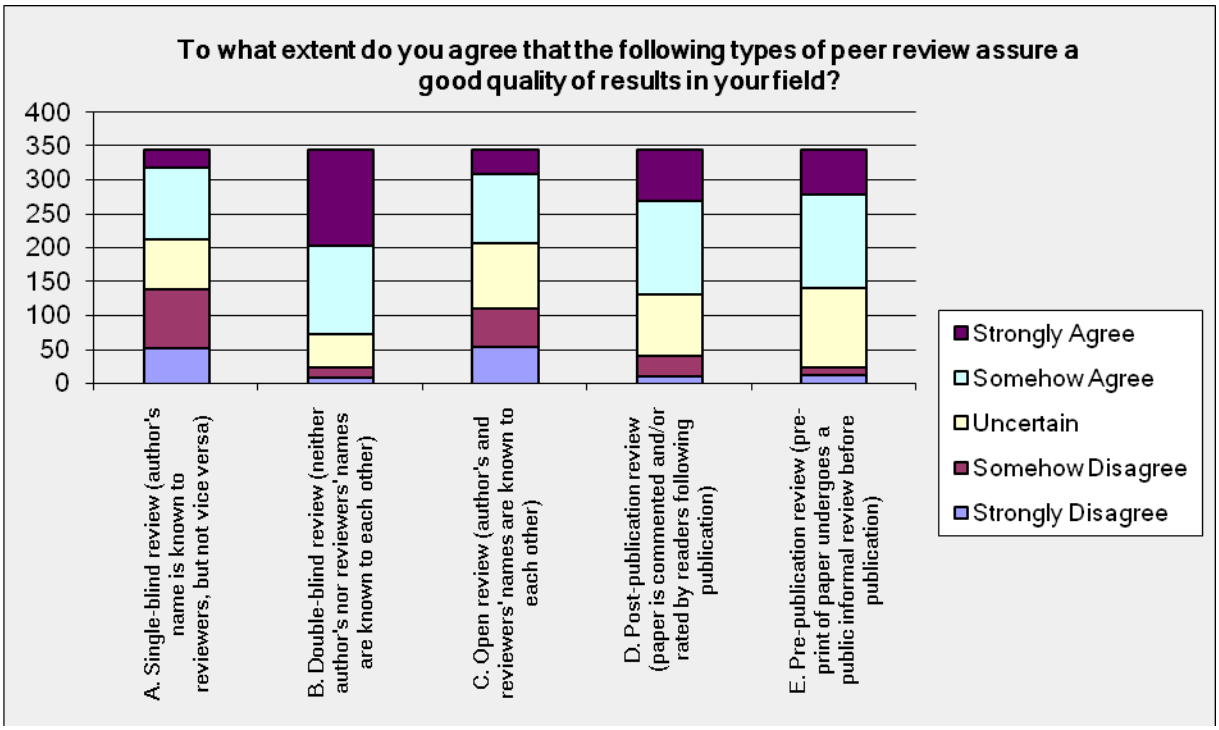

Figure 13: Type of peer review and expected ability to assess quality of scientific knowledge

As becomes obvious, peer review is a term that covers an abundance of different practices (cf. also (Wakeling et al. 2009). Moreover, the type of content also plays a major role. Our results show that there is a huge difference between the effort spent to review papers for conference or journal. Concerning conferences' papers, most of the respondents spent between 1 and 2 hours to review a paper for a conference (28.5%). Regarding journals papers, most of the people spend more than 2 hours in reviewing them (62.7%) (Figure 14).

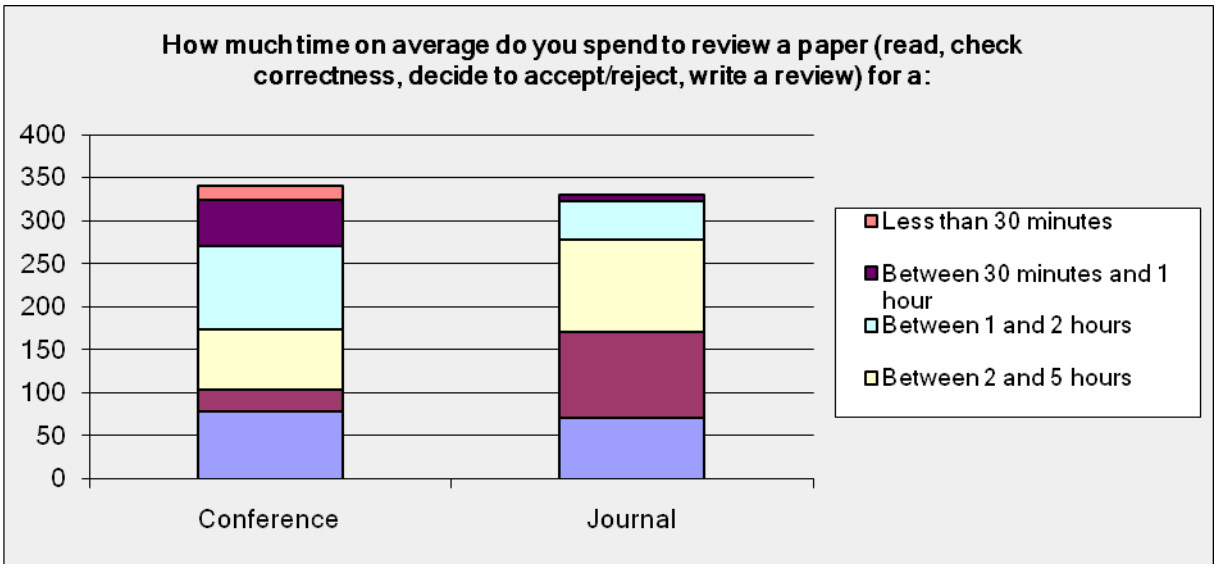

Figure 14: Time spent for review Conference vs. Journal paper

In **section five** we compared different publishing models, in particular, the differences between traditional subscription models to journals and the Gold and Green Open Access approaches. The currently most wide-spread publishing model is the subscription-only model, in which only those persons or institutions that pay fees to publishers can access articles, i.e. *readers* have to pay fees. In contrast, the Gold Open Access is a publishing model in which published articles are freely available and redistributable, but only if the *author* or his/her institution pays a fee to the publisher to cover the publication costs. Finally, the Green Open Access, also known as author self-archiving, involves the author posting a preprint version of their paper in a publicly-available online repository either before or after official journal publication. From our respondents, 40.9% of the respondents always self-archive their articles, while another 39.4% do this often or sometimes (Figure 15). This is a model which does not involve publishing companies.

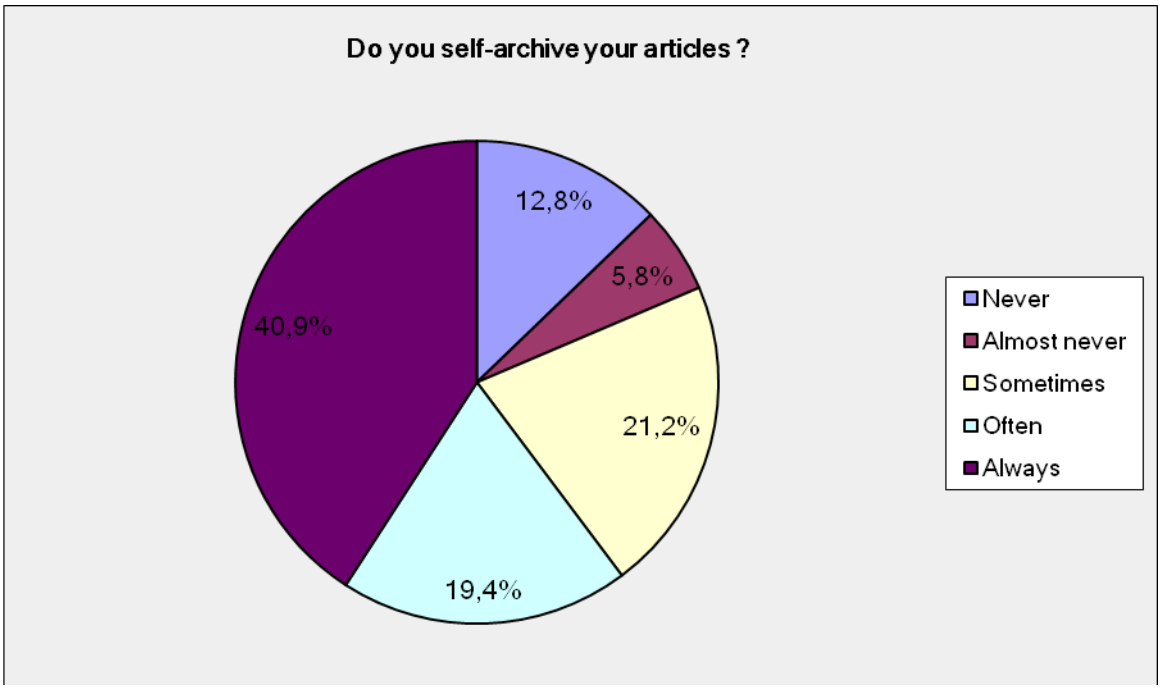

Figure 15: Self-archiving habits of respondents

Surprisingly, the Gold Open Access was known to less than half of our respondents (46.4%). Furthermore, while most of the people state that it would be desirable for the discipline (47.0%), a huge part of respondents (32.2%) does not know whether it might be useful or not. 20.9% explicitly do not consider the model to be desirable for their disciplines. One of the reasons for these divergent opinions may lay in the following: While 64.3% of the respondents would like to pay for keeping their papers openly accessible, only 16.5% of the respondents have the funds to do so.

Finally, the Green Open Access model is also perceived as having pros and cons. While it is believed to support a better dissemination of scientific work in the community (87.6%) and to increase the speed of feedback from the community (75.5%), the approach is seen as less effective in ensuring the quality of publications. 54,3% consider the lack of quality assurance as the major disadvantage of Green Open Access. For further information, confer figures 16 and 17 on the perceived advantages and disadvantages of the Green Open Access model.

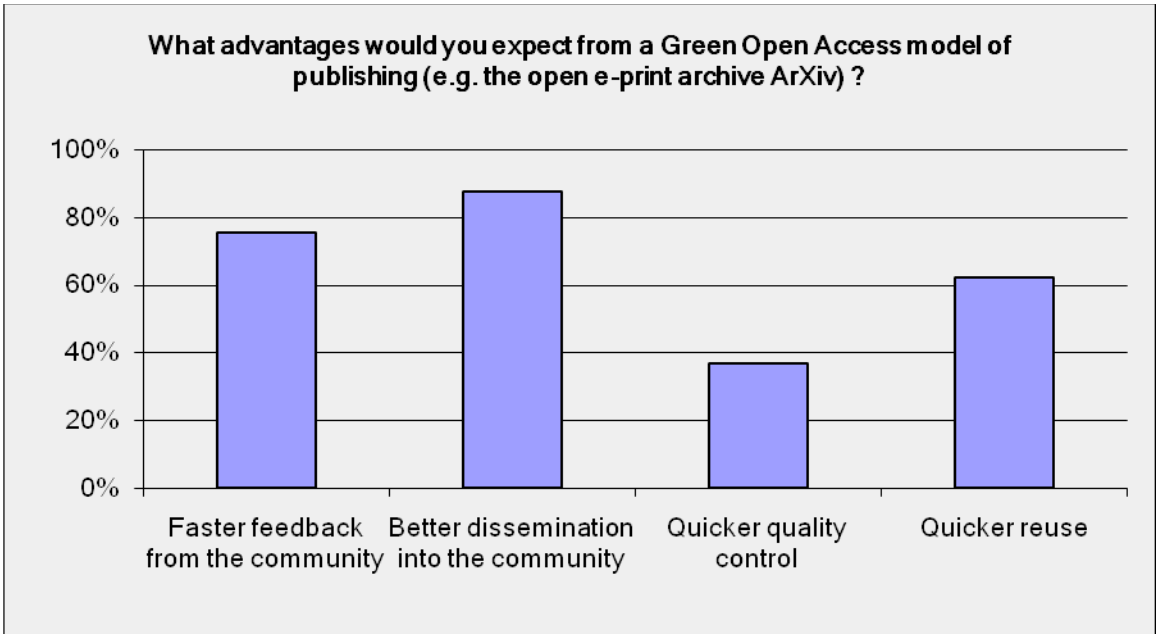

Figure 16: Advantages of the Green Open Access

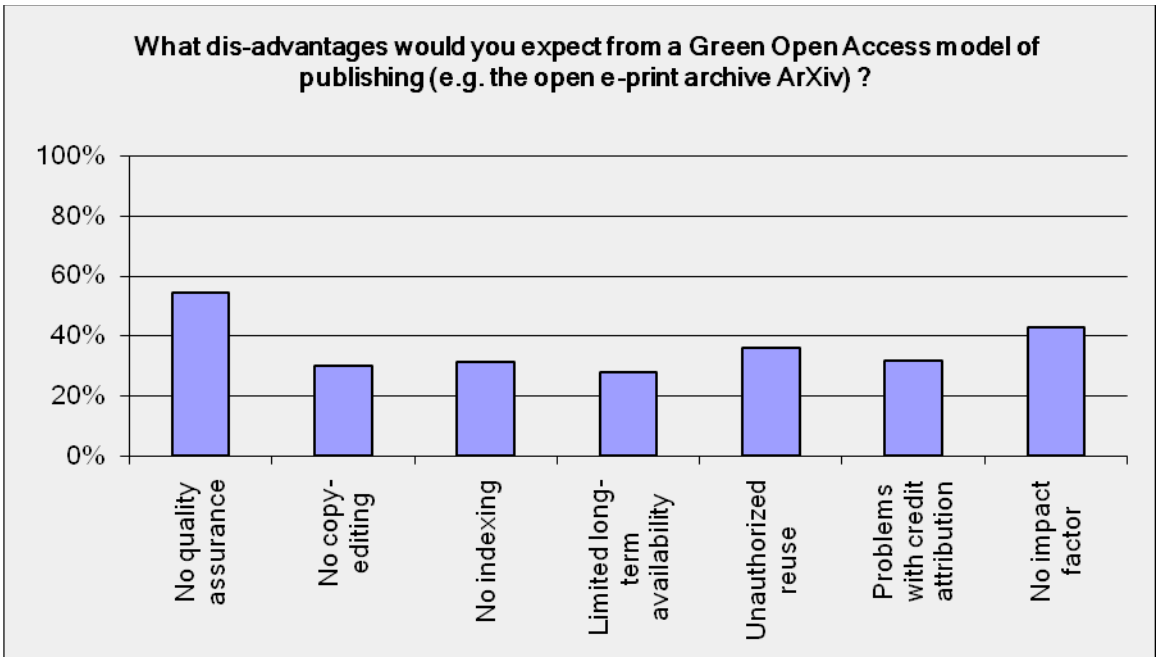

Figure 17: Disadvantages of the Green Open Access

Section six concerned copyright issues and the dissemination of scientific knowledge. The majority of the researchers would like to either dedicate the copyright of their work to a public domain (41.8%) or would like to retain copyright (42.1%). It is important to note, that only a 2.3% of respondents would like to transfer the copyright to a publisher as is current common practice (Figure 18). Hence there seems to be a strong request for a change of practice with respect to copyright practices in scientific publishing.

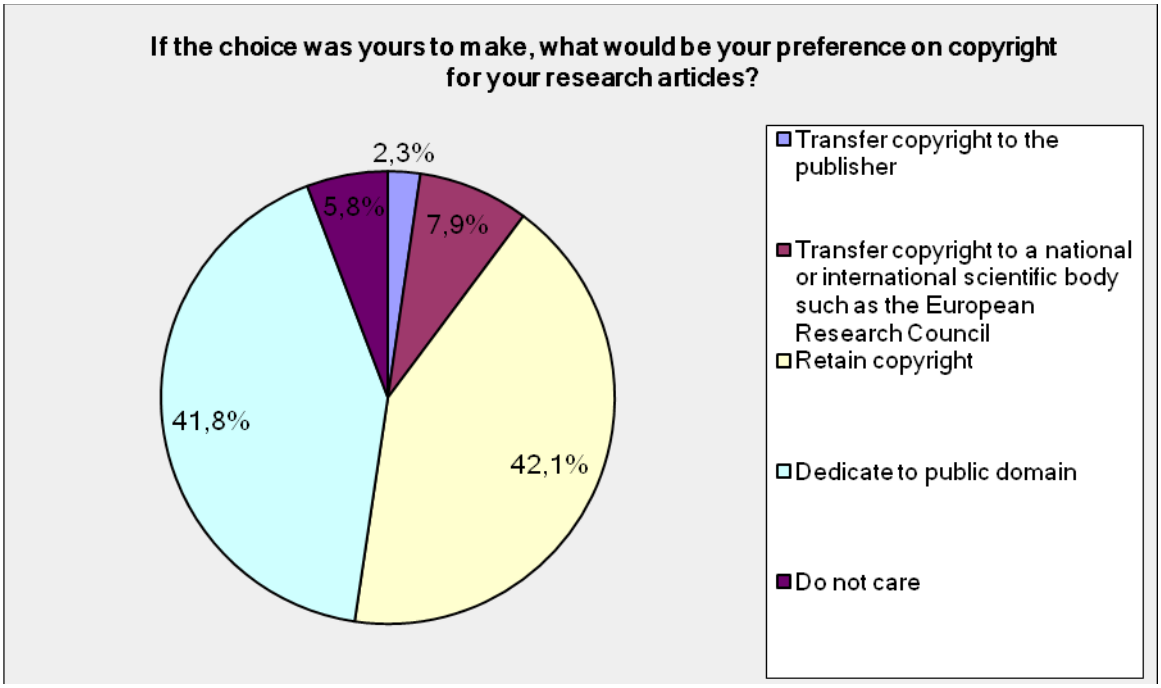

Figure 18: Preferences about copyright

With respect to possible licenses, most researchers prefer the free distribution of verbatim copies of their work (55.4%). 25.5% also accept modified re-distribution. Crucially, only 1.7% of the respondents prefer the currently common subscriber-only model managed by publishing companies (Figure 19), while 23.1% would prefer a free distribution which is managed by a publishing house.

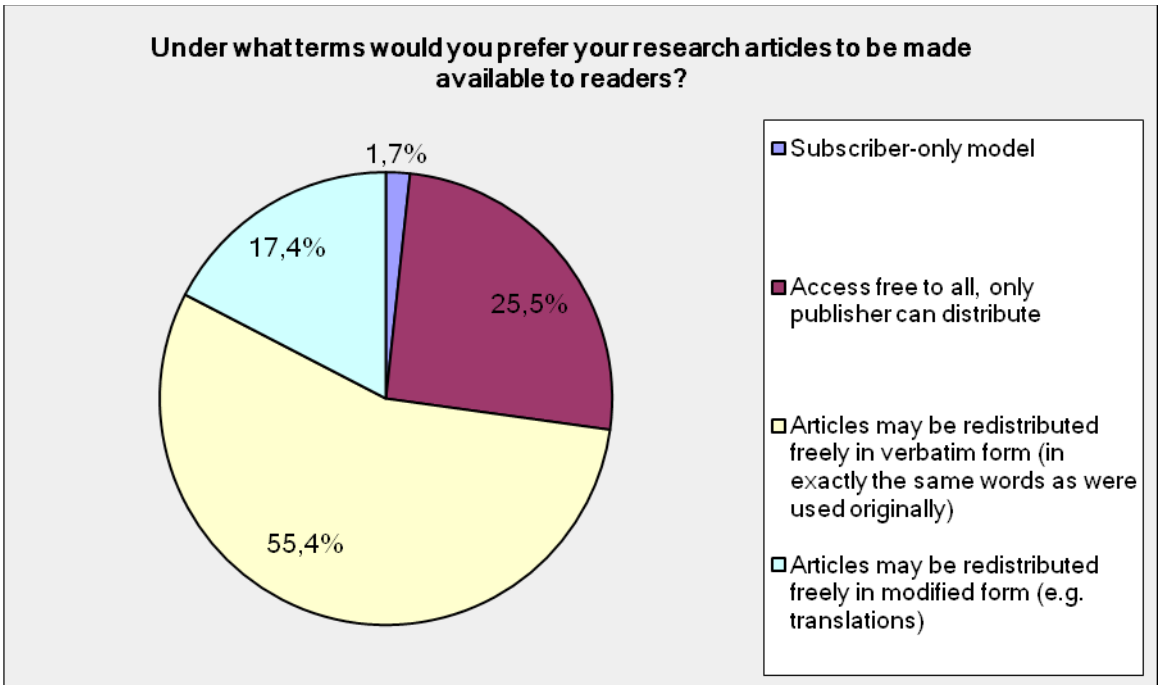

Figure 19: Preferences about the distribution of research work

DISCUSSION

The full spectrum of responses shows that researchers do have some experience in using Web 2.0 tools and it shows that they would like to use them even more in the future. In particular, the phases of review and dissemination appear to be most attractive for Web 2.0-based applications.

The results of our survey also show that it may be time to conceive new models of authorship attribution. Differentiating lab directors and idea givers from authors, distinguishing who wrote the introduction and who wrote – and did – the analysis, may be useful in disentangling authorship order and making scientific work distributions more transparent.

For the assessment of reputation of researchers and scientific content, it seems that Web 2.0-based proxies are still a lot less relevant than traditional criteria such as the quality and quantity of publication, fund raising abilities and citations index. This is a major point as the career of researchers is largely dependent from the recognition of their work.

When looking at the different types of review models, innovative and collaborative review models gather interest and may increase their role in the future of web-based scientific research. Double blind peer review is still seen as the most effective way to check the quality

of scientific knowledge, however merits of community-based review processes (pre- as well as post-publication) are also acknowledged.

With respect to the accessibility of scientific content, the pros of open access approaches in contrast to subscriber-only models are clearly indicative. However, both, the green as well as the gold model are perceived as having pros *and cons*. While the Gold Open Access model appears not affordable for many researchers or their institutions, sound and reliable quality control mechanisms for Green Open Access models are still to be conceived and/or validated.

Regarding the copyright, researchers prefer creative commons-type copyright models over those offered by most publishing houses to protect and disseminate their work. Indeed, if we take into account that 98.6% of the respondents would like to make their articles freely available, it seems that dissemination is much more crucial for them than their protection. Further, the current model of transferring the copyright to publishers is by far the least supported model. Since only 2.3% are satisfied with the current practice of transferring copyright to the publishers, a change of such practices seems to lie at the heart of the scientific community at least in so far as it is represented by our sample. Interestingly, although 41.8% would like to dedicate the copyright to a public domain, the European Research Council as well as other national or international scientific bodies are only considered by 7.9%. This raises the question of the visibility and trustworthiness of such institutions and the ideal form of such an obviously desired public domain institution.

CONCLUSIONS

We think that this survey on Web2.0 and scientific publishing raises many important issues which need to be addressed by us as researchers as well as by national and international research funding bodies. It is in particular the strong dissatisfaction with restrictive access to scientific content that needs to be targeted. Here, the issues are not of technological nature or can be solved technologically. Rather, policy intervention appears to be necessary to change the current situation for the better. On the other hand, developing robust and reliable means of community-based quality control mechanisms should rather be a task for us as researchers.

With respect to future research there are more analyses of the data to be done. One aspect we did not analyze yet, but plan to do in the near future concerns potential differences between

different groups of respondents. Are there substantial differences between researchers at different stages of their careers? Or even more importantly, are there significant differences between different disciplines? These questions we will address in the near future.

Finally, one has to take into account that our methodology may have induced a bias into the results. Clearly, by distributing a survey on Web 2.0 and scientific publishing via mailing lists first of all we reach only those researchers who are as ICT-affine as to use mailing lists. Secondly, only those interested in Web 2.0 may take the time to fill out the questionnaire. Hence, there may be a bias that those researchers who took part in our survey are more positively attuned to Web 2.0 than the average researcher. The reason behind our approach was to reach a diversity of scholars from different fields in a reasonable amount of time, but surely the downside of such an approach has to be kept in mind when interpreting the results.

REFERENCES

Casati, R., Origgi, G., Simon, J., (forthcoming). Micro-Credits in Scientific Publishing, *Journal of Documentation*.

Wakeling, J. R., Medo, M., Turrini, M., Birukou, A., Law, J., Marchese M. & Zhang, Y.-C. (2009) "Analysis of reviews and modelling of reviewer behaviour and peer review", LiquidPub Project Report D1.1 pp. 29-58.