Economic and Social Returns on Investment in Open Archiving Publicly Funded Research Outputs

Report to SPARC

By

John Houghton

With

Bruce Rasmussen and Peter Sheehan

July 2010

Centre for Strategic Economic Studies
Victoria University
(http://www.cfses.com)

Acknowledgements

The authors would like to acknowledge the support of the UK Joint Information Systems Committee (JISC), the Netherlands SURFfoundation, Denmark's Electronic Library (DEFF), Knowledge Exchange and the German Research Foundation (Deutsche Forschungsgemeinschaft) during the development and evolution of the modeling approach underpinning this study, and the Scholarly Publishing and Academic Resources Coalition (SPARC) for enabling its further development and application in the US. Thanks are also due to members of the various research teams from the previous studies.

Disclaimer

While every effort has been made to ensure its accuracy, the Centre for Strategic Economic Studies and Victoria University make no representations or warranties (express or implied) as to the accuracy or completeness of the information contained in this report. Victoria University, its employees and agents accept no liability in negligence for the information (or the use of such information) provided in this report.

Foreword

The Scholarly Publishing and Academic Resources Coalition (SPARC) provided support for a feasibility study, to outline one possible approach to measuring the impacts of the proposed US *Federal Research Public Access Act* (FRPAA) on returns to public investment in R&D. The aim is to define and scope the data collection requirements and further model developments necessary for a more robust estimate of the likely impacts of the proposed FRPAA open archiving mandate.

Preliminary modeling suggests that over a transitional period of 30 years from implementation, the potential incremental benefits of the proposed FRPAA archiving mandate might be worth around 8 times the costs. Perhaps two-thirds of these benefits would accrue within the US, with the remainder spilling over to other countries. Hence, the US national benefits arising from the proposed FRPAA archiving mandate might be of the order of 5 times the costs.

Exploring sensitivities in the model we find that the benefits exceed the costs over a wide range of values. Indeed, it is difficult to imagine any plausible values for the input data and model parameters that would lead to a fundamentally different answer.

These preliminary estimates are based on the information available to us at the time of writing. They are released in conjunction with an online model, which enables others to explore their own preferred values for the various parameters.

The model and this report can be found at http://www.cfses.com/FRPAA/.

Contents

FOREWORD .... **III**

CONTENTS .... **IV**

BACKGROUND AND AIMS **1**

THE MODEL AND ITS OPERATIONALIZATION **1**

PRELIMINARY RESULTS **4**

MODELED IMPACTS ON RETURNS TO R&D 4
COMPARING COSTS AND BENEFITS 5
POTENTIAL IMPACTS OF AN OPEN ARCHIVING MANDATE 7
SENSITIVITY IN THE MODEL 8

NEXT STEPS **9**

ANNEX I AN OUTLINE OF THE MODEL **12**

RETURNS TO R&D IN A SIMPLE SOLOW-SWAN MODEL 12
CONTRIBUTIONS TO GROWTH AND TOTAL FACTOR PRODUCTIVITY 13
ESTIMATING THE RATE OF RETURN TO R&D 14
INCORPORATING THE EFFICIENCY OF RESEARCH AND ACCESSIBILITY OF KNOWLEDGE 14
SOME METHODOLOGICAL NOTES ON THE MODEL 16

ANNEX II OPERATIONALIZING THE MODEL **19**

RETURNS TO R&D: SOURCES AND RATIONALE FOR THE RANGE TO BE MODELED 19
ACCESSIBILITY: SOURCES AND RATIONALE FOR THE RANGE TO BE MODELED 22
EFFICIENCY: SOURCES AND RATIONALE FOR THE RANGE TO BE MODELED 28
OTHER PARAMETERS: SOURCES AND RATIONALE FOR THE RANGE TO BE MODELED 30
DATA: SOURCES AND RATIONALE FOR THE BASE CASE VALUES 32

REFERENCES **34**

Background and aims

Public funding of scientific, technical and medical (STM) research is undertaken with the expectation that the economic and social returns to taxpayers will exceed the amount of the research investment. Because discovery is a cumulative process, with new knowledge building on earlier findings, the dissemination of research findings is crucial to ensuring that the returns on the investment are realized.

Traditionally, journals have been one of the primary channels for research dissemination. With the emergence of the Internet, it became possible to expand the sharing of research findings and thus to better serve scientists as well as the "long tail" of other potential users – such as educators and students, health clinicians and patients, businesses and the general public. It is now technically feasible to put knowledge to use far beyond the limited universe served by traditional toll-access or subscription journals. It is also possible for knowledge – research articles and other research outputs – to be used and integrated in new ways that further advance public purposes.

Responding to these opportunities, a growing number of public and private funders have implemented policies mandating deposit of their funded research outputs in open online archives, making it freely available to anyone with Internet access. While there is ample anecdotal evidence of the benefits of such policies, the extent of the resulting leverage or its relationship to costs has rarely been measured. As the availability of cost-benefit measures and quantitative data would inform public policymaking, this project seeks to identify metrics for demonstrating the return on investment in open access dissemination of publicly funded research. We outline one possible approach and identify areas in which further data collection is needed. This is intended to define and scope the data collection requirements and further model developments necessary to produce a more robust estimate of the potential impacts of an archiving mandate for publicly funded research outputs, such as the proposed *Federal Research Public Access Act* (FRPAA).

The following sections outline the model and data sources used for preliminary estimates. These estimates are based on the information available to us at the time of writing. They are released in conjunction with an online model, which enables others to explore their own preferred values for the various parameters. The model and this report can be found at http://www.cfses.com/FRPAA/.

The model and its operationalization

The task of fully quantifying the costs and benefits of the proposed FRPAA archiving mandate is daunting, but it is possible to gain some sense of the potential scale of impacts (Houghton *et al.* 2006; Houghton and Sheehan 2009; Houghton and Oppenheim *et al*. 2009).

The standard Solow-Swan model makes some key simplifying assumptions, including (*i*) that all R&D generates knowledge that is useful in economic or social terms (*the efficiency of R&D*), and (*ii*) that all knowledge is equally accessible to all entities that could make productive use of it (*the accessibility of knowledge*). Obviously, these assumptions are not realistic. In the real world, there are limits and barriers to access and limits to the usefulness of knowledge. So, we introduce *accessibility* and *efficiency* into the standard model as negative or friction variables, and then look at the impact on returns to R&D of reducing the friction by increasing *accessibility* and *efficiency.* Annex I presents details of the basis and development of the model.

Table 1: Summary of the base case parameter sources and values

Parameter	***Basis***	***Value***
ACCESSIBILITY		
Percentage change in accessibility (Reported access gaps)	Ware (2009): 10% to 20% of articles read presented access difficulties	Adjusting for share of difficulties due to toll access barriers
Percentage change in accessibility (OA citation advantage)	Hajjem *et al*. (2005): 25% to 250% more citations Gargouri *et al*. (2010); Zhang (2006): average 100%	Adjusting for what is already OA and articles as a share of the research stock of knowledge
Percentage change in accessibility (OA download advantage)	Davis *et al*. (2008): 42% to 89% more pdf and full text downloads (average 66%)	Adjusting for what is already OA and articles as a share of the research stock of knowledge
Combined estimate		*Taking the lower bound of the ranges above: Estimate for model 4.68%*
EFFICIENCY		
Percentage change in efficiency (wasteful expenditure: duplicative research and blind alleys)	Scenario 1, for illustrative purposes	1%
Percentage change in efficiency (new opportunities: collaborative opportunities and new methods)	Scenario 2, for illustrative purposes	1%
Percentage change in efficiency (Research time savings)	Scenario 3, for illustrative purposes	..
Combined estimate		*In the absence of a grounded metric 0%*
OTHER PARAMETERS		
Returns to R&D (per cent)	Conservative consensus from the literature (Arundel and Geuna (2003; Hall *et al*. 2009; etc.)	20% to 60% (estimate 20%)
Local share of returns to R&D (per cent)	Consensus from the literature (Jaffe 1989; Coe and Helpman 1995; Verspegan 2004, etc.), and national citation patterns (NSB 2010)	66%
Lag between R&D spending and impacts (years)	Mansfield (1991, 1998; Matsumoto 2008)	3 years to publication plus 7 years to impact, 10 years
Distribution of impacts (years)	Mansfield (1991, 1998; Sveikauskas 2007; Matsumoto 2008)	Normal distribution over 10 years
Depreciation of stock of research knowledge (per cent)	BLS method (Griliches 1995; Hall 2009; Sveikauskas 2007)	Applying the BLS method, estimate 8%
Discount rate / risk premium (per cent)	Conservative consensus from literature	10% per annum

Source: Authors' analysis (See Annex II).

To operationalize the model it is necessary to establish values for the *accessibility* and *efficiency* parameters, as well as rates of return to R&D and of depreciation of the underlying stock of research knowledge. Annex II presents details of the model's operationalization and explains the sources and rationale for the choice of base case values (Summarized in Table 1).

For the purposes of preliminary analysis we take 4.68% as a conservative estimate of the potential increase in *accessibility*. To put that into perspective, Ware (2009) reported that the equivalent of 10% to 20% of articles read by his survey respondents presented access difficulties, and on average across the studies reviewed citations doubled and downloads increased by around 66% when articles were made openly accessible (See Annex II).

Table 2: Summary of the base case data sources and values

Parameter	***Basis***	***Value***
Federal R&D Spending (USD billions)	NSB 2010 indicators: R&D expenditure by the 11 FRPAA departments in 2008	$61
Annual growth in federal R&D spending (per cent)	NSB 2010 indicators: reported growth over last 10 years	3.2%
Average annual salary of researchers (USD)	NSB 2010 indicators: reported average salaries in 2008	74,070
Number of articles published from federal R&D (2008)	NIH 2008: estimate based on the ratio of NIH expenditure to article output	170,000
Number of articles published from NIH funded research circa 2008	NIH (2008, p22)	80,000
Average annual growth in article output (per cent)	NSB 2010 indicators: over last 10 years	1.8%
Per article submission-based costs (USD)	ArXiv (2010)	$7
Per article submission-based costs (USD)	NIH (2008, p22)	$59
Per article life-cycle archiving cost in first year (USD)	LIFE^2 Project: Year 1 life-cycle costs	$34
Per article life-cycle costs per year in subsequent years (USD)	LIFE^2 Project: Subsequent year annual life-cycle costs	$12
Time for author deposit (minutes per article)	Reported average use of the NIHMS submission system (NIH 2008, p14)	10 mins
Annual growth in archiving costs (per cent)	BLS: Average US CPI over last 10 years	3%
Average level of compliance with mandate over 30 years (per cent)	Assumed full compliance for the base case	100%
Embargo period (months)	Assumed six month embargo for the base case	6

Source: Authors' analysis (See Annex II).

The third piece of the puzzle is the input data required for the modeling. The main requirements include the implied archiving costs, the volume of federally funded research outputs (*i.e.* articles), the levels of federal research funding and expenditure trends. For the purposes of preliminary analysis we have used publicly available sources and published estimates, and where necessary have derived estimates of our own from them. Annex II presents details of the data sources (Summarized in Table 2).

Data relating to federal research funding, activities and outputs are taken from the most recent National Science Board *Science and Engineering Indicators 2010* (NSB 2010). It should be noted that FRPAA agencies' funded article output is an estimate based on the ratio of NIH funding to articles produced and may overstate article output and, thereby, inflate archiving costs and lead to an underestimate of net benefits. We explore three sources of archiving costs:

- The LIFE^2 Project (Ayris *et al*. 2008), which reported life-cycle costs for articles and other items held on institutional archives in the UK, and found costs equivalent to up to $34 per article in the first year, and $12 per article held per annum in subsequent years;
- Reporting costs on a submissions equivalent basis, NIH (2008) estimated that it would cost $4.5 million per annum to host the estimated 80,000 articles from NIH funding circa 2008 and noted that they had spent a further $250,000 on policy-related staff costs, implying a per article cost of around $59 per submission; and
- Also reporting approximate costs on a submissions equivalent basis, arXiv (2010) noted that their annual budget was $400,000 rising to $500,000 by 2012 and that 64,047 articles had been submitted in 2009, implying a per article cost of around $7 per submission.

For the purposes of producing preliminary estimates, we explore this range of costs – noting that the mid-range NIH reported costing might be the best guide.

Preliminary results

This section presents the results from our preliminary modeling based on the base case values outlined above. These are not intended to provide a definitive answer, but rather to test the feasibility of the approach, define and scope the data collection requirements and further model developments necessary for more robust estimates of the potential impacts of an open access archiving mandate for federally funded research, such as that proposed in the FRPAA.

Modeled impacts on returns to R&D

Table 3 presents the preliminary modeled estimates of the impacts of a one-off increase in *accessibility* and *efficiency* on returns to 2008 R&D spending based on total federal expenditure on R&D and R&D expenditure by the 11 FRPAA agencies, with percentage changes in *accessibility* and *efficiency* shown cumulatively. For illustrative purposes, we present ranges of rates of social return to R&D of 20% to 60% (Arundel and Geuna 2003) and increases in accessibility and efficiency of 1% to 10%.[1]

With a 20% return to FRPAA agency R&D spending of $61 billion in 2008, a single percentage point increase in *accessibility* and *efficiency* would have been worth around $245 million (per

[1] We assume that a change in accessibility and efficiency would have no *net* impact on the rates of accumulation and obsolescence of the stock of R&D knowledge.

annum) in increased returns to R&D. Of this total, perhaps some $160 million might be expected to accrue within the US, with the remainder spilling over to other countries.

Table 3: Estimates of the impacts of a one-off increase in accessibility and efficiency on returns to R&D (USD millions, 2008)

Federal Funded R&D $139 billion	Rate of return to R&D				
	20%	30%	40%	50%	60%
Per cent change in accessibility and efficiency	Recurring annual gain from increased accessibility & efficiency (million)				
1%	560	840	1,120	1,400	1,680
2%	1,125	1,688	2,251	2,813	3,376
5%	2,855	4,283	5,710	7,138	8,566
10%	5,850	8,775	11,700	14,624	17,549

FRPAA Agency R&D $61 billion	Rate of return to R&D				
	20%	30%	40%	50%	60%
Per cent change in accessibility and efficiency	Recurring annual gain from increased accessibility & efficiency (million)				
1%	246	369	492	615	737
2%	494	741	988	1,235	1,482
5%	1,253	1,880	2,507	3,134	3,760
10%	2,568	3,852	5,136	6,420	7,704

Source: Authors' analysis.

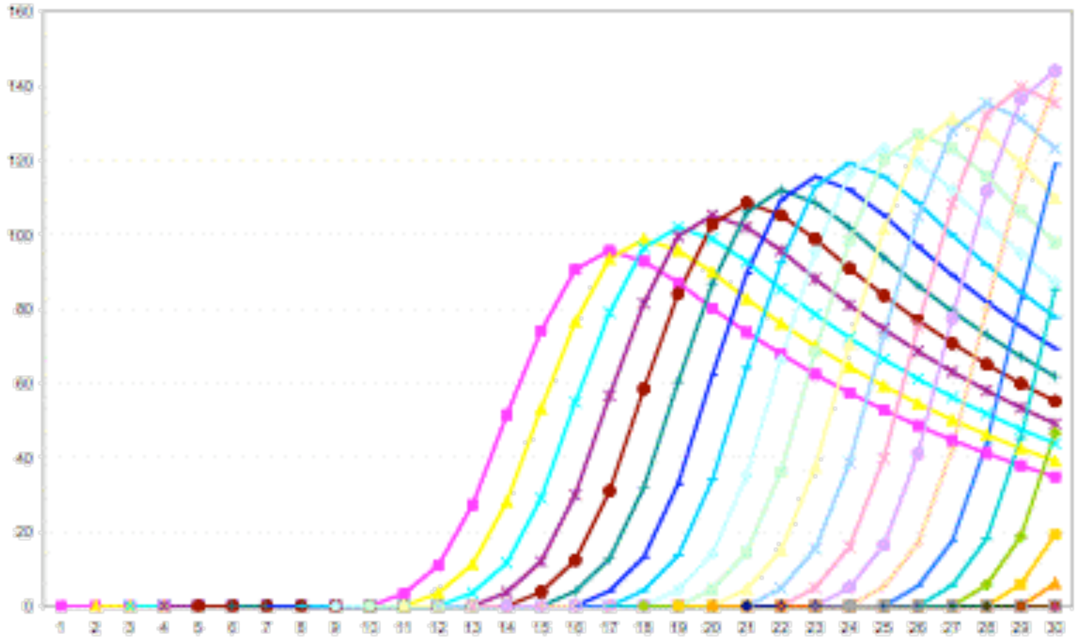
These are recurring annual gains from the effect of one year's R&D spending, such that if the change that brings the increases in *accessibility* and *efficiency* is permanent (*e.g.* the adoption of open archiving as proposed in the FRPAA) they can be converted to growth rate effects.

Comparing costs and benefits

In this section we attempt to compare the costs and benefits associated with the proposed FRPAA archiving mandate. Details of the model and its operationalization can be found in Annexes I and II.

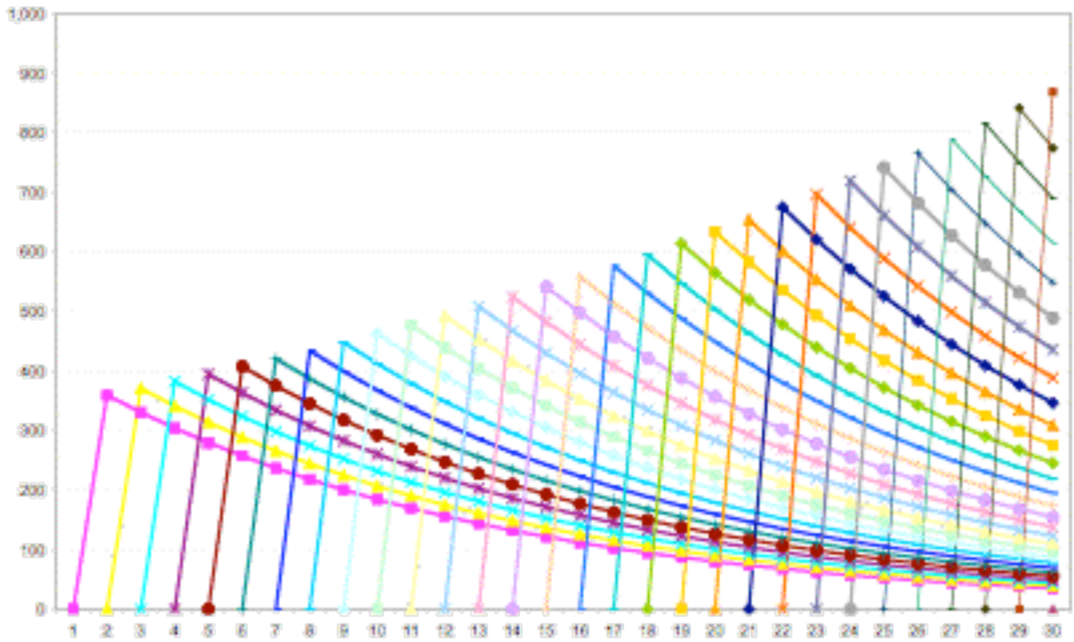
One thing to note is that we are modeling the transitional impact of open access archiving over 30 years. Because of the lag between research expenditure and the realization of economic and social returns to that research, the impact on returns to R&D is lagged (by 10 years in the base case) and the value of those returns are discounted accordingly. This reflects the fact that the impacts of open access archiving would be prospective and not retrospective, and that the economic value of impacts of enhanced *accessibility* and *efficiency* would not be reflected in returns to R&D until those returns are realized. Put simply, this has the effect that over a transitional period of 30 years we are comparing 30 years of costs with 20 years of benefits (Figure 1).

Figure 1: Indicative distribution of impacts over a transitional period of 30 years (USD millions in years 1 to 30)



Source: Authors' analysis.

Figure 2: Indicative distribution of impacts over 30 years in a simulated steady state period (USD millions in years 1 to 30)



Source: Authors' analysis.

An alternative approach would be to model a hypothetical 'steady-state' system in which the benefits of historical increases in *accessibility* and *efficiency* enter the model in year one. This would reflect the situation in an alternative system, after the transition had worked through and was no longer lagging returns to R&D. Put simply, in such a model one would be comparing 30 years of costs with 30 years of benefits (Figure 2).

We took the view that it was more realistic and of more immediate concern to model the transition. Nevertheless, it must be emphasized that a transitional model returns significantly lower benefit/cost ratios than would a hypothetical alternative 'steady-state' model.

Potential impacts of an open archiving mandate

The base case model parameters, their sources and rationale are outlined in Table 1 (above) and Annex II, and input data values are summarized in Table 2 (above) and Annex II. With these base case values, we model the impacts relating to R&D spending by the 11 departments affected by the FRPAA using reported arXiv, NIH and the upper bound LIFE[2] lifecycle archive costings. These archive costings vary significantly, but the mid-range NIH costing might provide a reasonable guide.

Table 4: Modeled estimates for the base case parameters (USD millions over 30 years in Net Present Value & Benefit/Cost Ratio)

Transitional Model	*Federal R&D (arXiv costing)*	***Federal R&D (NIH costing)***	*Federal R&D (LIFE[2] costing)*
Incremental Impacts			
Cost over 30 years (NPV)	68	**206**	400
Benefits over 30 years (NPV)	1,626	**1,626**	1,626
Worldwide benefit/cost	24	**8**	4
Local benefits over 30 years (NPV)	1,073	**1,073**	1,073
US national benefit/cost	16	**5**	3
Overall Impacts			
Cost over 30 years (NPV)	68	**206**	400
Benefits over 30 years (NPV)	2,587	**2,587**	2,587
Worldwide benefit/cost	38	**13**	6
Local benefits over 30 years (NPV)	1,707	**1,707**	1,707
US national benefit/cost	25	**8**	4

Notes: Using the base case parameters outlined in detail in Annex II and assuming a six-month embargo and 100% compliance with the proposed mandate.
Source: Authors' analysis.

Over a transitional period of 30 years from implementation, the potential *incremental* benefits of an open access archiving mandate for all FRPAA agencies' funded R&D[2] might be worth around $1.6 billion (Net Present Value), around 4 times the estimated cost using the higher end

2 Taking account of the share of articles that are already openly accessible through the NIH and other mandates.

lifecycle costing, 8 times the cost using NIH costing and more than 24 times the cost using arXiv costing. Perhaps some $1 billion of these benefits would accrue within the US, with the remainder spilling over to other countries. Hence, the US national benefits might be around 5 times the costs.[3] The overall impacts of openly archiving all FRPAA agencies' funded R&D article outputs would be greater than these incremental impacts, with likely US national benefits of around 8 times the costs (Table 4).

These estimates assume a six-month embargo period between publication and open accessibility. If there were no embargo, we estimate that incremental returns might be closer to $1.75 billion. Hence, a six-month embargo reduces the returns by around $120 million (NPV). Of course, the impact of an embargo delaying open accessibility will vary significantly between fields of research and disciplines, having greater impact in faster moving fields of research and practice than in those where the progress of knowledge, application and practice is slower.

It should be noted that these estimates are based on increased returns to R&D through increased accessibility and take no account of the potential activity cost impacts of more open access to federally funded research (*e.g.* possible savings in such areas as researcher time spent in search, discovery and access) or of potential efficiency impacts (*e.g.* in reducing duplicative research). They also focus on the transitional period following implementation and, as noted above (Figures 1 and 2), once established the benefits of open archiving would be substantially greater than immediately following implementation and during the transition. In a hypothetical 'steady-state' scenario, for example, estimated US national benefits might be more than 50 times the cost.

However, it should be noted that benefits in the form of increases in returns to R&D are diffuse in nature, occur throughout the economy and, indeed, throughout the world. They also accrue over time, sometimes lagging research expenditure and publication by many years. In contrast, the costs are both local and immediate. Hence, the costs must be met up-front, in order to maximize the return on public investment in research. It should also be noted that these estimates are preliminary in nature, intended to test the feasibility of the approach, define and scope the data collection requirements and further model developments necessary for more robust estimates of the possible impacts of an open archiving mandate for federally funded R&D. They come with many caveats (See Annexes I and II for details).

Sensitivity in the model

Among the caveats is the model's sensitivity, which we examine here in order to prioritize areas for further data collection and model development. Using the mid-range NIH reported archiving costs and changing individual parameters one-at-a-time we find that:

- The number of articles produced is an important driver of archiving costs. Nevertheless, at base case values the US national benefits would exceed the costs if

[3] It should be noted that these estimates are based on the most conservative assumptions (*e.g.* lower-bound values for returns to R&D and increases in accessibility). As such, they are likely to reflect the lower end of the benefits that might be expected.

current annual article output where more than one million – almost 6 times current article output – at the same level of R&D funding.

- Archiving costs are also important and further work is required to establish exactly what those costs would be. Nevertheless, the base case model returns net national benefits with per article submission costs of more than $375 (excluding author deposit costs).
- The potential increases in accessibility and efficiency resulting for open archiving are an important input, as the greater the increases the greater the benefits. However, a combined total change in accessibility and efficiency of less than 1% returns net benefits.
- The average rate of social return to publicly funded R&D is an important parameter, but average rates of as low as 4% produce US national net benefits.
- The rate of depreciation of the stock of research knowledge is set according to the established formula used by the US Bureau of Labor Statistics, and while the rate chosen makes a substantial difference (with lower rates producing higher benefits) the benefit/cost ratios remain greater than 1 for rates higher than 16% per annum – a rate that is higher than conventionally used for publicly funded R&D.
- The lag between research spending and its economic impacts and the distribution of those impacts over time have been set to work in the range of 5 to 15 years average lag, with a normal distribution of impacts over approximately 10 years. The shorter the lag the greater are the impacts, but a lag of 15 years still gives benefits that are more than double the costs.

Overall, even during a transitional period, the benefits appear to exceed the costs over a wide range of values and it is difficult to imagine any plausible values for the input data and model parameters that would lead to a fundamentally different answer.

Next steps

This section explores some of the key areas for data collection and refinement, and suggests possible next steps.

Levels of federally funded R&D activity and publication counts

While there are good data on the levels and types of federal R&D spending, we have not yet been able to fully match these to outputs (*i.e.* the number of articles produced from federally funded R&D by field and by funding agency), nor have we been able to explore the rate of growth of article outputs by field and by funding agency.

Further work in this area might involve conducting a more thorough review of existing data sources and, where necessary, undertaking targeted consultation with funding agencies to establish more informed estimates. The key issue is to obtain R&D expenditure and article output data relating to the specific agencies affected by the proposed FRPAA open archiving mandate.

Archiving costs and practices

Archiving costs are a key input. Unfortunately, relatively little is known about archiving and preservation costs, and what is known suggests that they vary greatly from case to case (*e.g.* centralized versus institutional archives, by field of research, etc.). Our initial estimates are based on the most recent and detailed studies available (Ayris *et al*. 2008, NIH 2008 and arXiv 2010). We have also included a cost for author deposit, based on NIH reporting on NIHMS experiences (NIH 2008) and recent UK studies (Houghton and Oppenheim *et al*. 2009; Swan 2010a). These costs will also vary with archiving practices (*e.g.* author deposit versus automated publisher submission processes).

Further efforts in this area might involve working towards better quantifying the costs of offering persistent access to US federally funded research outputs in open archives, building on existing information on archiving costs by conducting a more thorough review of published sources on archiving costs around the world, with a special focus on reported costs in the US, and consultation with a representative sample of archive operators and managers in the US, in order to refine preliminary estimates to ensure that the final estimates are representative of the potential mix of archives that might be required.

Accessibility and efficiency metrics

The potential increases in *accessibility* and *efficiency* resulting from an open archiving mandate, such as that proposed by the FRPAA, are also key parameters. Our initial estimates are based on studies reporting access gaps and possible open access citation and download advantages, and to be conservative we use the lower bound impacts reported. While these provide some foundation for preliminary estimates, further data collection and the development of additional metrics are required before more robust estimates can be made.

In relation to *accessibility*, further work might involve: (*i*) undertaking more focused surveys of research users in the US and elsewhere to better establish the extent and significance of the access difficulties and gaps they face in accessing journal articles of the type emerging from federally funded R&D; and (*ii*) undertaking a fuller review of studies of the possible citation and download advantages resulting from the open online accessibility of research articles, with a focus on what proportion might be a sustainable advantage.

In relation to *efficiency*, further work might focus on identifying examples of efficiency gains through consultation with experts (*e.g.* rejection of journal and conference papers or funding applications because the work is duplicative, case study examples of the pursuit of blind alleys due to incomplete information, known examples of unnecessarily duplicative research, etc.).

As a part of the consultation with archive operators and managers in the US, it would also be worthwhile asking if they can identify particular cases where the use of openly available research articles and/or data has had an impact and, where possible, follow up on the examples to ascertain the extent of the impacts.

Modeling parameters

While base case values have been sourced from the literature and are well grounded, further work on developing and refining the model might focus on the underlying evidence base for key parameters, including:

- The average rate of social return to publicly funded R&D, taking account of the mix of fields (*e.g.* the large share of life-science in US federally funded R&D, which would suggest higher rates of return that might otherwise the be case);
- The localization of returns and likely level of international spillovers;
- The average lags (*i*) between R&D funding and publication/archiving, and (*ii*) between publication/archiving and the realization of returns;
- The distribution of returns over time; and
- The most appropriate rate of depreciation to apply to the stock of R&D knowledge.

System cost impacts

As well as direct cost impacts, enhanced accessibility is likely to have indirect, intended and unintended impacts on the cost of research and scholarly communication activities. These might include such things as increased research library costs as librarians are faced with demands to help users with an additional information channel, and declining publisher revenues if archiving were to lead to subscription cancellations. Hence, in addition to the analysis outlined above, it would be desirable to explore both the direct and indirect, system-wide cost impacts of open archiving.

This might involve building on the activity modeling and costing approach used by Houghton and Oppenheim *et al*. (2009) and in subsequent studies, who looked at the system-wide cost implications of alternative scholarly publishing models. Such an approach requires detailed research activity data, research library and funder information, and while there are a number of existing sources for such information it is likely that some additional data collection would be required in the pursuit of such an approach.

***

In each of these cases, to further prioritize efforts the reader should consider the relative strengths and weakness of the sources used in the preliminary analysis presented herein.

Annex I An outline of the model

It is possible to gain some sense of the scale of potential costs and benefits arising from open archiving federally funded R&D articles output by using a modified Solow-Swan model (Houghton *et al.* 2006; Houghton and Sheehan 2009; Houghton and Oppenheim *et al*. 2009). This section presents an outline of the basis and development of the model used.

Returns to R&D in a simple Solow-Swan model

In the basic Solow-Swan model, the key elements are a production function:

(1) $Y = A^{\eta} K^{\beta} L^{\alpha}$

where A is an index of technology, K is the capital stock and L is the supply of labor, with both K and L are taken to be fully employed by virtue of the competitive markets assumption, and an accumulation equation:

(2) $\dot{K} = sY - \delta K,$

where $\dot{K}$ is the net investment or the change in the net capital stock, equal to gross investment less depreciation, and δ is a constant depreciation rate. Substituting (1) into (2) gives

(3) $= sA^{\eta} K^{\beta} L^{\alpha} - \delta K.$

From (3) it is possible to determine the conditions for steady state growth in the capital stock.

Re-arranging, taking logarithms, differentiating with respect to time and imposing the condition that for steady state growth:

$$d/dt(\ln \dot{K}/K) = 0$$

gives:

(4) $\dot{K}/K = \frac{\eta}{1-\beta} \dot{A}/A + \frac{\alpha}{1-\beta} \dot{L}/L$

where $\dot{K}/K = \dot{C}/C = \dot{Y}/Y$, is the single constant steady state rate of growth of capital stock, consumption and output, respectively.

The main features of the Solow-Swan model are apparent from equation (4). Firstly, if technology and labor supply are fixed, the steady state growth rate is zero. That is, there is no endogenous growth in the model, growth being driven in the steady state by change in the exogenous variables. Secondly, if one of technology and population show positive growth then the steady state growth rate of the economy is proportional to the growth rate in that variable; if both rates are positive the economy's growth rate is a weighted average of the two. Thirdly, the steady state growth rate does not depend on either the level of savings or of investment in the economy. An economy that continuously saves and invests 20% of national income will have a higher level of output than one investing 5%, but it will not have a higher steady state growth rate. Thus the broad economic message of the Solow-Swan model is that steady growth is possible in a purely competitive world, provided that there is growth in either population or technology, or both.

Contributions to growth and total factor productivity

Solow (1957) further developed this model in a way that provided the foundations for subsequent 'growth accounting'. Starting with total differentiation of the production function (1), and substituting for the partial derivatives of Y from (1) with respect to each of its arguments, yields:

(5) $\dot{Y}/Y = \eta\dot{A}/A + \beta\,\dot{K}/K + \alpha\dot{L}/L.$

Equation (5) can then be used in two main ways in the empirical study of growth.

Given that in the competitive model capital and labor are paid their marginal products and assuming constant returns to scale, β and α can be estimated from the relative shares of capital and labor. A variant of (5) with those weights can then be used to estimate the relative contribution of capital, labor, technology and other factors to growth. Solow made pioneering estimates in 1957, the results of which he later described as "startling" (Solow 1987), and these have been much refined and amplified by Denison (1985) and others. Solow found that 7/8th of the growth in real output per worker in the US economy between 1909 and 1949 was due to "technical change in the broadest sense" and only 1/8th to capital formation. Denison's 1985 estimates covered the US economy for the period 1929 to 1982. Of the growth in real business output of 3.1% per annum over that period, he found that the increase in labor input with constant educational qualifications accounted for about 25% and capital input for 12%. Most of the remainder is accounted for by technological progress and by the increased human capital of the workforce. What was "startling" about these results was the relatively minor contribution to output growth arising from the increase in the traditional factors of production, capital and labor.

The other related use of equation (5) is to estimate the "Solow residual", or total factor productivity. This is defined as the difference between output growth and the weighted sum of the growth rates of factor inputs (K and L), using constant return to scale weights. That is, total factor productivity growth (TFP) is given by:

(6) $\text{TFP} = \dot{Y}/Y - \beta\dot{K}/K - \alpha\dot{L}/L,$

where $\beta = 1 - \alpha$, and β and α are derived from the shares of capital and labor in total income.

Total factor productivity is thus the growth in output not accounted for, on these assumptions, by the growth in capital and labor inputs. This method is now used very widely around the world in measuring productivity. This recent use has confirmed the broad Solow-Denison findings, in that for most modern economies total factor productivity growth is significantly more important than expansion of inputs in explaining total output growth. However, it must be remembered that the method rests on the assumptions embedded in the Solow model and that, as a consequence, the finding that the larger proportion of growth is to be explained by an exogenous "technical change in the broadest sense" constitutes something of an admission of defeat for economic analysis.

Estimating the rate of return to R&D

While there are recognized limitations to the traditional growth model approach, this basic framework has been widely used in estimating the rate of return to R&D. The standard approach to estimating returns to R&D is to divide the technology variable A in (1) into two components, a stock of R&D knowledge variable R and a variable Z that represents a matrix of other factors affecting productivity growth. The production function then becomes:

(7) $Y = K^{\alpha} L^{\beta} R^{\gamma} Z^{\eta}$,

and the counterpart of equation (5) becomes:

(8) $\dot{Y}/Y = \alpha\, \dot{K}/K + \beta\, \dot{L}/L + \gamma\, \dot{R}/R + \eta\, \dot{Z}/Z$.

That is, the rate of growth of the R&D knowledge stock (*i.e.* accumulated R&D expenditure or R&D capital) contributes to output growth as a factor of production, with elasticity γ. The rate of return to knowledge ($\partial y/\partial R$) is that continuing average per cent increment in output resulting from a one per cent increase in the knowledge stock. This can be readily derived from the elasticity γ by

(9) $\partial y/\partial R = \gamma . (Y/R)$.

The normal approach to creating a measure of the stock of R&D knowledge, for a given industry or for the economy as a whole, is to use the perpetual inventory method to create the knowledge stock from the flows of R&D, using the relationship:

(10) $R_t = (1 - \delta) R_{t-1} + R\&D_{t-1}$,

where δ is the rate of obsolescence of the knowledge stock. This method also requires some starting estimates (R_0) of the knowledge stock, and estimates can be sensitive to that assumption.

Then the capital stock at time *t* is given by:

(11) $R_t = (1 - \delta)^t R_0 + \sum_{i=0}^{t-1} (1 - \delta)^I R\&D_{t-1}$

Given a series for R and for the variables Z, it is then possible to estimate γ by either of the two methods noted above: estimate equation (8) with the parameters α .. η unconstrained, or obtain estimates of the parameters α and β (constrained to be equal to one) from the factor shares of capital and labor, calculate TFP by a variant of (7) and regress R and Z on TFP to obtain γ.

Incorporating the efficiency of research and accessibility of knowledge

This standard approach makes some key simplifying assumptions. Here we note three in particular. It is assumed that:

- All R&D generates knowledge that is useful in economic or social terms (*efficiency of R&D*);
- All knowledge is equally accessible to all entities that could make productive use of it (*accessibility of knowledge*); and
- All types of knowledge are equally substitutable across firms and uses (*substitutability*).

A good deal of work has been done to address the fact that the substitutability assumption is not realistic, as particular types of knowledge are often specialized to particular industries and applications. Much less has been done on the other two assumptions, which are our focus.

We define an '*accessibility*' parameter ε as the proportion of the R&D knowledge stock that is accessible to those who could use it productively, and an '*efficiency*' of R&D parameter ϕ as the proportion of R&D spending that generates useful knowledge. Then starting with a given stock of useful knowledge R^*_0 at the start of period zero, useful knowledge at the start of period 1 will be given by:

(12) $$R^*_1 = (1 - \delta) R^*_0 + \phi \, R\&D_0,$$

where the contribution of R&D in period zero to the knowledge stock is reduced by the parameter ϕ to allow for unproductive R&D. This means that the stock of useful knowledge at period *t* is given by:

(13) $$R^*_t = (1 - \delta)^t R^*_0 + \phi \sum_{i=0}^{t-1} (1 - \delta)^i \, R\&D_{t-1}$$

If the period over which knowledge is accumulated is long, so that $(1 - \delta)^t R^*_0$ is small relative to R^*_t, then R^*_t can be approximated by ϕR. However, only a proportion of useful knowledge may be accessible, so that accessible useful knowledge at period *t* is εR^*_t, and hence approximately $\phi \varepsilon R_t$, where R_t is the stock of knowledge as calculated under the standard methods.

Using this approximation and noting that it is accessible useful knowledge that is the correct factor in the production function, (6) becomes:

(14) $$Y = K^\alpha L^\beta (\phi \varepsilon R)^\gamma Z^\eta$$

If ϕ and ε are independent functions of time, then the results of estimating a linearized version of (14) that excludes them will be misleading. However, if we assume that these parameters reflect institutional structures for research and research commercialization in a given country, and can hence be taken as fixed (and as less than or equal to one), then the standard results stand, but need to be reinterpreted. Again using R as the stock of knowledge calculated by the standard method (which assumes $\phi = \varepsilon = 1$) and R^* as the corresponding accessible stock of useful knowledge, then $R = R^*/\phi\varepsilon$, and the rate of return to useful and accessible knowledge becomes:

(15) $$\partial y/\partial R^* = \gamma . (Y/R^*) = \gamma/\phi\varepsilon . (Y/R) = \gamma . (Y/R).1/\phi\varepsilon.$$

Thus, if ϕ and/or ε are less than one, the rate of return to R^* is greater than that to R by the factor $1/\phi\varepsilon$. This does not imply that the measured rate of return to R is biased, because $R^* = \phi\varepsilon R$.

Assume now that there is a one-off increase in the value of ϕ and ε, from the constant values of ϕ_0 and ε_0 to new values of $(1 + \delta_\phi)\phi_0$ and $(1 + \delta_\varepsilon)\varepsilon_0$, respectively. Then the rate of return to R^*, that is:

(16) $\partial y/\partial R^* = \gamma. (Y/R). (1/\phi_0\varepsilon_0)$

is fixed, but the return to R will increase:

(17) $\partial y/\partial R = \gamma. (Y/R) = \phi_1\varepsilon_1 \, \partial y/\partial R^* = \gamma. (Y/R). (\phi_1\varepsilon_1 /\phi_0\varepsilon_0)$

$$= \gamma. (Y/R). (1 + \delta_\phi).(1 + \delta_\varepsilon)\varepsilon_0.$$

It follows from (17) that, because the increase in efficiency and accessibility leads to a higher value of R^* for a given level of R, the rate of return to R will increase by the compound rate of increase of the percentage changes in ϕ and ε.

The basic result of the foregoing is that, if *accessibility* and *efficiency* are constant over the estimation period, but then show a one-off increase, then, to a close approximation, the return to R&D will increase by the same percentage increase as that in the *accessibility* and *efficiency* parameters.

Some methodological notes on the model

While this model specification follows an established literature on the estimation of returns to R&D, there are a number of conceptual difficulties that need to be considered in applying this methodology to estimating the returns to knowledge generated by scholarly publications (*i.e.* journal articles). The first is that the measure of R&D used in the model is expenditure on R&D. This includes many activities that are broader than the creation of the stock of knowledge arising from the writing and publication of scholarly journal articles, which is the focus of this study.

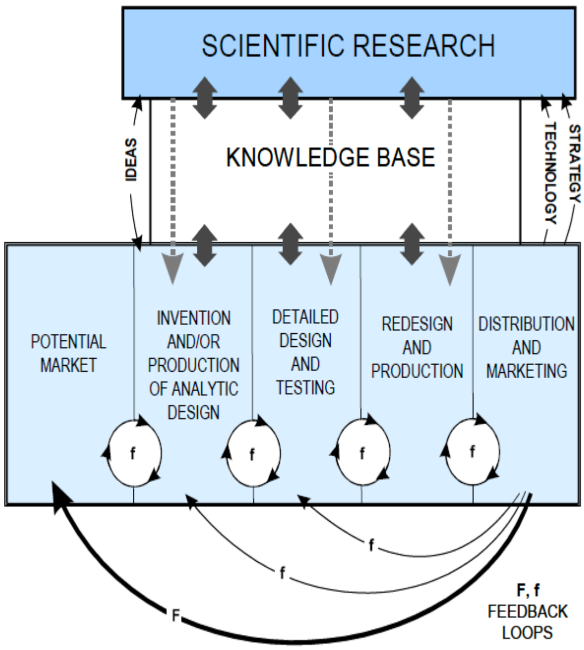
Martin and Tang (2007) explored seven mechanism or channels through which the benefits of publicly funded research may flow through to the economy or to society more generally, namely: (*i*) an increase in the stock of useful knowledge; (*ii*) the supply of skilled graduates and researchers; (*iii*) the creation of new scientific instrumentation and methodologies; (*iv*) the development of networks and stimulation of social interaction; (*v*) the enhancement of problem solving capacity; (*vi*) the creation of new firms; and (*vii*) the provision of social knowledge.[4] Hence, it is not sufficient to simply treat R&D expenditure as adding to the total stock of

[4] Although one could argue that enhanced access is important in all of these (arguably, with the exception of the third). More open access would effectively increase the stock of useful knowledge that is accessible to would-be users; contribute through impacts on education to enhancing the supply and skills of researchers; enable the development of networks on the basis of a shared, common and complete set of information; enhance problem solving capacity by providing necessary supporting information; enable the provision of a range of social knowledge (*e.g.* in health care); and provide opportunities for the emergence of new firms and new industries (*e.g.* text and data mining, metadata and discovery tools).

codified knowledge. Other channels, such as the enhancement of problem solving capacity, reflect increases in tacit knowledge. The training of skilled graduates in basic research is an important part of the R&D function as is their use of tacit knowledge to find and interpret specialized knowledge to solve problems as part of the innovation process.

The second and related issue is the complexity of the innovation process itself. The production function form of the returns to R&D equation proposed herein suggests a simple linear (science push) model of innovation, in which R&D is simply another factor of production. However, it is widely acknowledged that this fails to capture the complex feedback loops of the process, as suggested by the Kline and Rosenberg (1986) chain link model, which at least conceptually captures this complexity (Figure AI.1). It suggests that, in addition to the creation of new ideas and designs from research and their conversion into commercially available technologies, successful innovation depends on feedback from a myriad of actors in the innovation system, including customers, marketing departments, suppliers, etc.

Figure AI.1: Chain Link model of commercial innovation



Source: DEST (2003) *Mapping Australian Science and Innovation*, Canberra, p114. Based on Kline and Rosenberg (1986).

These two factors, the multiple mechanisms through which research impacts on innovation and the complexity of the innovation process itself, make it difficult to ascribe the results of research published in journal articles to particular innovations. This 'attribution problem' has resulted in some estimates of returns to R&D being upwardly biased because of the failure to properly match streams of research costs to streams of outputs (Alston and Pardy 2001). Various approaches have been adopted to deal with this problem. One approach is to introduce control variables for non-research factors into the equations used to estimate returns to R&D (Alston and Pardy 2001). Another is to selectively identify the influence of the stock of knowledge by substituting measures of the stock of publications for broader measures of R&D, such as expenditure, in the returns to R&D equation (Adams 1990, Verspagen 2004).

This suggests that where broader approaches to measuring the returns to R&D are used, such as in this study, some care is required to properly attribute the general returns to R&D to the development of the stock of knowledge represented by scholarly publications. For a single country, such as the US, this requires not only consideration of the extent to which scholarly publications relate to the returns to R&D from federally funded research in the US, but also how spillovers occur between countries (Jaffe 1989).

In this study, our approach is to: (*i*) estimate the proportion of total R&D activity devoted to the production and use of journal articles in terms of researcher time spent reading and writing journal articles; and (*ii*) estimate the extent of international spillovers from the localization of returns to R&D reported by economic studies, national versus international article and patent citation patterns, and national versus international downloads from existing open access archives (See Annex II for details).

Annex II Operationalizing the model

To operationalize the model we need to establish values for the parameters. This section outlines the sources and rationale for the choice of base case values. It also suggests where further data collection is required.

Returns to R&D: Sources and rationale for the range to be modeled

There have been many studies exploring the economic impacts of R&D at the firm, industry and national levels. A characteristic finding is that the returns to R&D are high – often in the region of 20% to 60%, and higher in some cases (Bernstein and Nadiri 1991; Griliches 1995; Industry Commission 1995; Salter and Martin 2001; Scott *et al.* 2002; Dowrick 2003; Shanks and Zheng 2006; Martin and Tang 2007; Sveikauskas 2007; Hall *et al*. 2009).[5] While there is considerable variation in the rates of return reported, those presented in Table AII.1 are indicative. Coe and Helpman (1993), Jones and Williams (1998) and others have shown that similar rates of return arise from endogenous growth models, and champions of the evolutionary approach suggest that, limited to seeing new knowledge as the output of research, simple growth models do not include other forms of economic benefit (*e.g.* skills development, development of instrumentation, development of networks, etc.) (Salter and Martin 2001; Scott *et al.* 2002; Martin and Tang 2007).[6] Hence, if anything, the approach used herein may understate the returns.

Table AII.1: Estimates of private and social rates of return to private R&D

Study	*Private rate of return (%)*	*Social rate of return (%)*
Minnasian (1962)	25	..
Nadiri (1993)	20-30	50
Mansfield (1977)	25	56
Terleckyj (1974)	27	48-78
Sveikauskas (1981)	10-23	50
Goto & Suzuki (1989)	26	80
Mohnen & Lepine (1988)	56	28
Bernstein & Nadiri (1988)	9-27	10-160
Scherer (1982, 1984)	29-43	64-147
Bernstein & Nadiri (1991)	14-28	20-110

Source: Salter, A.J. and Martin, B.R. (2001) 'The economic benefits of publicly funded basic research: a critical review,' *Research Policy* 30(3), p514.

5 Useful reviews of this literature include Griliches 1995; Salter and Martin 2001; Scott *et al.* 2002; Shanks and Zheng 2006; Martin and Tang 2007; Sveikauskas 2007; and Hall *et al*. 2009; 2010.

6 As Sveikauskas (2007, p6) noted: "measured benefits are limited to those that have a market evaluation. Some benefits, such as clean air or some types of medical advances, are not evaluated through market prices, and are typically not included in economic statistics."

There have been a number of studies showing the industry impacts of publicly funded science in general, and of scientific and scholarly publications in particular. Mansfield (1991; 1998) attempted to measure the returns to R&D for those innovations that are directly related to academic research. From a survey of R&D executives in US firms, he found that around 10% of new products and processes would not have occurred (without substantial delay) in the absence of recent academic research. In a follow-up study, Mansfield (1998) found that academic research was increasingly important for industrial innovation. Similarly, the PACE Survey of large European firms showed that firms rely heavily on scientific publications as a source of information about publicly funded research (Arundel *et al.* 1995). More recently, Ware (2009, p27) found that small and medium sized firms rated original research articles and review papers in journals as their most important sources of information (as did university and college based researchers). For large firms, technical information and standards were more important, but journal articles were still among the most important sources. Sveikauskas (2007, p26) concluded his review of the literature saying: "[t]hese articles show that, beyond the firm to firm transfers that comprise the core of the R&D literature, substantial technology is also transferred from government or universities to private firms."

In establishing what is a plausible range of rates of return to use, we take a lead from the literature. Arundel and Geuna (2003, p3) surveyed the literature, and reported that estimates of the rate of return to publicly funded research ranged between 20% and 60%. Martin and Tang (2007, pp6-7) noted that:

> *…there have been numerous attempts to measure the economic impact of publicly funded research and development (R&D), all of which show a large positive contribution to economic growth. For instance, the studies cited in OTA (1986) and Griliches (1995) spanning over 30 years of work find a rate of return to public R&D of between 20 and 50%…*
>
> *Mansfield (1991)… estimated the rate of* [private] *return for academic research to be 28%.*
>
> *Toole (1999) has shown… that firms appropriate a* [private] *return on public science investment of between 12% and 41%.*

Exploring the impact of research articles in the Netherlands, Verspagen (2004, pp10-11) concluded that:

> *In the optimistic scenario the rate of return is 81% and in the cautious scenario it is 2%. When applying the weights for domestic and foreign sources… we arrive at a point estimate of a 59% rate of return on academic research by universities in the Netherlands.*

In one of the most recent reviews of the literature, Hall *et al.* (2009; 2010) summarized the results of almost 100 studies, showing that the returns reported in the US studies ranged from 18% to 76%. Hall *et al.* (2009, p23) concluded that:

> *On the whole, although the studies are not fully comparable, it may be concluded that R&D rates of return in developed economies during the past half century have been*

strongly positive and may be as high as 75% or so, although they are more likely to be in the 20% to 30% range.[7]

In light of debates over difficulties attributing spillover and downstream returns to R&D or complementary investments, prices and measurement error (Sveikauskas 2007), we adopt the lower bound of 20% as a plausible conservative average rate of return to publicly funded research (including both private and spillover or social returns). In view of the large share of federally funded R&D in the US going to life sciences, wherein returns are often reported to be higher, this is likely to be erring on the conservative side.

Estimate for the model (lower bound of reported average returns): 20%.

How local are these returns?

There are various ways to explore the likely localization of returns within a country. Here we mention three.

1. **Economic studies on the localization of returns**: A number of studies have looked at the issue of the relative impact of local research on local returns and/or the international spillover of R&D. Jaffe (1989) suggested that domestic knowledge is twice as important as foreign knowledge (*i.e.* 66% was local). Coe and Helpman (1993; 1995) adopted a trade weighting approach, and concluded that approximately a quarter of the benefits from R&D in G-7 countries accrued to their trading partners, and 75% locally (Hall *et al*. 2009). Verspagen (2004, p10), citing Arundel and Guena (2004), suggested weights for domestic versus foreign sources of 73% for domestic and 27% for foreign sources.

2. **Article and patent citation patterns**: Article citations reflect just one specialized area of use, namely use in further published research, and do not reflect wider economic and social application. Nevertheless, citation patterns could be seen as an indicator of the local use of local research articles. The National Science Board (2010, pO-12) reported that 60% of the articles cited in US-authored articles are themselves US-authored articles. Looking at patent citation patterns, Sveikauskas (2007, p42) noted that between 30% and 53% of US patent citations were to non-US sources.

3. **Repository statistics**: Repository statistics are another possible source of information on the localization of use of scholarly work, especially that which is open access. Unfortunately they cannot be applied in the US and international data present a very mixed picture, with national downloads (*i.e.* those to the archive's country-code top level domain – ccTLD) varying from highs of 95% and more to lows of 20% and less. In the small international sample explored, however, the mean across repositories (N=12) was around 45%. Such download percentages will tend to understate the share of local use as there is likely to be a further share of local global top level domains (gTLDs) that remain unidentified in the data as well as a substantial number of unresolved domains. Indicatively, perhaps, one could add 45% of the gTLD and unresolved traffic to the ccTLD traffic. As such, the evidence of local use from repository download

[7] Differences in rates of return to R&D across the developed countries are not large (Hall *et al.* 2009; Cutler *et al*. 2008).

statistics is broadly in accord with the reported shares of local in total returns to R&D from the economic studies noted above.

Estimate for the model (average of reported national returns, citations and downloads): 66%.

Box AII.1: Diffusion of knowledge and returns to R&D

To illustrate the importance of the diffusion of knowledge from firms undertaking research to the broader community, consider technical change in pharmaceuticals. Statins are a new class of anti-cholesterol drugs which have contributed greatly to the decline of heart disease. A major pharmaceutical firm introduced the first commercial statin product in 1987, and conducted pioneering research demonstrating that statins were safe, lowered cholesterol, and successfully reduced the death rate from heart disease. Since 1987, several firms have introduced new and improved statins. A different firm now produces a new and greatly improved statin, which lowers cholesterol more effectively, and has therefore become the market leader.

Although the second firm now dominates the market for statins, it is not the case that the second firm's private investment in R&D is now the only relevant R&D. From the point of view of private returns, much of the early research which the initial firm carried out is indeed no longer profitable. However, in a broader sense all the initial research which demonstrated that statins were safe, highly effective, and reduced the incidence of heart disease still provides the core knowledge of the present day industry. The first firm's initial investment in R&D is still relevant to the industry and still provides important social returns, even though most of the private returns now go to the second firm.

To take another example, two leading firms have competed in producing microprocessors for many years. When the technology leader introduces a new chip, the second firm soon matches, and prices fall rapidly. As a result, microprocessor prices have declined sharply. Most of the benefits of innovation have been captured by consumers through lower prices. The profit of the innovators, obtained through returns to the R&D they conduct, is only a small part of the picture.

These examples illustrate how the knowledge and benefits obtained from R&D typically leak out from the original performers of R&D to competitors, to other firms, to consumers, and, eventually, to other countries. Many forms of knowledge are useful to other firms (and so have a social return) even when they no longer pay off to the firm initiating the research (no longer have a private return). Similarly, consumers obtain better or cheaper products (benefit from social returns) even if the private return to firms turns out to be low.

Source: Sveikauskas, L. (2007) *R&D and Productivity Growth: A Review of the Literature,* US Bureau of Labor Statistics Working Paper 408, BLS, Washington, DC. pp4-6.

Accessibility: Sources and rationale for the range to be modeled

Accessibility is defined as the proportion of the stock of knowledge generated by R&D that is accessible to those who could use it productively.

The key question is what impact might the open archiving of articles from federally funded research, as proposed under the FRPAA, have on accessibility? This can be unpacked to the following questions:

- What proportion of the stock of R&D knowledge produced by federally funded research is in journal articles?
- What proportion of the stock of R&D knowledge is likely or potentially available to archiving and alternative access?
- What measures are there of the potential impacts of the proposed FRPAA archiving mandate on accessibility?

We deal with each of these in turn.

What proportion of the stock of R&D knowledge produced by federally funded research is in journal articles?

Under the assumptions of the standard approach the stock of R&D knowledge is an output of the stream of expenditure on R&D, and whatever researchers do with the R&D funding they employ can contribute to the stock of R&D knowledge. Hence, a possible proxy for the proportion of the stock of R&D knowledge that is in journal articles is the proportion of researchers' time spent reading and writing articles (and, perhaps, peer reviewing and acting in journal editorial capacities).

Tenopir and King (2000), and the subsequent tracking studies, report the average time spent by researchers in industry and in universities on a number of tasks, including the time spent reading and writing journal articles. These studies suggest that researchers spend an average of around 90 to 100 hours writing journal articles. Both reading and writing habits vary between industry and university based researchers, but reading times for journal articles range from around 75 to 150 hours per year, suggesting that active researchers spend around 20% to 25% of their time reading and writing journal articles. Hence, on the basis of time spent, we could say that journal articles constitute some 20% of the stock of R&D knowledge.

What proportion of the stock of R&D knowledge is potentially available to archiving and alternative access?

Noting again that the stock of R&D knowledge is the output of the stream of expenditure on R&D, part of the answer to this question lies in the sectoral shares of R&D expenditure (*i.e.* the level of federal research funding).

In addition, some share of the article output from federally funded research will already be available open access, so we adjust according to the estimated share of articles produced in 2008 that are available open access (20%) (Björk *et al*. 2010) and for the level of compliance with the existing NIH mandate (56%) (NIH 2008, p26). These adjustments suggest that around 37% of federally funded article output is already openly accessible.

Hence, in order to focus on the *incremental impacts* of the proposed open archiving mandate, we limit analysis to FRPAA agency-related federally funded research and adjust for the share that is already openly accessible. To estimate the *overall impacts* we simply include the share of articles that are already openly accessible.

What measures are there of the potential impacts of the FRPAA archiving mandate on accessibility?

The proposed FRPAA archiving mandate relates to published research articles, so the crucial issue is the additional access that might be achieved through archiving. In the US and other developed countries, many researchers already have access to published articles through the journals concerned (*e.g.* via institutional library subscriptions). Outside the major institutions, however, such access can be more limited (*e.g.* for small firms, professionals, practitioners, educators and the general public). Such access can also be more limited for potential users in developing countries. While it is very difficult to estimate the potential increase in access, there are a number of possible proxy indicators.

Reports of access limitations, access gaps and difficulties

There have been a number of studies exploring access issues for researchers in various fields of research, institutional and sectoral settings. In a brief review of such studies, focusing very largely on research authors access to research and developed countries, Davis (2009) found that most indicated reasonably good and improving levels access for research authors employed in developed countries, although a significant number reported access difficulties and/or gaps, as did those from developing countries. Among the studies noted, Rowlands and Olivieri (2006) found that 67% respondents in immunology and microbiology reported having good or excellent access (33% did not); and Ware (2007) found that among an international sample, 69% of respondents reported having good or excellent access (31% did not), and outside the US and Canada 53% of respondents reported having good or excellent access (47% did not).

Ware (2009) looked at *Access by UK small and medium-sized enterprises to professional and academic literature*, although his study also included researchers and users in universities and colleges, hospitals and public health facilities, public research institutions and government departments, and other practitioners, professionals and individuals (Table AII.2). Ware found that 73% of small to medium sized firm (SME) respondents, 53% of large firm respondents and 27% of university or college respondents reported having difficulties accessing articles. Just 2% of SMEs, 7% of large firms and 17% of higher education-based researchers reported having access to all the articles they needed for their work. Amongst those experiencing access difficulties, those difficulties affected 6% to 10% of articles read. Of the entire sample, however, Ware (2009, p13) concluded that the percentage of articles with access difficulties ranged between 10% and 20%, of which between 21% and 55% related to the toll access barrier.[8] It should be noted that in Ware's survey, 71% of SMEs reported using open access journals and 42% reported using institutional repositories, so the reported access difficulties included current levels of open access availability (accessibility).

8 It should be noted that Ware's sample is reported to have been based on author, subscriber and pay-per-view transaction lists supplied by publishers, even though one might expect access gaps to be less prevalent amongst such groups than more generally. As such, the study may understate access gaps.

Table AII.2: Access to research articles, June 2009 (per cent)

Access to research articles	*SMEs* N=186	*Large Firms* N=111	*University College* N=470	*CIBER 2004* N=3,787
Excellent (I have access to all the articles I need)	2%	7%	17%	15%
Good (I have access to most of the articles I need)	26%	39%	55%	46%
Varied (I sometimes have difficulty getting the articles I need)	56%	37%	22%	29%
Poor (I frequently have difficulty getting articles)	14%	13%	4%	8%
Very Poor (I always have great difficulty getting articles)	3%	3%	1%	2%
Experiencing access difficulties	73%	53%	27%	39%
Have access to all I need	2%	7%	17%	15%
Number reporting recent access difficulties	55%	34%	24%	..

Source: Ware, M. (2009) *Access by UK small and medium-sized enterprises to professional and academic literature,* Publishing Research Consortium, Bristol.

RIN (2009, pp8-9) reported that more than 80% of survey respondents said that the difficulties they encountered in gaining access to content had an impact on their research, and nearly a fifth (16%) said that the impact was 'significant'. The most common impacts reported were delays in research, and inconvenient and disruptive interruptions to workflow. Lack of access is also a hindrance to collaborative working, and can lead to delays in the submission of papers and of bids for funding. Peer reviewers are also hindered when they cannot access sources cited by an author, and scientists worry that lack of access to the latest findings and methodologies may lead them to undertake redundant work.

Estimates of the Open Access citation and download advantages

There are many studies of, and active discussion about, a possible open access (OA) citation advantage, with general agreement that there does seem to be an observable advantage and argument focusing mainly on why (EPS *et al*. 2006).[9] The observed advantages vary considerably. In his brief review, Davis (2009) noted that: Davis *et al*. (2008) reported that freely-accessible articles received no more citations than subscription access articles, but they did receive significantly more downloads (*i.e.* an 89% increase in full text downloads, suggesting wider access and use); and Evans and Reimer (2009) found that freely accessible articles received about 8% more citations on average, and twice that for the poorer countries. In one of the more widely cited studies, Hajjem *et al*. (2005) concluding that:

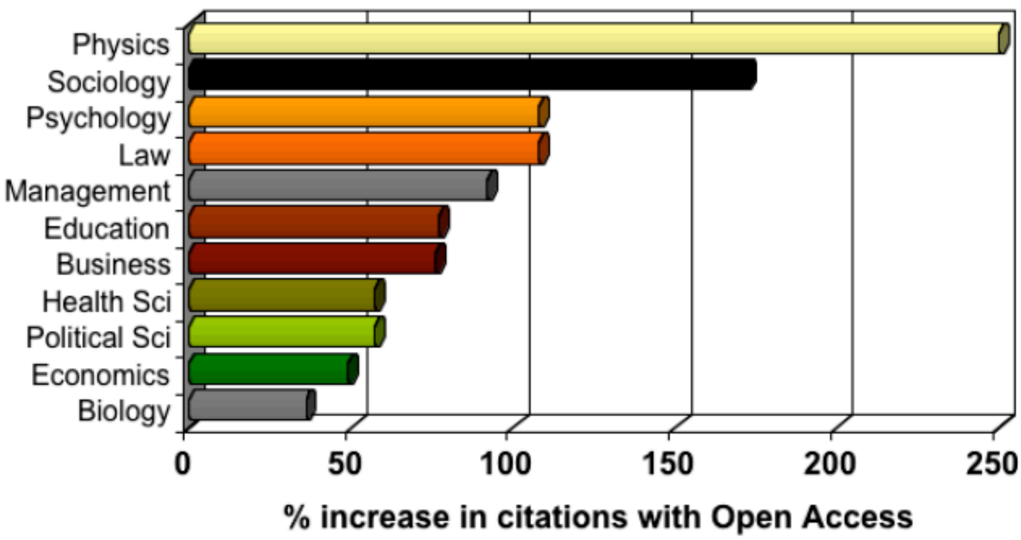
> *In 2001, Lawrence found that articles in computer science that were openly accessible (OA) on the Web were cited substantially more than those that were not. We have since replicated this effect in physics. To further test its cross-disciplinary generality, we used 1,307,038 articles published across 12 years (1992-2003) in 10 disciplines (Biology, Psychology, Sociology, Health, Political Science, Economics, Education, Law, Business, Management). The overall percentage of OA (relative to total OA + NOA)*

9 See http://opcit.eprints.org/oacitation-biblio.html for a bibliography.

articles varies from 5%-16% (depending on discipline, year and country) and is slowly climbing annually. Comparing OA and NOA articles in the same journal/year, OA articles have consistently more citations, the advantage varying from 25%-250% by discipline and year.

More recently, and most directly relevant to the proposed FRPAA archiving mandate, Gargouri *et al*. (2010) found that articles whose authors have supplemented subscription-based access to the publisher's version by self-archiving their own final draft to make it accessible free for all on the web (archived) average twice as many citations as articles in the same journal and year that have not been made Open Access (archived).

Figure AII.1: Average citation ratios for articles in the same journal and year that were and were not made OA by author self-archiving (1992-2003)



Source: Harnad, S. *et al.* (2004) 'The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update,' *Serials Review 34*, pp36-40.

In the most recent review of the literature, Swan (2010b) summarized the findings of 36 studies noting that 27 reported finding an open access citation advantage and 4 found no advantage, and that the citation advantage found ranged from 31% to 400%. Four of the studies focused on one of the most established archiving forums, finding that, on average, articles on arXiv received twice as many citations.

Few studies have looked at the sources of citations for subscription and open access materials (*i.e.* identifying the users of the content and the nature of use). However, looking at the evidence from two journals Zhang (2006) found that: on average open access articles received twice as many citations as non-open access articles; the largest increase in open access article citations came from non-scholarly documents, such as academic essays, encyclopedia, online discussions and research reports, and from course and teaching materials; and a major source of the observed citation boost came from developing countries. Zhang (2006, p155) concluded:

> *The Web citation advantage of OA journal JCMC was demonstrated. Published online, OA articles are freely accessible to any user having Internet access so that they may potentially have a much larger size of readership than traditional access journal articles, and consequently receive far more citations. The classification of Web citation sources shows that traditional access journal articles have a significantly smaller proportion of citations from non-formally published academic materials than OA articles. This indicates that the OA articles' impact advantage over traditional access counterparts in informal academic communication is even more distinct than in formal communication, which is represented by formal publication and school education. The classification of Web citations by countries shows that JCMC articles receive a higher proportion of Web citations from developing countries and from a wider international scope. A convincing interpretation is that open access could effectively improve the articles' impact in developing countries and contribute to decreasing the academic gap between developing countries and developed countries.*

Of most immediate relevance are the studies relating to archiving, which show that articles that are openly archived receive around twice as many citations. Some adjustment of the citation and download advantages is necessary to take account of what might be the temporary and what a permanent advantage (Harnad 2005) and of what is already openly accessible. Björk *et al*. (2010) found that some 20% of articles published in 2008 were available open access, and NIH (2008, p26) reported a 56% compliance rate for their archiving mandate. Hence, we adjust the citation and download advantages to take account of an estimated average of 37% of articles from FRPAA-related federally funded research that are already openly accessible.

Box AII.2: Model parameter: Percentage change in accessibility

Parameter	***Basis***	***Value***
Percentage change in accessibility (Reported access gaps)	Ware (2009): 10% to 20% of articles read presented access difficulties	Adjusting for share of difficulties due to toll access barriers
Percentage change in accessibility (OA citation advantage)	Hajjem *et al*. (2005): 25% to 250% more citations Gargouri *et al*. (2010); Zhang (2006): average 100%	Adjusting for what is already OA and articles as a share of the research stock of knowledge
Percentage change in accessibility (OA download advantage)	Davis *et al*. (2008): 42% to 89% more pdf and full text downloads (average 66%)	Adjusting for what is already OA and articles as a share of the research stock of knowledge
Combined estimate of the percentage change in accessibility to be modeled		*Taking the lower bound of the ranges above: Estimate for model 4.68%*

Source: Authors' analysis.

For preliminary estimates we take 4.68% as a conservative estimate of the potential incremental increase in accessibility – based on the lower bound reported impacts (Box AII.2).

Efficiency: Sources and rationale for the range to be modeled

Efficiency is defined as the proportion of R&D spending that generates useful knowledge, and can have a number of dimensions relating to wasteful, inefficient and/or poorly directed research expenditure. The key question is what impact might the open archiving of research articles from federally funded research, as proposed under the FRPAA, have on efficiency?

Drawing on a previous analysis of the literature (Houghton and Oppenheim *et al*. 2009) suggested that key dimensions of impact might include:

- Researcher and research system cost savings that might arise from the additional/easier access facilitated through mandatory archiving, with those savings being spent doing more research for the same level of R&D spending, and thereby producing more useful knowledge for a given spend;
- The potential avoidance of duplicative and ill-informed research, and of scientific fraud and plagiarism, and thereby the reduction of wasteful expenditure;
- The potential for reduction in accessibility delays leading to a speeding up of the research and discovery process (subject to the embargo limitations imposed), thereby producing more useful knowledge for a given cost;
- The potential for better and/or more informed review and evaluation of funding proposals and research outputs leading to better allocation of grants and other funding, thereby providing support to more productive and useful research and thereby producing more useful knowledge for a given cost;
- The potential for greater support for interdisciplinary research and collaborative research (*e.g.* inter-sectoral collaborations) leading to greater research focus on problems in areas of greater and/or more immediate impact (*e.g.* clean energy, climate change, etc.), thereby producing more useful knowledge; and
- The potential to enable greater research participation from developing countries, thereby unlocking new potential to generate more useful knowledge.

With many possible impacts on efficiency (RIN 2009), but few immediately available metrics, the best we can do is to explore plausible scenarios as a way to get a sense of the potential scope and scale of possible impacts (for illustrative purposes only).

Scenario 1: Less risk of duplicative research being done and of pursuing blind alleys through greater access and more complete dissemination

If just 1% of total federally funded research time were spent performing duplicative research and pursuing blind alleys that could have been avoided if researchers had had more complete access to the findings of others, then the annual 'saving' would have been around $1.4 billion – equivalent to around 12 million researcher hours. With returns to publicly funded R&D of 20%, the implied lost annual returns (*i.e.* from the same amount of research expenditure that was not duplicative) would have been around $280 million annually.

Scenario 2: Collaborative research and new research opportunities made possible by greater access to research publications brings higher returns to R&D

It is widely held that there are advantages to collaborative research and greater use of the findings from collaborative work (Katz and Hicks 1997; Katz and Martin 1997; Walsh and Maloney 2001). Enhanced access through centralized archives can offer greater support for collaboration, on the basis of a share common base of materials. Enhanced access through centralized archives can also increase opportunities for new research approaches (*e.g.* text mining). If greater and easier collaboration and new research opportunities increased the returns to federally funded R&D by just 1%, then it too would be worth around $1.4 billion per annum.

Scenario 3: Enhanced accessibility saves research time, allowing more research to be done for the same R&D expenditure

Exploring the potential research activity time and cost savings relating to: (*i*) reduced search and discovery time through enhanced discoverability and greater access, and less use of proprietary silo access systems; (*ii*) less time spent on seeking and obtaining permissions to use (copyright and licensing); (*iii*) less time spent on checking during peer review through greater access, in turn making for better quality review; and (*iv*) less time spent on writing and preparation through greater access making reference checking etc. easier, Houghton and Oppenheim *et al.* (2009) reported potential annual research activity savings from open access of GBP 73 million in UK higher education – equivalent to around 1.2% of higher education research expenditure. Scaling these scenarios to the scope of the FRPAA article archiving mandate and translating to US research activity and expenditure levels, would suggest potential US federally funded research activity savings of $43 million per annum – equivalent to around 380,000 research hours per annum. Of course, these sorts of savings might be available to all research, not just that funded federally in the US.

Box AII.3: Model parameter: Percentage change in efficiency

Parameter	***Basis***	***Value***
Percentage change in efficiency (wasteful expenditure: duplicative research and blind alleys)	Scenario 1, for illustrative purposes	1%
Percentage change in efficiency (new opportunities: collaborative opportunities and new methods)	Scenario 2, for illustrative purposes	1%
Percentage change in efficiency (Research time savings)	Scenario 3, for illustrative purposes	..
Combined estimate of the percentage change in efficiency to be modeled		*In the absence of a grounded metric* ***0%***

Source: Authors' analysis.

However, given the lack of a grounded metric we have not included any increase in efficiency in our preliminary estimates (Box AII.3).

Other parameters: Sources and rationale for the range to be modeled

There are a number of other parameters required in the modeling of impacts, for which we have adopted conservative values so as not to risk overstating the potential benefits.

Rate of growth of R&D spending

Various subsets of federal R&D spending are examined and there are differences in spending trends between sectors and agencies. However, the National Science Board (2010) reported 5.8% per annum growth in US R&D spending over the last 10 years in current values (3.3% per annum real), and that federal spending on R&D had increase by 3.2% per annum over the last 10 years.

Estimate for the model: 3.2% per annum.

Lag between R&D spending and impacts

Lags between research spending and impacts being felt can be very long in some fields, perhaps 20 to 30 years, and short in others, perhaps 1 to 2 years or less. Mansfield (1991; 1998) reported that for US firms the average lag between the publication of academic research and the timing of subsequent commercial innovation relying on it was around seven years (falling to 6.2 in the later study). One might expect some further speeding up of the research and commercialization process since that time, but we model an average lag of 10 years for the base case to take account of the seven years reported by Mansfield (1991; 1998) and allowing a further three years for the lag between project funding/expenditure and publication.

Estimate for the model: lag 10 years.

Distribution of impacts over time

As well as being lagged, impacts occur over time. Mansfield (1991; 1998) reported that for US firms the lag between the publication of academic research and the timing of subsequent commercial innovation relying on it ranged from a minimum of 4.2 years to a maximum of 9.8 years, falling to 5.2 years to 8.5 years in the later study. However, these are private returns. Sveikauskas (2007, p6) noted that: "as knowledge gradually leaks out, private benefits decline and spillover effects increase. Consequently, private and spillover returns follow different time paths… spillover effects are considerably more long lived than private effects." Hence we distribute the impacts over approximately 10 years.

Estimate for the model: normal distribution over 10 years.

Rate of inflation (cost increase)

Costs change differently in different areas, but overall inflation (Consumer Price Index) gives an approximate guide, and reported CPI over the last 10 years has averaged around 3% per annum.

Estimate for the model: 3% per annum

Discount rate

There is active discussion of the appropriate discount rate to use in cost-benefit calculation, with some suggesting very low rates and others much more conservative rates (Evans and Sezer 2002; Harrison 2007). Again, we adopt the more conservative approach.

Estimate for the model: 10% per annum.

Box AII.4: Model parameter: Rate of return to R&D and other parameters

Parameter	***Basis***	***Value***
Returns to R&D	Conservative consensus from the literature (Arundel and Geuna (2003; Hall *et al.* 2009; etc.)	20% to 60% (estimate 20%)
Local share of returns to R&D	Consensus from the literature (Jaffe 1989; Coe and Helpman 1995; Verspegan 2004, etc.)	66%
Rate of growth in R&D spending	National Science Board, S&E Indicators 2010	3.2% per annum
Lag between R&D spending and impacts	Mansfield (1991, 1998)	3 years to publication plus 7 years to impact, 10 years
Distribution of impacts	Mansfield (1991, 1998), Sveikauskas (2007)	Normal over 10 years
Discount rate (risk premium)	Conservative consensus from literature	10% per annum
Depreciation of Stock of Research Knowledge	Griliches (1995), Hall (2009), Sveikauskas (2007)	Less than 9% and up to 15%, (estimate) 8% pa
Rate of archiving cost increase	Set to average CPI over last 10 years	3% per annum

Source: Authors' analysis.

Rate of depreciation of the underlying knowledge stock

Looking at the most appropriate rate of depreciation to apply, Hall *et al.* (2009, p16) noted that most researchers use the 15% that Griliches had settled on in his early work. However, this may be more suitable for private returns than publicly funded research. Sveikauskas (2007, p6) noted that:

> *"Okubo et al. (2006) calculate R&D asset stocks assuming a 15 percent (or greater) annual depreciation rate. In contrast, the Bureau of Labor Statistics (1989), measuring the longer lasting spillover effects, assumes 10 percent depreciation for applied research and development and zero depreciation for basic research, which implies an overall depreciation rate of less than 9 percent."*

If we apply these BLS rates to the balance of federally funded R&D in 2008, which was approximately 20% basic research, it implies an average depreciation rate of 8%.

Estimate for the model: 8% per annum.

Data: Sources and rationale for the base case values

The third piece of the puzzle is the input data required for the modeling. The main requirements include the implied archiving costs, the volume of federally funded research outputs (*i.e.* journal articles), and the levels of FRPAA-related federal research funding and expenditure trends. For the purposes of preliminary analysis we have used publicly available sources and published estimates (Box AII.5).

Box AII.5: Model parameters: Base case data sources and values

Parameter	***Basis***	***Value***
Federal R&D Spending (USD billions)	NSB 2010 indicators: R&D expenditure by the 11 FRPAA departments in 2008	$61
Annual growth in federal R&D spending (per cent)	NSB 2010 indicators: reported growth over last 10 years	3.2%
Average annual salary of researchers (USD)	NSB 2010 indicators: reported average salaries in 2008	74,070
Number of articles published from federal R&D (2008)	NIH 2008: estimate based on the ratio of NIH expenditure to article output	170,000
Number of articles published from NIH funded research (2008)	NIH (2008, p22)	80,000
Average annual growth in article output (per cent)	NSB 2010 indicators: over last 10 years	1.8%
Per article submission-based costs (USD)	ArXiv (2010)	$7
Per article submission-based costs (USD)	NIH (2008, p22)	$59
Per article life-cycle archiving cost in first year (USD)	LIFE^2 Project: Year 1 life-cycle costs	$34
Per article life-cycle costs per year in subsequent years (USD)	LIFE^2 Project: Subsequent year annual life-cycle costs	$12
Time for author deposit (minutes per article)	Reported average use of the NIHMS submission system (NIH 2008, p14)	10 mins
Annual growth in archiving costs (per cent)	BLS: Average US CPI over last 10 years	3%
Average level of compliance with mandate over 30 years (per cent)	Assumed full compliance for the base case	100%
Embargo period (months)	Assumed six month embargo for the base case	6

Source: Authors' analysis.

Data relating to federal research funding, activities and outputs are taken from the most recent National Science Board *Science and Engineering Indicators 2010* (NSB 2010). We explore three sources for archiving costs:

- The LIFE^2 Project (Ayris *et al*. 2008), which reported life-cycle costs for articles and other items held on institutional archives in the UK, and found costs equivalent to up to $34 per article in the first year, and $12 per article held per annum in subsequent years;
- Reporting costs on a submissions equivalent basis, NIH (2008) estimated that it would cost $4.5 million per annum to host the estimated 80,000 articles from NIH funding

circa 2008 and noted that they had spent a further $250,000 on policy-related staff costs, implying a per article cost of around $59 per submission; and

- Also reporting costs on an approximate submissions equivalent basis, arXiv (2010) noted that their annual budget was $400,000 rising to $500,000 by 2012 and that 64,047 articles had been submitted in 2009, implying a per article cost of around $7 per submission.

For the purposes of producing preliminary estimates, we explore this range of costs – noting that the mid-range NIH reported costing might be the best guide.

Box AII.6: A brief description of the model

Main characteristics: A spreadsheet model to estimate the impacts of possible increases in accessibility and efficiency on returns to R&D over 30 years, with two major data inputs: (*i*) federal R&D expenditure levels and trends, and (*ii*) the costs associated with the proposed archiving.

Assumptions and parameters: All the parameters can be changed in order to explore various scenarios and test sensitivities. As outlined above, they include: (*i*) the rate of return to R&D and share of returns that are national, (*ii*) the rate of depreciation of the underlying stock of research knowledge, (*iii*) the discount rate applied to costs and benefits to estimate net present value, (*iv*) the level and rate of growth of federal R&D expenditure, (*v*) the level and rate of growth of costs associated with the archiving of articles resulting from federally funded research, (*vi*) the average lag between funding, publication or archiving and returns to R&D in years, (*vii*) the effective embargo period between publication and archiving, (*viii*) the level of compliance with the FRPAA mandate, and (*ix*) the share of articles produced by federally funded R&D that are currently available open access.

Transition versus steady state alternative: Because of the lag between research expenditure and the realization of economic and social returns to that research, the impact on returns to R&D is lagged (by 10 years in the base case scenario) and the value of those returns discounted accordingly. This reflects that fact that the archiving would be prospective and not retrospective, and the economic value of impacts of enhanced accessibility and efficiency would not be reflected in returns to R&D until those returns are realized. An alternative approach would be to model a hypothetical alternative 'steady state' system in which the benefits of historical increases in accessibility and efficiency enter the model in year one. This would reflect the situation in an alternative system, after the transition had worked through and was no longer affecting returns to R&D. The model used herein to estimate impacts focuses on the transition and explores impacts over a 30 year transitional period.

Note: See Annexes I and II for details.
Source: Authors' analysis.

We have created a simplified model in MS Excel format, in order to enable anyone to examine a range of values for the various parameters, test sensitivities and explore the issues for themselves. It is available at http://www.cfses.com/FRPAA/. We encourage people to experiment with it and we would welcome any feedback.

References

AAAS (2008) *AAAS Report XXXIII: Research and Development FY 2009*, AAAS Publication Number 08-1A. Available www.aaas.org/spp/rd/rd09main.htm (January 2010).

ACMI (2008) *Innovation Measurement: Tracking the State of Innovation in the American Economy*, Report to US Department of Commerce. Available www.innovationmetrics.gov/Innovation%20Measurement%2001-08.pdf (February 2010).

Acs, Z., Audretsch, D. and Feldman, M. (1992) 'Real effects of academic research: comment', *American Economic Review* 82(1), pp363-367.

Adams, J.D. (1990) 'Fundamental stocks of knowledge and productivity growth,' *Journal of Political Economy* 98(4), 673-702.

Allen Consulting Group (2004) *Measuring the impact of publicly funded research*, Department of Education, Science and Training, Canberra.

Alston, J.M. and Pardy, P.G. (2001) 'Attribution and other problems in assessing the returns to agricultural R&D,' *Agricultural Economics* 25, pp141-152.

Antelman, K. (2004) 'Do open-access articles have a greater research impact?', *College & Research Libraries News,* 65(5), pp372-382. Available http://eprints.rclis.org/archive/00002309/ (January 2010).

Arundel, A. and Geuna, A. (2003) *Proximity and the Use of Public Science by Innovate European Firms,* Royal Economic Society Annual Conference 2003 86, Royal Economic Society.

Arundel, A. and Geuna, A. (2004) 'Proximity and the use of public science by innovative European firms,' *Economics of Innovation and New Technology* 3(6), pp559-580.

Arundel, A., van de Paal, G. and Soete, L. (1995) *PACE Report: Innovation strategies of Europe's largest firms: Results of the PACE survey for information sources, public research, protection of innovations and government programmes*, MERIT, University of Limberg, Maastricht.

arXiv.org (2010) *arXiv Business Model White Paper*, Cornell University. Available http://arxiv.org/help/support/whitepaper (February 2010).

Ayris, P. *et al*. (2008) *The LIFE[2] Final Report*, Joint Information System Committee (JISC), London & Bristol. Available http://www.life.ac.uk/2/documentation.shtml (February 2010).

Bernstein, J.I. and Nadiri, M.I. (1991) 'Product demand, cost of production, spillovers and the social rate of return to R&D,' NBER Working paper 3526.

Björk, B-C., Roos, A. and Lauri, M. (2008) 'Global annual volume of peer reviewed scholarly articles and the share available via different Open Access options,' Proceedings of ELPUB 2008, Toronto, June 2008.

Björk B-C., Welling P., Laakso M., Majlender P., Hedlund T., *et al*. (2010) 'Open Access to the Scientific Journal Literature: Situation 2009', *PLoS ONE* 5(6): e11273. doi:10.1371/journal.pone.0011273. Available http://www.plosone.org/article/info:doi/10.1371/journal.pone.0011273 (July 2010)

Bureau of Economic Analysis (2007) Research and Development Satellite Account, BEA, Washington DC. Available http://www.bea.gov/national/newinnovation.htm (January 2010).

Bureau of Labor Statistics (2009) *Contribution (1987-2007) and Stocks (1948-2007) of Research and Development (R&D)*, Bureau of Labor Statistics, Washington DC.

Coe, D.T and Helpman, E. (1993) *International R&D Spillovers,* NBER Working Paper 4444, National Bureau of Economic Research, Cambridge MA.

Coe, D. T. and E. Helpman (1995) 'International R&D Spillovers', *European Economic Review* 39, pp859-887.

Crespi, G. and Geune, A. (2004) *The Productivity of Science*, SPRU, University of Sussex. Available www.sussex.ac.uk/spru/documents/crespiost2.pdf (January 2010).

Cutler, T. *et al*. (2008) *Venturous Australia: building strength in innovation*, Cutler and Company, Melbourne.

Davis, P.M., Lewenstein, B.V., Simon, D.H., Booth, J.G. and Connolly, M.J.L. (2008) Open access publishing, article downloads and citations: randomised trial, *BMJ* 337: a568. Available http://dx.doi.org/10.1136/bmj.a568 (March 2010).

Davis, P.M. (2009) *Studies on Access: A Review*, Unpublished. Available http://arxiv.org/abs/0912.3953 (March 2010).

Denison, E.F. (1985) *Trends in American Economic Growth, 1929-1982,* Brookings Institution, Washington DC.

Department of Trade and Industry (DTI) (2007) *Measuring economic impacts of investment in the research base and innovation: a new framework for measurement*, Department of Trade and Industry, London.

DEST (Department of Education, Science and Training) (2003) *Mapping Australian Science and Innovation: Main Report*, Canberra.

Dowrick, S. (2003) *A Review of the Evidence on Science, R&D and Productivity.* Department of Education, Science and Training, Canberra.

Dowrick, S. and Rogers, M. (2002) 'Classical and Technological Convergence: Beyond the Solow-Swan Growth Model,' *Oxford Economic Papers* 54, pp369-385.

EPS, *et al.* (2006) *UK scholarly journals: 2006 baseline report – An evidence-based analysis of data concerning scholarly journal publishing*, Research Information Network, Research Councils UK and Department of Trade and Industry. Available at http://www.rin.ac.uk/data-scholarly-journals (December 2009).

Evans, D. and Sezer, H. (2002) 'A time preference measure for the social discount rate for the UK,' *Applied Economics* 34, pp1925-1934.

Evans, J. A. and Reimer, J. (2009) Open Access and Global Participation in Science, *Science* 323: 1025-. Available http://dx.doi.org/10.1126/science.1154562 (January 2010).

Frantzen, D. (2000) 'R&D, Human Capital and International Technology Spillovers: A cross-country analysis,' *Scandinavian Journal of Economics* 102(1), pp57-75.

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T. and Harnad, S. (2010) 'Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research,' Available http://arxiv.org/abs/1001.0361 (February 2010).

Geuna, A. and Arundel, A. (2003). *Proximity and the Use of Public Science by Innovate European Firms*, Royal Economic Society Annual Conference 2003 86, Royal Economic Society.

Graig, I.D., Plume, A.M., McVeigh, M.E., Pringle, J. and Amin, M. (2004) *Do Open Access Articles Have Greater Citation Impact? A critical review of the literature*, Publishing Research Consortium.

Griffith, R., Redding, S. and Van Reenen, J. (2004) 'Mapping the two faces of R&D: Productivity growth in a panel of OECD industries,' *Review of Economic Statistics* 8, pp883-895.

Griliches, Z. (1986) 'Productivity, R&D and basic research at the firm level in the 1970s,' *American Economic Review* 76, pp141-154.

Griliches, Z. (1995) 'R&D and productivity: Econometric Results and Measurement Issues,' In Stoneman, P. (Ed.) (1995) *Handbook of The Economics of Innovation and Technological Change.* Blackwell, Oxford, pp52–89.

Guellec, D. and van Pottelsberghe de la Potterie, B. (2001) 'R&D and Productivity Growth: Panel data analysis of 16 OECD countries,' *OECD Economic Studies 33*, pp103-126.

Hajjem, C, Harnad, S. and Gingras, Y. (2005) 'Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact,' *IEEE Data Engineering Bulletin* 28(4), pp39-46.

Hall, B.H. (2007) *Measuring the Returns to R&D: The Depreciation Problem*, NBER Working Paper 13473, NBER, Cambridge MA.

Hall, B.H., Mairesse, J. and Mohnen, P. (2009) *Measuring the returns to R&D*, NBER Working Paper 15622, NBER, Cambridge MA.

Hall, B.H., Mairesse, J. and Mohnen, P. (2010) Measuring the returns to R&D, in eds. Hall, B.H. and Rosenburg, N. (2010) *Handbook of the Economics of Innovation*, North Holland (forthcoming).

Harnad, S. and Brody, T. (2004) 'Comparing the impact of open access (OA) vs. non-OA articles in the same journals', *D-Lib Magazine* 10(6), June 2004. Available http://www.dlib.org/dlib/june04/harnad/06harnad.html (February 2010).

Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y, Oppenheim, C., Stamerjohanns, H., and Hilf, E. (2004) 'The Access/Impact Problem and the Green and Gold Roads to Open Access,' *Serials Review* 30(4) 2004. Available http://dx.doi.org/10.1016/j.serrev.2004.09.013 (February 2010).

Harnad, S. (2005) 'OA Impact Advantage = EA + (AA) + (QB) + QA + (CA) + UA.' Available http://eprints.ecs.soton.ac.uk/12085/1/OAA.html (March 2010).

Harrison, M. (2007) 'Valuing the future: Choosing the discount rate in cost-benefit analysis,' OPBR Cost Benefit Analysis Conference, November 2007.

Houghton, J.W. (2005) 'Economics of Publishing and the Future of Scholarly Communication' in Eds. Gorman, G.E. & Rowland, F. (2005) *International Year Book of Library and Information Management 2004-2005: Scholarly Publishing in an Electronic Era*, Facet Publishing, London.

Houghton, J.W. Steele, C. and Sheehan, P.J. (2006) Research Communication Costs in *Australia, Emerging Opportunities and Benefits*, Department of Education, Science and Training, Canberra. Available http://eprints.vu.edu.au/538/ (February 2009).

Houghton, J.W. and Sheehan, P.J. (2009) 'Estimating the Potential Impacts of Open Access to Research Findings,' *Economic Analysis and Policy* 39(1), March 2009. Available http://www.eap-journal.com/ (February 2010).

Houghton, J.W., Rasmussen, B., Sheehan, P.J., and Oppenheim, C., Morris, A., Creaser, C., Greenwood, H., Summers, M. and Gourlay, A. (2009) *Economic Implications of Alternative Scholarly Publishing Models: Exploring the Costs and Benefits*, The Joint Information Systems Committee (JISC), London & Bristol. Available http://eprints.vu.edu.au/15222/ (February 2010).

Houghton, J.W., de Jonge, J. and van Oploo, M. (2009) *Costs and Benefits of Research Communication: The Dutch Situation, S*URFfoudation, Utrecht. Available http://www.surffoundation.nl/wiki/display/economicstudyOA/Home (December 2009).

Houghton, J.W. (2009) *Costs and Benefits of Alternative Publishing Models*: *Denmark*, DEFF, Copenhagen. Available http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fDK_Costs_and_benefits_of_alternative_publishing_models.pdf (December 2009).

Houghton, J.W. (2009) *Open Access: What are the economic benefits? A comparison of the United Kingdom, Netherlands and Denmark*, Knowledge Exchange, Brussels. Available http://knowledge-exchange.info/Default.aspx?ID=316 (December 2009).

Industry Commission (1995) *Research and Development,* Report No 44, Industry Commission, Canberra.

Jaffe, A. (1989) 'Real effects of academic research,' *American Economic Review* 79, pp957-970.

Jones, C.I., and Williams, J.C. (1998) 'Measuring the Social Return to R&D,' *Quarterly Journal of Economics* 113(4), pp1119-1135.

Katz, J.S. and Hicks, D. (1997) *How much is collaboration worth? A calibrated bibliometric model*, Brighton: University of Sussex (SPRU).

Katz, J.S. and Martin, B.R. (1997) 'What is Research Collaboration?,' *Research Policy 26*, pp1-18. Available www.sussex.ac.uk/Users/sylvank/pubs/Res_col9.pdf (March 2003).

King, D.W. and Alvarado-Albertorio, F.M. (2008) 'Pricing and other means of charging for scholarly journals: a literature review and commentary,' *Learned Publishing* 21(4), pp248-272. Available http://dx.doi.org/10.1087/095315108X356680 (September 2009).

King, D.W. and Tenopir, C. (2004) 'An evidence-based assessment of the author-pays model', *Nature,* Web Focus: Access to the Literature. Available www.nature.com/nature/focus/accessdebate (September 2009).

King, D.W., Aerni, S., Brody, F., Herbison, M. and Kohberger, P. (2004) *Comparative cost of the University of Pittsburgh electronic and print library collections,* University of Pittsburgh, May 2004.

King, D.W., Lancaster, F.W., McDonald, D.D., Roderer, N.K. and Wood, B.L. (1976) *Statistical Indicators of Scientific and Technical Information Communication (1960-1980)*, 3 Volumes, Report to the National Science Foundation C-878. King Research, Inc., Center for Quantitative Sciences (GPO 083-000-00295-3).

King, D.W., Montgomery, C.H., Boyce, P. and Tenopir, C. (2003) 'Library economic measures: Examples of the comparison of print & electronic journal collections,' *Library Trends 51*(3), pp376-400.

King, D.W., Tenopir, C. and Clarke, M. (2006) 'Measuring total reading of journal articles,' *D-Lib Magazine* 12(10), October 2006. Available www.dlib.org/dlib (March 2009).

Kline, S., Rosenberg, N., (1986) 'An overview of innovation,' In Landau, R., Rosenberg, N. (Eds.), *The Positive Sum Strategy: Harnessing Technology for Growth*, National Academy Press, Washington DC.

Mansfield, E. (1991) 'Academic research and industrial innovation,' *Research Policy* 20(1), 1991, pp.1-12.

Mansfield, E. (1998) 'Academic research and industrial innovation: an update of empirical findings,' *Research Policy* 26(7/8), 1998, pp.773-776.

Martin, B.R. and Tang, P. (2007) *The benefits of publicly funded research*, SWEPS Paper No. 161, Science Policy Research Unit, University of Sussex, Brighton. Available http://www.sussex.ac.uk/spru (December 2008).

Martin, B.R., Salter, A., Hicks, D., Pavitt, K., Senker, J., Sharp, M. and von Tunzelmann, N. (1996) *The Relationship between Publicly Funded Basic Research and Economic Performance*, A SPRU Review, HM Treasury, London.

Matsumoto, M. (2008) *Assessing the impacts of public R&D in National Institute of Advanced Industrial Science and Technology (AIST)*, OECD, Paris. Available http://www.oecd.org/document/10/0,3343,en_2649_34273_40854218_1_1_1_1,00.html (February 2010).

National Science Board (NSB) (2010) *Science and Engineering Indicators 2010*, National Science Foundation (NSB 10-01), Arlington, VA. Available http://www.nsf.gov/statistics/ (February 2010).

National Institutes of Health (NIH) (2008) *Analysis of Comments and Implementation of the NIH public access policy*, NIH, Washington DC. Available http://publicaccess.nih.gov/analysis_of_comments_nih_public_access_policy.pdf (February 2010).

OECD (2005) *Digital Broadband Content: Scientific Publishing*, Organisation for Economic Cooperation and Development, Paris. Available http://www.oecd.org/dataoecd/42/12/35393145.pdf (February 2010).

OTA (1986) *Research Funding as an Investment: Can We Measure the Returns?* Office of Technology Assessment, Washington DC.

Productivity Commission (2007) *Public support for science and innovation,* Productivity Commission, Canberra.

Rand, D. G. and Pfeiffer, T. (2009) 'Systematic Differences in Impact across Publication Tracks at PNAS,' *PLoS ONE*, 4(12). Available http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0008092 (February 2010).

RIN (2009) *Overcoming barriers: access to research information content*, Research Information Network, London. Available http://www.rin.ac.uk/system/files/attachments/Sarah/Overcoming-barriers-report-Dec09_0.pdf (February 2010).

Rowlands, I. and Olivieri, R. (2006) *Journals and Scientific Productivity: A case study in immunology and microbiology*, Publishing Research Consortium, Bristol. Available http://www.publishingresearch.net/documents/Rowland_Olivieri_summary_paper.pdf (March 2009).

Salter, A.J. and Martin, B.R. (2001) 'The economic benefits of publicly funded basic research: a critical review,' *Research Policy* 30(3), pp509-532.

Scott, A., Steyn, G., Geuna, A., Brusoni, S. and Steinmueller, E. (2002) *The Economic Returns to Basic Research and the Benefits of University-Industry Relationships.* Report to the Office of Science and Technology, London.

Shanks, S. and Zheng, S. (2006) *Econometric modeling of R&D and Australia's productivity,* Staff Working Paper, Productivity Commission, Canberra.

Solow, R.M. (1957) 'Technical Change and the Aggregate Production Function,' *Review of Economics and Statistics* 39, pp312-320.

Solow, R.M. (1987) Growth Theory and After. *R.M. Solow – Prize Lecture*, Nobel e-Museum Laureates.

Sveikauskas, L. (2007) *R&D and Productivity Growth: A Review of the Literature,* US Bureau of Labor Statistics Working Paper 408, BLS, Washington, DC. Available www.bls.gov/osmr/pdf/ec070070.pdf (February 2010).

Swan, A. (2010a) *Modelling Scholarly Communication Transitions: Costs and benefits for universities,* Report to the Joint Information Systems Committee, Bristol and London. Available http://ie-repository.jisc.ac.uk/442/ (February 2010).

Swan, A. (2010b) The Open Access citation advantage: Studies and results to date. Available http://eprints.ecs.soton.ac.uk/18516/ (March 2010).

Tenopir, C. and King, D.W. (2000) *Towards Electronic Journals: Realities for Scientists, Librarians and Publishers,* Special Libraries Association, Washington D.C.

Tenopir, C. and King, D.W. (2002) 'Reading behavior and electronic journals,' *Learned Publishing* 15(4), pp259-265.

Tenopir, C. and King, D.W. (2007) 'Perceptions of value and value beyond perceptions: measuring the quality and value of journal article readings,' *Serials* 20(3), pp199-207. Available http://www.uksg.org/serials (September 2009).

Tenopir, C., King, D.W., Edwards, S. and Wu, L. (2009) 'Use of Scholarly Journals and Articles by University Faculty: Changes in Information Seeking and Reading Patterns over Nearly Three Decades,' *Aslib Proceedings*, 2009.

Tewksbury, J.G., Crandall, M.S. and Crane, W.E. (1980) 'Measuring the Societal Benefits of Innovation,' *Science* 209(4457), pp658-662.

Verspagen, B. (2004) *The Impacts of Academic Knowledge on Macroeconomic Productivity Growth: An Exploratory Study*, Eindhoven Centre for Innovation Studies, Eindhoven.

Walsh, J.P., and Maloney, N.G. (2001) 'Computer Network Use, Collaboration Structures and Productivity.' In P. Hinds and S. Kiesler, (eds.) *Distributed work,* Cambridge, MA: MIT Press. Available http://tigger.uic.edu/~jwalsh/Collab.html (March 2008).

Ware, M. (2007) *Peer review in scholarly journals: Perspectives of the scholarly community – an international study*, Publishing Research Consortium, Bristol. Available http://www.publishingresearch.net/documents/PeerReviewFullPRCReport-final.pdf (January 2010).

Ware, M. (2009) *Access by UK small and medium-sized enterprises to professional and academic literature,* Publishing Research Consortium, Bristol. Available http://www.publishingresearch.net/SMEaccess.htm (January 2010).

Wheatley, P., Ayris, P., Davies, R., McLeod, R. and Shenton, H. (2007) 'LIFE: costing the digital preservation lifecycle,' iPRES Annual Conference, 10 October 2007, Beijing, China. Available http://www.life.ac.uk (January 2010).

Zhang, Y. (2006) 'The Effect of Open Access on Citation Impact: A Comparison Study Based on Web Citation Analysis,' *Libri* 56(3), pp145-156. Available http://www.librijournal.org/pdf/2006-3pp145-156.pdf (March 2010).