EnablingOpenScholarship
Université de Liège - Rectorat, place du 20-Août, 7 (Bât. A1), B-4000 Liège, Belgium

COM(2009)647
CONSULTATION ON THE FUTURE "EU2020" STRATEGY

Response from Enabling Open Scholarship (EOS)

A. Opening statement

We concur with the Commission's premise that knowledge is the engine for sustainable growth. Europe must become a sustainable knowledge-based economy or fall behind other economic regions. Open Access to publicly-funded research will advance European science and technology, will speed up research, will reduce duplication, will increase the usage and impact of research, will facilitate interdisciplinary research, will enable the deployment of new semantic technologies to create new knowledge from existing research findings, will provide the wherewithal for better, smarter research assessment and management and will provide greater payoff for the taxpayer from the funds invested in research across the ERA. A number of constituencies will benefit: as well as the research community itself, which will have immediate and untrammelled access to the information it needs to do its work, the professional, practitioner and lay public communities will also benefit. Access must be made available in ways that permit full re-use of research results and through services that maximise ease of use and convenience for the relevant user constituencies. We elaborate on these points in our detailed response below.

B. Enabling Open Scholarship (EOS)

EOS is an organisation of universities and research institutes worldwide whose managers have come together to discuss, shape and promote the principles of open scholarship. EOS has members on six continents, from the largest, broad-based universities and research institutes to some of the smallest, most specialised research-based institutions. As well as universities and research institutes, EOS also has government departments and research councils in several different countries as members. The EOS Board has collectively developed and signed this submission: Board members are listed at the foot of the document.

C. EOS response to the Consultation document

We focus our comments on a particular priority identified by the Commission – "creating value by basing growth on knowledge", although the benefits of open scholarship are also felt in the other two priority areas as well. Within this priority area, our key concerns are with the principles of open scholarship and the issue of the accessibility of research information. Open scholarship is about the opening up of scholarship and research that we are now seeing through the growing open access, open data, open education, open science and open innovation movements. These, and other, 'open' approaches to scholarship are changing the way research and learning are done and there are profound implications for universities and research institutions. EOS has been established to help guide developments and to assist others in understanding the issues and their implications. We support the conclusions[1] reached by public consultation and presented at the recent ERA 2009 conference, “Working Together to Strengthen Science in Europe”[2] and look forward to working with the Commission wherever possible to help the recommendations become manifest.

The debate around open access has matured, after a decade, from one of whether there should be public availability for publicly-funded research findings, into one about how this should be made possible. It is, however, appropriate that we rehearse the main arguments and describe the beneficiaries and effects of open access as a preliminary to our points on how the EU can improve its scientific communication system.

1. Who benefits from publicly-available scientific research results in Europe and what results should be available to them?

There will be a heightened return to European taxpayers from their investment in research if all results from that research are immediately and freely available. The return on the investment accrues to a number of stakeholder constituencies, namely:

- Researchers in all disciplines within the academy, who not only spend less time searching for, locating and accessing information but who can better contribute to solutions to major global issues such as poverty alleviation, health improvement and climatic studies
- Research managers, for monitoring research progress and direction
- Research funders, for monitoring research progress and direction
- Research communities outside the academy (e.g. R&D-based industries, military research establishments, etc.)
- Professional communities outside of the academy (e.g. medical services, the legal community, accounting professionals, etc.)
- Practitioner communities outside the academy (e.g. civil engineering and energy sectors, architecture and the built-environment sector, etc.)
- Cultural heritage organisations (museums and galleries, in both public and private sectors)
- Freelance and private scholars
- The lay public (interested individuals who can use research information in their daily lives)

The difficulty most frequently expressed by ***researchers within universities*** is their inability to access journal articles they have identified as useful because of a subscription barrier (Jubb et al, 2007) and this issue heads the list of top concerns of researchers across all disciplines reported by Swan (2008a). The recent report by the UK’s Research Information Network concludes that ‘many researchers are encountering difficulties in getting access to the content they need and that this is having a significant impact on their research’ (RIN, 2009). This is evidence that even the largest and best-funded universities cannot afford to buy in all the research information that their researchers need, as inter-library lending budgets continue to testify. For example, Imperial College London, one of the UK’s elite universities, still needs to spend around €35,000 per annum on the purchase of articles from journals to which the institution does not subscribe.

---

[1] http://blogs.ec.europa.eu/ERAconference09/tag/conclusions-15/
[2] http://ec.europa.eu/research/conferences/2009/era2009/index_en.htm

So long as journal articles are obtainable only through Toll Access (via subscription or pay-per-view), it will continue to be difficult and costly for would-be users *outside* institutions with well-stocked libraries to find and use all the research material that they need.

Into this category of would-be users fall the ***R&D-based SMEs***, the bedrock of the knowledge economy and innovation-based industry. The EU's own Community Innovation Statistics[3] have shown that SMEs are not able to source information needed to innovate easily from universities. Indeed, universities and government research institutes are at the bottom of the list of cooperative partners that SMEs use for sourcing research information. A recent survey by Ware (2009) reported that 73% of UK SMEs had difficulty in accessing journal articles they need, as do 53% of large companies. The same study found that just 2% of SMEs, 7% of large firms and 17% of higher education-based researchers report having access to all the articles they need for their work. These numbers correlate with those from a much larger study conducted five years previously (CIBER, 2004), indicating that access has not improved in any meaningful way in that half-decade. And Swan (2008b) showed that UK SMEs have equal difficulty locating the grey literature (theses, dissertations, technical reports, working papers) that universities have created.

A great knowledge economy cannot be built on this kind of foundation. Perhaps the best way to bring the point home is to quote the managing director of a small biotechnology company, who said:

> *"With a small oncology company … it is imperative that I have access to the literature. But small companies do not have the "deep pockets" necessary... The for-profit journal publishers have effectively barred access to key scientific information except to those who can afford their outrageous fees. Much of the most innovative work is being done at companies like mine that cannot afford to pay $30+ per paper or pay per-search charges in abstracts or journal collections."*[4]

If companies cannot afford access to the journal literature essential for their business, access problems are all the more hindering for institutions in the less wealthy cultural heritage and practitioner communities, let alone for freelance and independent scholars whose numbers are not inconsiderable in Europe.

These communities would certainly use journal articles if they were available to them, as would the research community itself. Data from university repositories shows the level of usage that freely-available articles enjoy over an above the usage they get from publisher websites by those who have access through subscriptions. For example, the University of Liege's ORBi repository has 14,000 items at the time of writing, and sees 4,000-5,000 downloads a month by users. Southampton University's School of Electronics & Computer Science repository has around 18,000 items with an average download figure of over 30,000 downloads *per month*. This is usage that is *in addition* to usage from the journals themselves and illustrates the latent demand for the fruits of scholarship that is not satisfied through

---

[3] http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-07-081/EN/KS-SF-07-081-EN.PDF

[4] Terence Dolak, SDR Pharmaceuticals, Andover, NJ: Chemistry Central blog, Thursday May 24, 2007: http://blogs.openaccesscentral.com/blogs/ccblog/entry/unemployed_retired_might_lose_touch

the access-restriction (Toll Access) model that has hitherto dominated the scholarly communication arena.

Along with journal articles, research datasets must also be made publicly available in an appropriate way. This means that data created from publicly-funded research must be placed in the public arena in a form that permits re-use by others. The precise way in which this is done most appropriately will vary from discipline to discipline and presupposes that the infrastructure is available to accept, store and preserve the data. This issue is complex and is the focus of many studies and trials. There are also considerable cost implications in undertaking to curate and preserve research data in the long term. Universities and research institutes can do so much, but the technical demands of long-term preservation of data from across the broad academic spectrum are not going to be met adequately by individual institutions nor should be. There is an argument that this should not even be done at national level but that there should be a coordinated European approach to data management, especially given that a majority of Member States signed the *OECD Declaration on Access to Research Data from Public Funding* in 2004[5].

Public databanks already exist in some disciplines, and principles for optimal data curation and preservation are being worked out by careful study of the needs and practices of different disciplinary communities: see, for example, the JISC 'Managing Research Data' programme[6] in the UK, the Data Audit Framework[7] and studies on aspects of data management, including costs: (Swan and Brown 2008a, 2008b).

Our recommendation on this issue is that the Commission uses this foundation work to develop plans for a scientific data management policy for the nations of the EU and to accompany the policy with plans to establish a suitable, robust, funded infrastructure for data curation and preservation.

2. Is policy development needed to achieve satisfactory public access and, if so, what kind?
Policy is needed to ensure that a publicly-available corpus of scientific literature and data accumulates at the hoped-for levels. The concept of open, public access is new and turns on its head the traditional way of communicating research findings, which is based on access restriction, not access maximisation. Despite the opportunity the Web brings to open up research communication processes, researchers do not instinctively turn to such practices because of strong pressures to adhere to access restriction-based norms. A policy adopted by their employer or research funder raises their awareness of the issue, helps overcome their fears and anxieties, and provides the framework within which they can then act.

The interests of users of research information (researchers, libraries, universities, national and European governments, the public) are maximally served by *provision of research information on an Open Access basis, immediately available after peer review, and with minimal conditions on re-use*. Any open access policy should therefore make this the primary aim.

---

[5] http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html
[6] http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx
[7] http://www.data-audit.eu/

There should not be a delay period past the moment when peer review-required corrections have been made. The world is not yet so perfect, however, and while all rights to dissemination pass routinely to publishers there will always be a difficulty in enabling early accessibility due to publisher desires to restrict access, at least for a period (which may be long) to protect their revenues under an access-restriction business model.

There are a number of ways that a policy covering publicly-funded research might overcome this problem to enable the early access that is so important:

- The policy could be worded to require that manuscripts are deposited in pubic archives as soon as they are finalised after peer-review (i.e. as postprints). If the publisher of the journal in which a paper is to be published requires an embargo period on access, the full-text of the article may be kept 'dark' for the embargo period. The metadata, however, over which the publisher has no rights, should be made openly accessible from the time of deposit. This at least allows potential users to know that the article exists and they will be able to contact the author about it if they need it during the embargo period. Such a policy has been implemented in a number of our member institutions and it works extremely well as a pragmatic solution to accommodating publisher embargoes whilst enabling the earliest possible access.
- Policies can usefully be modelled on the existing examples (e.g. European Commission, European Research Council, NIH), where assigning the funder the right to disseminate a publicly available copy is a condition of grant. The terms and conditions of grant require grantees to retain that right and can assign the funder a non-exclusive license to disseminate the author's final manuscript. It is simple, elegant and has proved itself in practice.

An open access policy needs three elements:
(i) It must be mandatory. The policy from the National Institutes of Health in the USA was originally voluntary, with researchers encouraged and exhorted to make their publications openly accessible through PubMed Central (PMC). During the first year or so under these conditions, only around 4% of qualifying articles were deposited by their authors in PMC. Once the policy articulated a requirement (that is, it became a mandatory policy) for authors to ensure their articles were made openly accessible, deposit levels increased rapidly. After one year, the deposit level was 57% and has been growing since. Other studies also demonstrate that only mandatory policies bring the kind of level of open literature that is hoped for by research funders (Sale, 2006).

(ii) The effects of the policy must be monitored to see what the level of compliance with the requirement is, so a system to measure deposit levels against expectations needs to be put in place. The Wellcome Trust began measuring compliance in the first year after its own mandatory policy came into effect. The Trust was able to see that levels of deposit were not as great as had been expected and was able to understand what the causes of this were (partly author recalcitrance and partly publisher non-compliance). As a result of this understanding, steps were taken to make authors more aware of the requirement on them and to remind publishers that immediate deposit of articles that the Trust had paid to be made openly accessible was expected. Compliance by both constituencies rose markedly.

(iii) Sanctions, at least implied ones, must be available to the policymaker to use against non-compliers. The research community is not one where heavy-handedness goes down well, and the general evidence shows that once researchers have begun to see the benefits of making their work openly accessible they readily continue the process, but a new mandatory policy needs to be supported by clear guidelines about expectations and what will happen if those expectations are not met. Many universities, including some of our member institutions, have simply stated that the repository is the place where research managers will look when they wish to see the outputs of any member of research staff (i.e. for promotion boards, appraisals, etc). This brings about the desired effect, for researchers do not wish to be disadvantaged by not compiling a complete record of their outputs in the institutional repository[8]. The University of Liege, for example, implemented its mandatory policy from the end of November 2008 when the number of full-text articles in the University's repository was 974, and by the end of October 2009 the number of full-text articles stood at 14,090.

To enhance open access to peer-reviewed papers arising from publicly-funded research, the Commission should extend its own policy on open access to cover 100% of FP7-funded research outputs from the current 20%. It should also do more to support Member States in developing policies for their own national funding agencies. There is some progress here: the United Kingdom's seven Research Councils, for example, all have policies on Open Access to research articles and datasets. The Research Foundation Flanders has one, as do the Irish Research Council for Science, Engineering & Technology, the Swedish Research Council, the Agence Nationale de la recherche (ANR) in France and others. There are currently 42 mandatory policies on Open Access from research funders and the list of existing policies is given in appendix 1. The policies can be viewed in the Registry of Open Access Policies[9]. There is not a coherent picture across the EU, though, and this must be tackled.

Numbers have grown year on year: there were none in existence in 2005, but by the end of 2006 there were 9, by end-2007 there were 23, by end-2008 there were 30 and by end-2009 there were 42. European agencies would therefore not be treading new, experimental ground by establishing open access policies: indeed, from our viewpoint, representing the interests of universities and research institutes, we would welcome the swift development and implementation of suitable policies across the ERA.

3. The version of an article that should be made available for public access is the 'postprint', and access should be available immediately after peer review.

The version that should be made available under a public access policy is the author's final version after the corrections and changes required as a result of the peer review process have been incorporated. This version is normally called the '***postprint***'. The '***preprint***' is the version before peer review and although this is the version that the high-energy physics community makes available for public access in arXiv[10], this solution is not generally favoured across other disciplines. Indeed, there is a good argument for considering it to be bad or even dangerous practice in some fields (notably biomedicine), where inaccurate or

---

[8] http://listserver.sigmaxi.org/sc/wa.exe?A2=ind09&L=american-scientist-open-access-forum&D=1&O=A&F=l&P=866

[9] http://www.eprints.org/openaccess/policysignup/

[10] www.arxiv.org

misleading articles placed in the public domain before peer-based quality control may have undesirable consequences.

The ***published version*** of an article is laid out and presented in the publisher's house style. While this may be attractive, and the inclusion of page numbers may be convenient for citation purposes, the published version is almost always in PDF format. This inhibits the re-use of the article by semantic technologies (text-mining and data-mining software), thus reducing the utility and value of the article. Moreover, many publishers still insist on an embargo period before even the postprint can be self-archived and most do not ever permit the published PDF to be made publicly accessible. Delay in accessibility reduces the potential impact of an article for all time, as shown in a recent study by Gentil-Beccot et al (2009). Conversely, early open accessibility gives an article an increased impact over comparable articles that are available only through Toll Access: this impact advantage is persistent and continues to diverge positively from the impact of Toll Access counterparts. Below is Figure 4 from that study which shows the data on which the statements above are based.

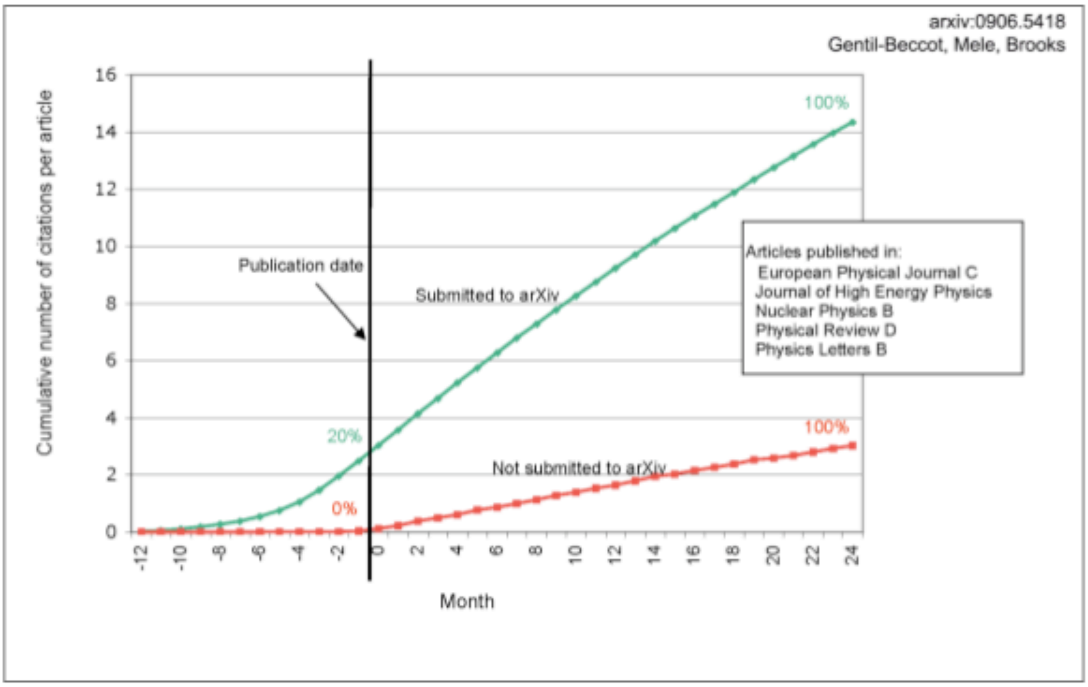

Figure 4. Cumulative citation count as a function of the age of the paper relative to its publication date. 4839 articles from 5 major HEP journals published in 2005 are considered.

(Source: Gentil-Beccot *et al*, 2009: published in *Scientometrics*, December 2009)

Other studies in other disciplines have also shown the same early access advantage (e.g. Henneken *et al*, 2006). It is therefore essential that a public access policy specifies that articles be made accessible as early as possible after peer review. The relative advantages and disadvantages of access to different versions of an article are summarised below:

Version	Advantages	Disadvantages
Preprint	• Very early access, useful in fast-moving	• Not appropriate for fields of

	fields of research and in small research communities where familiarity and trust between researchers negate the possibility of unreliable information misleading others	research where errors and misleading information in uncorrected manuscripts may have serious consequences if used by others
Postprint	• Early access is possible • Quality control has been performed • Article is to all intents and purposes the same as the published version and is an acceptable alternative for those (the majority) without Toll Access	• Has not undergone editorial quality control so may contain grammatical or other errors that should be picked up at proofing and copy-editing stages of the publishing process • Can be subject to delay in access if a publisher's embargo is not counteracted by a properly-worded policy
Published version	• Fully formatted and copy-edited by publisher	• PDF format inhibits machine re-use • Virtually always delayed-access, even if use is permitted • Use is usually not permitted by the publisher

As argued – and supported by empirical evidence – above, the earlier an article is made publicly accessible the greater its impact and the better value it represents for author, funder and European society. The optimal point for articles to be made publicly available if public interests are paramount is immediately after the changes required by peer review have been made.

Publishers will argue for an embargo period (of varying length depending upon discipline), saying that providing immediate access after peer review will harm subscription sales of their journals. We wish to set out quite clearly that this supposition is not supported by any evidence, at least as yet.

The evidence shows the opposite is true. Two society publishers have been doing business in the high-energy physics arena during the whole period when the arXiv has been in operation (since 1991). Effectively, <u>all</u> articles in high-energy physics and allied fields have been made publicly available through the arXiv over that time period, while at the same time being published as usual in subscription journals. If this form of open access were to have an effect on subscriptions, it would be expected to have shown up in high-energy physics during that period.

In 2005, two important society publishers in physics, the American Physical Society (APS) and the Institute of Physics Publishing (IOPP) in the UK, were asked about their experiences of living alongside the arXiv for 14 years[11]. There are two salient points to note from their responses:

[11] http://listserver.sigmaxi.org/sc/wa.exe?A2=ind05&L=american-scientist-open-access-forum&D=1&O=A&F=l&P=6121

1. Neither can identify any loss of subscriptions to the journals that they publish as a result of the arXiv.
2. Subscription attrition, where it is occurring, is the same in the areas that match the coverage of the arXiv as it is across any other areas of physics that these societies publish in.

Both societies, moreover, see actual benefits for their publishing operations arising from the existence of arXiv. The APS spokesman stated that the American Physical Society has *"cooperated closely with arXiv including establishing a mirror (jointly with Brookhaven National Laboratory)... We also revised our copyright statement to be explicitly in favor of author self-archiving. These efforts strengthened (rather than weakened) Physical Review D [an APS journal that covers high-energy physics] ...I would say it is likely we maintained subscriptions to Physical Review D that we may otherwise have lost if we hadn't been so pro-arXiv .."*

In answer to the question "Does arXiv worry or threaten your business?" the APS answered:

> *"We don't consider it a threat. We expect to continue to have a symbiotic relationship with arXiv. As long as peer review is valued by the community (and it seems to be), we will be doing peer review. While the APS aspires to open access and is not threatened by arXiv.org, we do have strong reservations about government requirements for open access."*

The Institute of Physics Publishing's response was:

> *"IOPP's experience as a learned society publisher illustrates the strong synergies and mutual benefits that currently exist between major peer-reviewed journals, such as our Classical and Quantum Gravity, and the arXiv e-print server. Both systems continue to serve the scientific community very effectively. Journals act as the "brand", setting standards for scientific quality. Our authors and editors tell us that they value publishing in a peer-reviewed journal because this continues as an essential requirement for establishing reputation and authority of the research they publish. Whilst posting an pre-print or post-print is becoming more of an essential in some areas of the physics community for immediate and wide dissemination, we do not see the arXiv or repositories threatening our business."*

Gene Sprouse, Editor-in-Chief at the American Physical Society, recently contributed further comments[12] on this issue as follows:

> *"1) One piece of information in the recent work of Gentil-Beccot, Brooks and Mele http://arxiv.org/abs/0906.5418v1 is that papers put on the arXiv and never published in a peer reviewed journal are cited 1/6 as much as those that are eventually published. In addition, papers submitted to peer review journals that were never put on the arXiv also suffer the same low citations.(see Figure 2). The American Physical Society, publisher of Physical Review D, has embraced and supported the arXiv, as it provides the "newspaper" for the field, while the journals provide the validation. Both are necessary.*

[12] http://listserver.sigmaxi.org/sc/wa.exe?A2=ind09&L=american-scientist-open-access-forum&D=1&O=A&F=l&P=70112

2) Our experience at APS is that we have not seen significant loss of subscriptions of PRD (Physical Review D) as a result of the arXiv. PRD is the largest of the HEP journals, with currently more than 45% of the HEP literature."

4. The Commission should encourage efforts to begin measuring increased return on investment gained from open access.

Measuring return on investment from open access is a new challenge to society. Benefit to research – and thus to society – can have a payoff far into the future. Better dissemination of research information now from a team working on blood rheology may take 20 years to manifest as a new medical engineering technology for diagnosing disease. Such an example serves to illustrate both the necessity to take a long-term view when developing methods for measuring payoff and benefit, and highlights the fact that open access to research findings facilitates exactly the kind of interdisciplinary approach now becoming so much a necessary characteristic of modern science and technology.

Preliminary work has begun on modelling the cash benefit to society from open access to research findings for the research community alone (i.e. this initial work has not attempted to tackle the modelling of benefit to *all* of the constituencies, listed in point 1 above, that can benefit from open access). Initial modelling work has been carried out by Houghton (2008, 2009a, 2009b) and Houghton *et al* (2009a, 2009b). These investigators developed an economic model for calculating the benefits of open access and have so far applied it to the national situations in Australia, the UK, Denmark and The Netherlands (and are currently doing so for Germany).

In all cases, the findings are that there is a considerable cash saving from an open system of dissemination of research findings. If open access prevailed, savings over the current subscription-only system would be around €30 million *per annum* for Denmark, €50 million *per annum* for the Netherlands and €125 million *per annu*m for the UK (Houghton 2009d). Of particular interest in regard to the question posed here is the element of benefit that economists call 'return to R&D', which is effectively the savings made by increased efficiency and efficacy in research that arise from increased accessibility to information. This is the saving for the research community as a whole, (including to funders, institutions and researchers) described in monetary terms, arising from better accessibility of research information; savings derive from less duplication, reducing plagiarism, greater overall accessibility, facilitation of interdisciplinary research, etc.

These economists modelled recurring annual gains in efficiencies and accessibility from one year's research expenditure and, of course, if these changes are permanent – as they should be with the implementation of suitable policies across the ERA – then they can be converted to growth rates. Such calculations could be used by European economists to calculate the benefits to the European research economy, of open access. Moreover, the Houghton model has been adopted and adapted by Swan (2010), who has applied it to a number of universities in the UK, showing the sort of cash and efficiency savings that individual institutions could enjoy from open dissemination of research information.

This type of economic modelling is just the start of what will undoubtedly be a growing field of endeavour in economics. As well, benefit of public access to research information will come in non-economic forms, and methods to measure these benefits will need to be developed. Already, some workers are attempting to model the impact of developments in basic research fields in applied areas many years down the line (for example, Lewison, 2005), and we anticipate such approaches growing over time as accessibility matures as a friction factor in research dynamics.

5. The Commission should take steps to ensure that peer-reviewed papers arising from public investment be made available in formats that make them easy to search, find, and retrieve be linked to

There are existing standards for archiving and interoperability and any system for exposing the results of publicly-funded research must adhere to those standards if it is to achieve the objective of maximising usage and impact of that research.

(a) Interoperability

i) Search, find and retrieve:

The foundation standard for harvesting content between repositories is ***OAI-PMH*** (Open Archives Initiative-Protocol for Metadata Harvesting). This enables content to be collected from open access research repositories and amassed in others. It thus enables a system whereby original documents are deposited in local, distributed repositories based in research institutions and then these articles – or just their metadata – are harvested for use by other services. These services, such as OAIster or Google Scholar, provide the search-and-discover and retrieve functions that help would-be users to locate and access articles of interest. Both of these examples harvest content across all disciplines from institutional repositories and present a complete corpus to the user. Other services harvest specific types of content, such as all articles in a particular subject area or all electronic theses in Europe. They then present these to the user as a particular 'view' of the research corpus.

A system for making publicly-funded research publicly accessible might offer such convenient 'views' as well as constructing a comprehensive European-level collection of research and presenting that as a service, too. A number of countries[13] already have such national-level 'shop window' services: specific examples are *Intute Repository Search* in the UK[14], *JAIRO* in Japan[15] and *NARCIS* in The Netherlands[16].

ii) National access models:

There is an optimal model for how a nation could develop a system for making research publicly accessible in these ways[17]. It was developed in a study for the JISC (Joint Information Systems Committee) in the UK (Swan *et al*, 2005). The study recommends the so-called 'harvesting model', much as described above, where institutions collect the content in their local repositories and this content is harvested by national-level services that provide the 'front-end' for users. One of the advantages of such a system is that institutions benefit from collecting a complete record of their research outputs, and have a shared interest with their researchers in the enhanced visibility and impact that open accessibility brings to these local

---

[13] http://repinf.pbworks.com/Harvesters+-+national+and+international
[14] http://www.intute.ac.uk/irs
[15] http://jairo.nii.ac.jp/en/
[16] http://www.narcis.info/background
[17] http://www.openscholarship.org/jcms/c_6801/national-oa-model

collections. It is thus in their interest to ensure that content is collected. Institutions themselves also have policies on open access (there are currently 79 mandatory institutional policies), so national policies that cover research carried out under public funding could be enhanced by institutional policies that would develop alongside them if the policy recommended that universities collect the articles in their own repositories for harvesting. Not only would publicly-funded research outputs be collected, then, but also outputs from research funded by other bodies and from unfunded research.

iii) Exchange of digital objects:
A further OAI standard, ***OAI-ORE*** (Open Archives Initiative-Object Re-use and Exchange) has been developed to facilitate the exchange of items between digital collections and enable them to be re-used in their new contexts. Details of both standards are on the OAI website[18]. It is critically important that publicly-funded research outputs are stored in digital collections that operate according to this standard, otherwise they will not be maximally accessible.

It is worth noting that the use of OAI-PMH and ORE – while practicable – is not optimal because they rely on a metadata standard named DC (Dublin Core). Since 1999 this has attracted criticism from the euroCRIS community[19] involved in systems concerning research information. The European Union has recommended to member states a standard named CERIF (Common European Research Information Format) and has requested euroCRIS to maintain, develop and promote it. It is being taken up rapidly throughout Europe and has the advantage over traditional repository technologies of a formal syntax and declared semantics including full multi-character set/media type, full multimedia and full multilingual features. CERIF provides much richer metadata than DC not only for publications but also for other outputs (products, patents) and the whole research context (persons, organisations, projects, funding, events, facilities, services, equipment and so forth). CERIF can interoperate comfortably with repositories based on DC and interoperation based on OAI-PMH/OAISTER as a value-added overlay. CERIF has – for exchange and interoperation – a XML variant.

(b) Archiving: preservation is an area receiving much attention and there are a number of preservation standards for digital research material. The current picture is described in reference material developed for a project (ongoing) on repository interoperability, International Repositories Infrastructure[20]. There is a wiki for this project that contains extensive background information on the state of play of interoperability developments for digital research repositories. This information, presented under a number of categories, can be found by following links under 'Briefing Materials' from the front page of the wiki[21]. We recommend this resource is consulted if further detail is required on interoperability, since the resource is trusted and has been peer reviewed by the international repository community. We are also appending to this document map versions of these pages. . The Alliance for Permanent Access to the Records of Science[22] brings together research funders,

[18] http://www.openarchives.org/
[19] www.eurocris.org
[20] http://repinf.pbworks.com/Preservation
[21] http://repinf.pbworks.com/
[22] www.alliancepermanentaccess.eu

publishers, national libraries, large research institutions with publications and data/software to address curation and preservation.

In our view, it is unlikely that these standards will change materially, though it is highly likely that refinements and modifications will be developed over time as new needs arise and better ways of serving needs are identified.

New developments in the field of interoperability include work on unique identifiers (for authors, institutions, and objects), and technologies to improve ingest of items into repositories, including enabling deposit into more than one repository at a time, thus saving authors from having to make multiple deposits. These sorts of technological developments will undoubtedly lead to new standards in future.

Our overall opinion on these technical issues is, though, that the answers to these questions do not need to be determined for the purpose of writing law, or even mandated across the board. Some experimentation should be encouraged so that the different agencies can find the most appropriate solutions for their needs. Systems may well evolve over time accordingly.

5. Access demands not only availability, but also meaningful re-usability

i) Use by machines:

If the opportunity for re-use is to be maximised, peer-reviewed papers need to be available in non-PDF format. The ideal format is XML, and mark-up is best done at source. Most authors are not yet accustomed to providing XML documents, however, though the newest edition of Microsoft's Office software produces XML output as default, so the situation will improve. XML conversion can be done after the paper has been deposited by the author. The practice at UK PubMed Central (UKPMC), for example, is for XML conversion (plus additional mark-up to enable particular text-mining technologies that have been developed specifically for application on the biomedical research literature) to be carried out after deposit wherever needed. The process has a cost attached, but the UKPMC sponsors consider that the payoff, in terms of optimising re-use potential, is worth the cost.

An open access system for publicly-funded research might implement a similar arrangement since it has the merit of standardising the technical presentation of documents and optimising this for re-use. We anticipate that technical standards for maximising re-use may diverge between different disciplines over time as semantic mining technologies develop to best serve the needs of those disciplines. However, such a system should interlink strongly with CRIS (Current research Information Systems) based on CERIF (Common European Research Information Format) which provides contextual (meta)data allowing more effective and efficient machine management of the information.

ii) Use by people:

Use can be measured in a number of ways, including by metrics such as download numbers, visitor numbers and page views. The Commission should create a system of monitoring use of European public access collections that provides as much information about users as possible without contravening data protection and privacy rules. For example, most university repositories collect data on page views and download numbers both for the whole collection, and by author and article.

Some of our own repositories additionally record:

- the domain from which users originate
- the referrer (which service refers the user to the article (e.g. Google or a certain blog posting, etc)
- the key words or search terms that the user employed in their search algorithm
- the time of the 'hit' on an article
- the pattern of hits for the whole collection and for each article (thus providing useful information about capacity requirements for the system as well as for more general usage monitoring)

And of course with an interconnected CRIS then much additional information about the research is available providing a rich contextual content to assist the end-user.

It is possible also to add user feedback in a variety of ways and these will increase and change with time. The high-energy physics arXiv permits 'trackbacks' where blog comments can be linked to from the article in arXiv that sparks the comment[23] though this system has not been without problems or criticism. Inclusion of the facility to add comments by users to items in the ERA collection is worth considering, though due attention should be paid to the issue of quality control for such a system. It may become burdensome or controversial. The arXiv uses a team of editors to verify the validity of any Trackbacks (not without controversy) before allowing them to be added to the database and it would seem that some sort of verification system would also need to be used to ensure quality and propriety of open commentary systems for a corpus of ERA research outputs.

This submission
Submitted to the European Commission by the Board of *Enabling Open Scholarship*, January 2010:

Professor Bernard Rentier (chair), Rector of the University of Liège, Belgium
Professor Tom Cochrane, Deputy Vice Chancellor, Queensland University of Technology, Brisbane, Australia
Dr William Dar, Director General of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India
Professor Martin Hall, Vice Chancellor, University of Salford, UK
Professor Stevan Harnad, Canada Research Chair, Université du Québec à Montréal (UQAM), Montreal, Canada
Professor Keith Jeffery, Director of IT and International Strategy at the Science & Technology Facilities Council, Swindon, UK
Dr Sijbolt Noorda, President of VSNU, the Association of Dutch Research Universities, The Hague, Netherlands
Professor Stuart Shieber, James O. Welch, Jr. and Virginia B. Welch Professor of Computer Science, School of Engineering and Applied Sciences at Harvard University and Director of Harvard's Office of Scholarly Communication, Cambridge, USA

[23] http://arxiv.org/help/trackback/

Professor Ian Simpson, Deputy Principal for Research and Knowledge Transfer, and Professor of Environmental Science, University of Stirling, UK
Professor Peter Suber, Berkman Center for Internet & Society, Harvard University, Cambridge, USA
Professor John Willinsky, Khosla Family Professor of Education at Stanford University and director of the Public Knowledge Project at the University of British Columbia and Simon Fraser University
Dr Alma Swan (Convenor/Coordinateur), Director of Key Perspectives Ltd, Truro, UK

Appendix 1: research funders with mandatory policies on Open Access:

Australia	Australian Research Council
Australia	National Health & Medical Research Council
Austria	Fonds zur Foerderung der wissenschaftlichen Forschung
Belgium	Research Foundation Flanders
Canada	Michael Smith Foundation for Health Research
Canada	Canadian Health Services Research Foundation
Canada	Fonds de la recherche en sante Quebec (FRSQ)
Canada	Canadian Breast Cancer Research Alliance
Canada	National Research Council (NRC) Canada
Canada	Ontario Institute for Cancer Research (OICR)
Canada	Canadian Cancer Society Research Institute
Canada	Canadian Institutes of Health Research (CIHR)
Europe	European Commission
Europe	European Research Council
France	Agence Nationale de la recherche (ANR) (Humanities and Social Sciences Branch)
Hungary	Hungarian Scientific Research Fund (OTKA)
Ireland	Science Foundation Ireland (SFI)
Ireland	Irish Higher Education Authority (HEA)
Ireland	Irish Research Council for Science, Engineering & Technology
Norway	Norwegian Research Council
Spain	Government of the Principality of Asturias
Spain	Madrid Autonomous Community of Spain
Sweden	Swedish Research Council
Switzerland	Swiss National Science Foundation (SNF)
UK	Engineering & Physical Sciences Research Council (EPSRC)
UK	JISC (Joint Information Systems Committee)
UK	Arts and Humanities Research Council (AHRC)
UK	Arthritis Research Foundation
UK	British Heart Foundation
UK	Cancer Research UK
UK	Chief Scientist Office (Scottish Executive Health) Department
UK	Department of Health (UK)
UK	Science & Technology Facilities Council
UK	Natural Environmental Research Council (NERC)
UK	Biotechnology and Biological Sciences Research Council (BBSRC)
UK	Economic and Social Research Council (ESRC)
UK	Medical Research Council (MRC)
UK	Wellcome Trust
USA	Institute of Education Science
USA	Autism Speaks
USA	National Institutes of Health (NIH)
USA	Howard Hughes Medical Institute (HHMI)

References

CEPA (Cambridge Economic Policy Associates) (2008) Activities, costs and funding flows in the scholarly communications system in the UK: http://www.rin.ac.uk/costs-funding-flows

CIBER (2004) *Scholarly communication in the digital environment: What do authors want?* CIBER Group, UCL, London. www.publishing.ucl.ac.uk/papers/2004Rowlands_etal.pdf

Gentil-Beccot A, Mele S and Brooks T (2009) Citing and Reading Behaviours in High-Energy Physics. How a community stopped worrying about journals and learned to love repositories. http://arxiv.org/abs/0906.5418

Henneken E, Kurtz M, Eichhorn G, Accomazzi A, Grant C, Thompson D and Murray S (2006) Effect of E-printing on Citation Rates in Astronomy and Physics http://arxiv.org/abs/cs/0604061

Houghton, J *et al* (2008) Economic Implications of Alternative Scholarly Publishing Models: Exploring the costs and benefits. http://www.jisc.ac.uk/publications/documents/economicpublishingmodelsfinalreport.aspx

Houghton, J., Rasmussen, B. and Sheehan, S.; with Oppenheim, C., Morris, A., Creaser, C., Greenwood, H., Summers, M. and Gourlay, A. (2009a), Economic implications of Alternative Scholarly Publishing Models: Exploring the Costs and Benefits, JISC EI-ASPM Project, Report to the Joint Information Systems Committee (JISC) (UK), CSES and Loughborough University, January 2009.
http://www.jisc.ac.uk/publications/documents/economicpublishingmodelsfinalreport.aspx
http://alpsp.publisher.ingentaconnect.com/content/alpsp/lp/2007/00000020/00000002/art00002
http://www.jisc.ac.uk/whatwedo/topics/opentechnologies/openaccess/reports/learnedsociety.aspx

Houghton, JW, de Jonge, J and van oploo, M (2009b) Costs and benefits of research communication: the Dutch situation. SURF Foundation, Utrecht. http://www.surffoundation.nl/en/publicaties/Pages/CostsandBenefitsofOpenAccessPublicationlTheDutchSituation.aspx

Houghton, JW (2009a) Costs and benefits of alternative publishing models: Denmark. Danish Electronic Research Library (DEFF), Copenhagen. http://www.deff.dk/content.aspx?itemguid={EACA73FB-2EFE-44CA-92CD-91C4416C0370}

Houghton J (2009b) Open Access – what are the economic benefits? A comparison of the United Kingdom, Netherlands and Denmark. A report to Knowledge Exchange. Knowledge Exchange, Copenhagen.
http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fOA_What_are_the_economic_benefits_-_a_comparison_of_UK-NL-DK__FINAL_logos.pdf

Jubb M, Look H and Sparks S. (2007) Researchers' use and perceptions of discovery services. *Learned Publishing* **20 (2),** pp147-153.

RIN (2009) Overcoming barriers: access to research information content. Research Information Network, London. http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/overcoming-barriers-access-research-information

Sale, AHJ (2006) Comparison of IR content policies in Australia. *First Monday*, 11 (4). http://eprints.utas.edu.au/264/

Swan, A (2008a) The 'Big Picture' and researchers' top concerns about the scholarly communication process. A report for the JISC Scholarly Communications Group. http://eprints.ecs.soton.ac.uk/15450

Swan A (2008b) Study on the availability of academic 'grey literature' to UK SMEs. Report to the JISC Scholarly Communications group. http://eprints.ecs.soton.ac.uk/17667/

Swan A (2010) Modelling scholarly communication transition: costs and benefits for universities. In press, to be published early February 2010 by the Joint Information Systems Committee, London, UK.

Swan, A and Brown, S (2008a) To share or not to share: Publication and quality assurance of research data outputs. Report to the Research Information Network. http://eprints.ecs.soton.ac.uk/16742/

Swan, A and Brown, S (2008b) The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs. A report to JISC. http://eprints.ecs.soton.ac.uk/16675/

Ware, M. (2009) Access by UK small and medium-sized enterprises to professional and academic literature. Publishing Research Consortium, London. http://www.publishingresearch.net/documents/SMEAccessResearchReport.pdf

The UK Research Data Service Feasibility Study: report and recommendations to HEFCE (2008). http://www.ukrds.ac.uk/resources/download/id/16