



COMMISSION OF THE EUROPEAN COMMUNITIES

Brussels, 14.2.2007
SEC(2007) 181

COMMISSION STAFF WORKING DOCUMENT

Document accompanying the

**COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN
PARLIAMENT, THE COUNCIL AND THE EUROPEAN ECONOMIC AND SOCIAL
COMMITTEE**

on scientific information in the digital age: access, dissemination and preservation

{COM(2007) 56 final}

CONTENTS

1.	Introduction	5
1.1.	Aim of the staff working paper	5
1.2.	Preparatory work and consultations	5
2.	Access to and dissemination of scientific information	6
2.1.	The importance of efficient access to and dissemination of scientific information.....	6
2.1.1.	Access, dissemination and the Lisbon Agenda.....	6
2.1.2.	The European Union as a big investor in research.....	7
2.2.	Access and dissemination issues.....	8
2.2.1.	Opportunities of the digital age.....	8
2.2.2.	Scientific publishing: central role and rising costs	10
2.2.3.	Scientific publishing and research excellence.....	12
2.2.4.	Disciplinary differences	12
2.2.5.	Financing available for dissemination as a percentage of total R&D spending	13
2.2.6.	VAT on scientific publications	13
2.2.7.	Intellectual property rights issues.....	16
2.3.	Access and dissemination: initiatives and developments.....	17
2.3.1.	Publishers' new business models	17
2.3.2.	The Open Access movement and resulting trends	18
2.3.3.	Open repositories	21
2.3.4.	Activities of legislatures, funding bodies and international organisations	24
3.	Preservation issues	26
3.1.	The importance of digital preservation	26
3.2.	Preservation: initiatives and developments.....	28
4.	Overview of relevant Community initiatives.....	30
4.1.	i2010 digital libraries	30
4.2.	EU research policy	30
4.3.	Investigation, public consultation and interaction with stakeholders	30

Executive Summary

The present staff working paper accompanies the Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on scientific information in the digital age: access, dissemination and preservation. The aim of the Communication is to signal the importance of and launch a policy process on (a) improving access to and dissemination of scientific information, and (b) developing strategies for the preservation of scientific information across the Union. This staff working paper aims to present facts, evidence and examples relating to the scientific publishing and information systems that are relevant to the issues dealt with in the Communication.

The working paper highlights the importance of widespread and efficient access to scientific information for research and innovation and for the public at large. It shows that the new information society tools have changed and will continue to change the way in which researchers can access, share and use scientific information, i.e. journal articles and research data.

The paper stresses that the importance of research data is likely to grow in the coming years. Information society tools make it possible to access data directly. New information services are combining journal articles and data, and applying new search techniques such as data mining. The paper gives a number of concrete examples and describes how these developments are linked with e-science infrastructures and with relevant intellectual property rights issues (e.g. issues related to the use of digital rights management systems and to the database directive).

Much of the debate revolving around access to scientific information has focused on scientific publications in journals. Journal articles play a key role in the scientific information system. The peer review function undertaken by scientific publishers is crucial for the scientific community as it represents a key quality control mechanism. Moreover, publishing an article in a journal enjoying a high reputation is very important for a scientist's career as it determines potential promotions and chances of future research funding. Scientific publishing is also a significant economic activity in its own right.

With the advent of the internet, publishers' dissemination models have changed. Some 90% of journals are now available online (in most cases accessible on a subscription basis), although there are differences between the disciplines (in particular between STM journals and arts and humanities journals). Over the last few years, the number of journals has increased steadily, by some 3.5% a year, reflecting the growing R&D output. The total budgets for the dissemination of information and relevant library budgets have not increased at the same pace. New business models — including “big deals” with individual libraries and country deals — have been devised to respond to libraries' budgetary constraints. However, these deals, by tying considerable parts of libraries budgets, have had side-effects on their capacity to purchase monographs and other journals. The staff working paper also describes the effects of higher VAT rates for digital publications — as opposed to paper publications — on libraries' budgets, as well as mechanisms used in some Member States to alleviate these effects.

One development of the digital age has been the “open access” movement, which pursues the goal of making scientific articles freely accessible on the web. The working paper outlines the main policy statements on open access and the main types of experiments with open access:

open access publishing and self-archiving. A key funding model for open access is the “author pays” scheme, where the author (often the institution or research funding body behind the author) pays for publication costs instead of readers doing so through subscriptions. It is still too early to draw conclusions on the economic viability of this model.

A parallel development is self-archiving, whereby authors post a version of the article (typically the revised manuscript after peer review) on an open web-based repository. Some funding bodies (Wellcome Trust, UK Research Councils) have started requiring deposit in an open repository after a specified embargo period and are also prepared to fund publication costs under the author pays model to allow immediate open access. In this context, the European Organisation for Nuclear Research (CERN) is experimenting with a move towards the open access model in the area of particle physics.

In recent years universities and research communities have set up open repositories — institutional or domain-based — to make the information they produce more accessible. The number of repositories, as well as their size and quality, varies widely among the Member States. Some of the repositories contain mainly journal articles, others combine articles with data, audiovisual and other digital material. An example of this integrated approach is the DARE programme in the Netherlands.

Digital preservation concerns the long-term preservation and accessibility of “born-digital” journals and underlying research data. The digital files may become inaccessible in future as software and hardware changes. The working paper gives examples of the importance of data from the past for present research (e.g. on climate change), and underlines the potential of the area for economic activity. A survey of 1 200 organisations in industrialised countries indicated that 70% of respondents considered the management (and preservation) of electronic information to be of critical importance.

Between 2007 and 2013, the European Community has agreed to invest some €50 billion in the 7th Framework Programme for research and development. By the sheer size of this financial contribution, it is a major stakeholder in the debate on the access to and preservation of scientific information. Over the last few years, the 5th and 6th R&D Framework Programmes have co-funded several experiments to improve access to and dissemination of scientific information, and to develop strategies for digital preservation. The staff working paper describes several such projects to exemplify developments in this area.

1. INTRODUCTION

1.1. Aim of the staff working paper

The present staff working paper accompanies the Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on scientific information in the digital age: access, dissemination and preservation.

The aim of the Communication is to signal the importance of and launch a policy process on (a) improving access to and dissemination of scientific information¹, and (b) developing strategies for the preservation of scientific information across the Union.

This staff working paper aims to present facts, evidence and examples relating to the scientific publication and information systems² that are relevant to the issues dealt with in the Communication.

1.2. Preparatory work and consultations

The Communication on scientific information in the digital age: access, dissemination and preservation and this staff working paper have greatly benefited from consultations and interaction with relevant stakeholders.

A public online consultation on the EC-commissioned “Study on the economic and technical evolution of the scientific publication markets in Europe” was held from March to June 2006³. The study generated 170 responses from publishers, individual researchers, academic organisations, libraries and information organisations. While respondents from the research community and libraries welcomed the study and its recommendations, publishers gave very critical appraisals⁴. An online consultation on the digital libraries initiative, held between September 2005 and January 2006, also yielded a number of comments in relation to scientific and scholarly information, although the consultation did not specifically address this area⁵.

¹ For the purposes of this document, “scientific information” comprises publications and research data.
² The “scientific publication system” can be understood as the practices, rules and mechanisms governing the process of scientific publication, as well as its exploitation. The term “scientific information system” comprises the same type of mechanisms as “scientific publication system”, but is broader in that it also covers the research data underpinning publications and research results not published in journals or monographs.

³ http://ec.europa.eu/research/science-society/page_en.cfm?id=3185.

⁴ An overview of the replies to the consultation can be found at http://ec.europa.eu/research/science-society/document_library/pdf_06/synthesis-consultation_en.pdf. Individual responses can be read at http://ec.europa.eu/research/science-society/page_en.cfm?id=3186.

⁵ An overview of the replies to the consultation can be found at http://europa.eu.int/information_society/activities/digital_libraries/doc/results_online_consultation/en.pdf. Individual responses can be read at http://europa.eu.int/information_society/activities/digital_libraries/cultural/actions_on_consultations/online_consultation/index_en.htm.

A number of workshops on scientific information questions was organised with stakeholder groups (in particular commercial publishers, learned societies publishers, and organisations in favour of open access). In parallel, a considerable number of bilateral contacts with publishers, the research community, libraries and funding agencies were developed.

Discussions within relevant groups such as the European Research Advisory Board (EURAB) were also held. Within the High Level Expert Group on Digital Libraries, which brings together stakeholders on digital libraries issues, a sub-group addresses scientific information questions.

2. ACCESS TO AND DISSEMINATION OF SCIENTIFIC INFORMATION

2.1. The importance of efficient access to and dissemination of scientific information

2.1.1. Access, dissemination and the Lisbon Agenda

In March 2000, European leaders announced the Lisbon Agenda, which can be summarised as the goal for the EU of becoming, by 2010, “the most dynamic and competitive knowledge-based economy in the world capable of sustainable economic growth with more and better jobs and greater social cohesion, and respect for the environment”, where others compete with cheap labour or primary resources. In 2005, the European Commission presented the Growth and Jobs Strategy, a revised version of the Lisbon Agenda. Sustained economic growth and more and better jobs are its priorities.

One of the main objectives for the achievement of the Lisbon Agenda is the investment of three percent of GDP in research and development by EU Member States by 2010. In connection with the fact that all research builds on earlier work, and depends on scientists’ possibilities to share and access scientific publications and research data, the efficiency of the system for dissemination of and access to research results and data significantly contributes to overall technological advance, and is essential for innovation and economic performance⁶.

The environment in which research is conducted and disseminated is undergoing profound change, with new technologies offering new opportunities and changing research practices demanding new capabilities. New opportunities and new models could enhance the dissemination of research findings and maximise the returns on investment in R&D. The potential benefits of better and quicker access to scientific information for the efficiency of research include⁷:

- acceleration of the research and discovery process, leading to increased returns on R&D investment;
- avoidance of duplicative research efforts, leading to savings in R&D expenditure;

⁶ Access to and use of data must respect the rules on the protection of personal data as laid down in particular in EU Directives 95/46/EC and 2002/58/EC (OJ L 226, 22.9.1995, p. 1, and OJ L 201, 31.7.2002, p. 37).

⁷ On the following, see J.Houghton, C.Steele and P.Sheehan, Research Communication Costs in Australia: Emerging Opportunities and Benefits, 2006.

- enhanced opportunities for multi-disciplinary research, as well as inter-institutional and inter-sectoral collaborations.

Potential benefits of enhanced access for innovation include:

- broader and faster opportunities for adoption and commercialisation of research findings, generating increased returns on public investment in R&D;
- the potential for the emergence of new industries based on scientific information⁸.

Finally, enhanced access helps inform citizens about scientific progress and results. For example, eighty percent of US internet users, or some 113 million adults, searched for information on health topics in 2005⁹.

2.1.2. *The European Union as a big investor in research*

As a significant investor in R&D, the European Union's stakes in efficient access and dissemination are high. Under the Lisbon Agenda, the EU's goal in the area of research and development is to achieve an R&D intensity (i.e. expenditure as a percentage of GDP) of at least 3% for the EU by 2010, and to have two thirds of R&D expenditure financed by the business sector. In 2004, the EU25 spent nearly €200 billion on R&D. Its R&D intensity stood at 1.90% as against 1.92% in 2003. R&D intensity remains significantly lower in the EU25 than in other major economies.

The budget of the Framework Programme for research and development amounted to around €19 billion (at 2004 prices) for the four-year period 2002-2006 under FP6. It will reach about €50 billion (or €48 billion at 2004 prices) for the seven-year period 2007-2013 under FP7. In 2006, research accounted for 4 percent of the EU budget. The Framework Programme accounts for about 6 percent of EU15 non-military governmental RTD expenditure¹⁰.

In line with the EU's high investment in research, Framework Programme projects generate many publications and thereby contribute directly to Europe's total output in terms of scientific publications. Some indications on the publication output of Community (co-)funded research can be given on the basis of earlier programmes. In the majority (79 percent) of projects under the 4th and 5th Framework Programmes with UK participation surveyed, at least one peer-reviewed publication was produced, while in around 10 percent of those projects more than 20 such outputs were produced¹¹. An analysis of impact variables resulting from research projects in the fisheries and aquaculture domain of the FAIR programme under FP4 identified 711 publications in 219 peer-reviewed journals by the participants in 82

⁸ Examples of new industries built on publicly accessible data include weather derivatives based on meteorological data.

⁹ Online Health Search 2006, PEW Research 2006.

¹⁰ Court of Auditors, Special Report No 1/2004 on the management of indirect RTD actions under the Fifth Framework Programme (FP5) for Research and Technological Development (1998 to 2002), together with the Commission's replies (pursuant to Article 248(4) second subparagraph EC) (2004/C 99/01), 23 April 2004, paragraph 5.

¹¹ DTI – Office of Science and Technology, EU Framework Programmes, p. 22.

projects, an average of 8.7 peer-reviewed publications per project¹². Participants in the transport programme under FP4 reported 3 766 publications for 269 shared-cost projects, an average of 14 publications per project¹³.

2.2. Access and dissemination issues

2.2.1. Opportunities of the digital age

The advent of the internet in the mid-1990s has revolutionised the ways in which scientific information can be accessed and disseminated, and opened up many new opportunities.

In terms of publication in journals, the digital age encourages publishers to adopt digital delivery and to provide online access to their journals. The majority of journals are now available online (most are also in parallel published in print). A study by Cox in 2005 (based on a publisher survey) found that 90% of all journals were online (93% of STM and 84% of arts and humanities journals). Data in Ulrich's Periodicals Directory suggest a lower proportion of 62% online. The likely reason for the divergence is that Ulrich's is more representative of the total global situation, while Cox's sample reflects the more advanced development of the US and European publishing industry¹⁴. According to LISU, library expenditure on printed serials is declining as a proportion of total spend and now accounts for around 37 percent of serials spend, while electronic-only journals account for 26 percent and joint electronic/print subscriptions for 37 percent of spend¹⁵.

Despite evidence that new technologies and the internet have considerably improved access to scientific information¹⁶, some studies suggest that it could still be much improved. For example, the UK scholarly journals 2006 report entitled "An evidence-based analysis of data concerning scholarly journal publishing" states that just under 50% of researchers experience problems accessing research resources¹⁷. While this assessment covers all resources, a particular problem stems from the fact that individual libraries often do not stock specific journals. In this respect, the access issue is particularly acute for less well-endowed institutions and in countries with lower income levels.

In the area of research data, new technologies have opened and are still opening new frontiers to high-speed communication, data storage capacity, distribution and sharing of resources and high-performance computation. In certain areas of science (radio-astronomy, environment, health), new technologies and instrumentation are at the basis of "data factories" producing very large amounts of data, which in turn become the raw material on which new experiments

¹² Gesche Pluem, Analysis of Impact Variables Resulting from Research Projects in the Fisheries and Aquaculture Domain of the European Commission's Fair Programme (1994-1998), 2003, pp. 12, 14.

¹³ European Commission, Competitive and Sustainable Growth, June 2000, p. 28.

¹⁴ Scientific publishing in transition: an overview of current developments, Mark Ware Consulting, 2006.

¹⁵ UK scholarly journals: 2006 baseline report. An evidence-based analysis of data concerning scholarly journal publishing, 2006.

¹⁶ <http://www.dlib.org/dlib/october03/king/10king.html>.

¹⁷ UK scholarly journals: 2006 baseline report. An evidence-based analysis of data concerning scholarly journal publishing, 2006.

are carried out. The full “life-cycle of science”, from data acquisition to modelling and development of technology, is heavily influenced today by the availability of new ICT technologies.

One crucial change can be summarised by the term “e-science”. E-science is computationally intensive science conducted in highly distributed network environments or using large datasets that require shared computing facilities and high connectivity. It enables remote interaction with national and international instrument-based facilities. Examples of the kind of disciplines in which e-science is becoming common practice include social simulations, particle physics, earth sciences and bio-informatics.

For example, particle physics has a particularly well developed e-science infrastructure due to its need for adequate computing facilities for the analysis of results and storage of data originating from the European Organisation for Nuclear Research (CERN)’s Large Hadron Collider (LHC), which is due to start taking data in 2007. The wealth and diversity of data collected and stored is growing rapidly as automation increases and technological costs diminish. Collecting, preserving and making data available for re-use is a crucial mechanism in collaborative scientific research. Many researchers share the effort and pool their data and knowledge in each research community. This provides a rapid mechanism for distributed communication. In many cases, the database is then published: within an organisation, within a community or publicly.

The **UK National Crystallography Service (NCS)** has developed a prototype e-science infrastructure for the provision of a small molecule crystallography service, from sample receipt to results dissemination. Access to the NCS facilities and expertise and a mechanism to submit samples is granted through a secure grid infrastructure, which seamlessly provides instantaneous feedback and the ability to remotely monitor and guide diffraction experiments and stage the diffraction data to a securely accessible location. Publication of all the data and results generated during the course of the experiment, from processed data to analysed structures, is then enabled by means of an open access data repository. The repository publishes its content through established digital library protocols, which enable harvester and aggregator services to make the data searchable and accessible¹⁸.

Another new technique is data mining. New information services are emerging that build on the results of individual research projects, gathering together scientific literature and data, and using data mining techniques. More and more examples of the use of text mining technology in the academic field and in the commercial sector demonstrate the value of this technology for such diverse applications as content production for bio-databases, mining of electronic health records and mining the abstracts of journal articles. Peer-reviewed journals play a critical role in the selection of relevant raw material and are frequently the starting point for the development of new information resources. An interesting example is the Cochrane Library.

The Cochrane Library. In the healthcare sector it is a real challenge to keep up to date with the relevant evidence in any field of interest. The Cochrane Library is a growing source of

¹⁸ www.ukoln.ac.uk/ukoln/staff/e.j.lyon/publications.html.

reliable evidence about the effects of health care. Evidence-based results from among the major bibliographic databases are collected and evidence is put together for and against the effectiveness and appropriateness of treatments (medications, surgery, education, etc.) in specific circumstances¹⁹.

Patents are also an important scientific and technical information resource with great potential. Given their public domain nature, facilitating its digitisation and online access would also contribute to enhancing access to and preservation of scientific information. Some initiatives have been taken to this end by both institutional and private entities. The European Patent Office – a major player in the digital patent databases and related services – has set up espace@net. More recently Google²⁰ has launched a service that aims at enabling the wider public to access patent information in a structured and user-friendly way.

2.2.2. *Scientific publishing: central role and rising costs*

The importance of scientific publishing lies in its role in the selection, production and spread of scientific and technical knowledge, and how this dissemination of knowledge drives economic growth and further research. Scientific publishers traditionally have a series of functions, including registration (establishing the author's precedence), peer review (aiming to ensure quality control), dissemination (communicating findings to the intended audience), and archival record (preserving a fixed version of the paper for future reference).²¹

With the advent of the digital revolution scientific publishers have also started to provide new added value services (e.g. Science Direct²²) to researchers as well as citation navigation, that is, providing filters and signposts to relevant work amid the huge volume of published material.

All of the functions mentioned are important, but the peer review process is critical for the scientific community as it rubber-stamps the quality of an article. It is the systematic assessment of submitted papers by independent experts from within the relevant research community with the objective of appraising whether the methodologies used, as well as the reasoning and evidence presented in the paper meet the interest and quality standards of the scientific discipline. Several studies indicate that researchers concentrate their reading on peer-reviewed published articles²³.

There are between 20 000 and 25 000 scholarly peer-reviewed journals worldwide²⁴, collectively publishing about 1.4 million articles a year²⁵. The number of peer-reviewed journals published annually has been growing at an annual rate of about 3.5% over the past years.

¹⁹ <http://www.cochrane.org>.

²⁰ <http://www.google.com/patents>

²¹ Scientific publishing in transition: an overview of current developments. Mark Ware Consulting, 2006.

²² <http://www.sciencedirect.com>.

²³ NOP & CIBER Research for Elsevier.

²⁴ Ulrich's web directory.

²⁵ Scientific publishing in transition: an overview of current developments. Mark Ware Consulting, 2006.

Scientific publishing of academic papers is a profitable and sustainable business, which traditionally operates on a subscription-based model. 780 publishing houses based in Europe are responsible for publishing 49% of all research articles²⁶; 36 000 full-time staff plus 10 000 freelancers, editors and staff working for suppliers are employed in this industry in Europe. European researchers publish 43% of the world's research papers²⁷, and it is estimated that Europe accounts for 24-32% of world expenditure on journals.

Over the last twenty years, journal subscription prices have on average increased much faster than inflation levels. This has put publicly funded libraries, their main clients, under financial pressure and has led to subscription cancellations in certain cases. Publishers argue that price increases are due mainly to the growth in numbers of articles submitted. As a result, in recent years, libraries' purchasing decisions have become more difficult as the number of scientific journals has increased steadily, by some 3.5% annually throughout the past decades. The growth in the number of scientific journals has been in line with the increasing expenditure for research. This constant growth, plus the rising prices of subscription journals, which have increased well above inflation over the last decades, combined with declining or stagnant library budgets, have made the acquisition of scientific publications unsustainable for most academic institutions. This is at the root of what is termed the "serials crisis". The statistics reproduced below illustrate this crisis:

- In the US, from 1986 to 2005, inflation reached 78%, journal prices rose by 167% and book prices rose by 81%. The typical research library spent 302% more on serials in 2005 than in 1986, but the number of titles purchased increased by only 42%. The average annual increase on journal spending from 1986 to 2005 was 7.6%, whilst the annual inflation increase reached 3.1%. These means that the annualised price increase for journals over the 1986–2006 periods was 4.5% above inflation.²⁸
- In the UK, the average price of an academic journal rose by 58% between 1998 and 2003, while the inflation rate was 11% over the same period. Although the proportion of university library expenditure on serials increased it could not maintain serials purchasing power²⁹.
- The Chartered Institute of Library and Information Professionals (CILIP) reported that between 1996/1997 and 2000/2001 the average journal price increased by 41%, while over the same period the information resource budget of UK university libraries decreased by 29% in real terms.

²⁶ Reed Elsevier submission to the EC-commissioned study on scientific publishing markets in Europe.

²⁷ National Science Foundation. Science and Engineering Indicators. 2006.

²⁸ <http://www.arl.org/stats/arlstat/05pub/05intro.html>

²⁹ Memorandum from CURL (Consortium of University Research Libraries) and SCONUL (Society of College, National and University Libraries), presented before the UK House of Commons Science and Technology Committee, 2004.

2.2.3. *Scientific publishing and research excellence*

The impact of the scientific publication system on research excellence is a widely recognised but as yet insufficiently studied phenomenon.

For example, the so-called “publish or perish” imperative, signifying that researchers must publish in order to develop their scientific careers, remains dominant. Moreover, it is not enough to simply publish: in many disciplines, researchers must also publish in journals with high “impact factors” in order to be successful in academia.

Recent studies³⁰ indicate that communicating the results of their research to peers remains the primary reason for researchers to publish their work. Researchers wish to share their findings with their peers, with the aim of making an impact in their field. It is widely perceived that publishing in high impact journals provides researchers with increased opportunities for career advancement. A further reason for researchers to publish their work is to establish their personal prestige. By publicising their findings, researchers boost their eligibility for future funding. Finally, scientists publish to establish precedence for their work.

There is no ideal way to assess the quality of published content. The most common method is to use science citation impact factors (Thompson ISI factor). A science citation impact factor is determined for a particular journal according to the number of times scientific papers in that journal are cited (quoted) elsewhere in the scientific literature (in other journals). Important and influential journal articles are typically cited many times. Examples of journals with a very high impact factor — as assessed by science citation results — include the *New England Journal of Medicine*, *Cell* and the *British Medical Journal*.

The EC-commissioned “Study on the economic and technical evolution of the scientific publication markets in Europe” suggested the need to establish alternative evaluation mechanisms to supplement the Thompson ISI citation data. The responses to this idea in the public consultation on the Study’s recommendations showed some support for supplementary evaluation mechanisms such as web citations and analysis of journal article downloads, although other respondents argued that the current system needs no change.

2.2.4. *Disciplinary differences*

Many differences exist amongst disciplines; this issue was highlighted in the online consultation on the EC-commissioned “Study on the economic and technical evolution of the scientific publication markets in Europe”. Survey evidence suggests that journal articles are most important in the sciences and social sciences, but that books and monographs are more important in the arts and humanities. Moreover, digital technologies and the network have been taken up at a slower pace in humanities than in science, in part due to cultural barriers.

Another important difference is the language in which journals are published. The worldwide community in natural sciences and engineering communicates almost exclusively in English. As a result there is a huge potential readership base for publications in these areas. On the

³⁰ NOP & CIBER Research for Elsevier, Open access self-archiving: An author study. May 2005, Alma Swan and Sheridan Brown.

other hand, the circulation of journals in the humanities is more limited and the English language is not predominant.

2.2.5. *Financing available for dissemination as a percentage of total R&D spending*

Industry approximations put the total STM publishing market (which includes journals, books, secondary information services) at between €6.8 and 9 billion³¹. It is estimated that journal sales account for about 45-50 percent of the total STM publishing market and provide a 2004 market value of €3.6-4.1 billion. This figure should, however, be interpreted with caution as it does not include non-English language journals in the social sciences, humanities and arts.

The total European STM publishing market is estimated at between 24 and 32 percent of the total STM publishing market, namely €1.7-2.2 billion in 2004³². Using the average of the journals market estimates, this suggests a European STM scholarly journal market value of between €0.9 billion and €1.2 billion in 2004.

OECD countries spent €586 billion on R&D in 2004³³. In 2004, the EU25 spent nearly €200 billion or USD 260 billion on R&D. Combined with the figures on the publishing market presented above this would imply that budgets spent on journal subscriptions represent some 0.6% of the total European R&D budget. This is less than the Reed Elsevier estimate that 1% of the total R&D budget is spent on journal subscriptions.

2.2.6. *VAT on scientific publications*

Within the European Union, printed and electronic versions of publications are treated differently. While books, newspapers and periodicals are subject to a reduced rate of VAT (see Table 1), electronic publications are charged at the standard rate. This has consequences for European libraries and the market for scholarly publications in Europe, with special reference to the ongoing switch from printed to electronic publications.

The higher costs for electronic publications influence libraries' decisions when purchasing resources. As the EC-commissioned "Study on the economic and technical evolution of the scientific publication markets in Europe" indicates, the different VAT rates lead to a situation where it is cheaper for libraries to order printed versions of resources in addition to the electronic version even if only the electronic version is considered valuable.

³¹ UK scholarly journals: 2006 baseline report. An evidence-based analysis of data concerning scholarly journal publishing.

³² UK scholarly journals: 2006 baseline report. An evidence-based analysis of data concerning scholarly journal publishing.

³³ <http://www.oecd.org/dataoecd/39/19/37685541.pdf>

Table 1

<i>Country VAT</i>	<i>VAT rate for print journals</i>	<i>VAT rate for e-journals</i>
<i>Belgium</i>	<i>0%³⁴ / 6% / 21%³⁵</i>	<i>21%</i>
<i>Bulgaria</i>	<i>20%</i>	<i>20%</i>
<i>Czech Republic</i>	<i>5%</i>	<i>19%</i>
<i>Denmark</i>	<i>0%³⁶ / 25%</i>	<i>25%</i>
<i>Germany</i>	<i>7%</i>	<i>19%</i>
<i>Greece</i>	<i>4,5%</i>	<i>19%</i>
<i>Spain</i>	<i>4%³⁷/16%</i>	<i>16%</i>
<i>France</i>	<i>2,1%³⁸ / 19.6%</i>	<i>19,6%</i>
<i>Ireland</i>	<i>13,5%</i>	<i>21%</i>
<i>Italy</i>	<i>4%³⁹ / 20%</i>	<i>20%</i>
<i>Cyprus</i>	<i>5%</i>	<i>15%</i>
<i>Latvia</i>	<i>5%</i>	<i>18%</i>
<i>Lithuania</i>	<i>5%</i>	<i>18%</i>
<i>Luxembourg</i>	<i>3%</i>	<i>15%</i>
<i>Hungary</i>	<i>20%</i>	<i>20%</i>
<i>Malta</i>	<i>5%</i>	<i>18%</i>
<i>Netherlands</i>	<i>6%</i>	<i>19%</i>
<i>Austria</i>	<i>10%</i>	<i>20%</i>
<i>Poland</i>	<i>0%⁴⁰ / 7%⁴¹ / 22%</i>	<i>22%</i>

³⁴ Applies to daily and weekly publications of general information only.

³⁵ Publications which are published for purposes of advertisement or whose main purpose is publicity

³⁶ Newspapers normally published at a rate of more than one issue per month.

³⁷ Journals and magazines, which do not contain mainly advertisements (i.e. when the Editor's income from publicity does not exceed 75% of the total income).

³⁸ If registered with the publication agency.

³⁹ Newspapers, and daily news bulletins, dispatches from press agencies, magazines, except for pornographic periodicals and magazines

<i>Portugal</i>	5%	21%
<i>Romania</i>	9%	19%
<i>Slovenia</i>	8,5%	20%
<i>Slovakia</i>	19%	19%
<i>Finland</i>	0 ⁴² / 22%	22%
<i>Sweden</i>	ex ⁴³ 6%	25%
<i>United Kingdom</i>	0%	17,5%

These differences in the rates of VAT charged within Europe place European libraries in different situations compared with each other. Moreover, most libraries cannot recover VAT on their purchases because they are part of public, academic or non-profit organisations. Declining budgets, high VAT rates and rising prices of scientific publications have had an adverse impact on their purchase of publications. This affects not only the provision of knowledge within the EU; it also may affect the development of science in the EU as a whole, compared with countries without VAT charges on electronic publications like the United States.

Some EU countries have taken action to address this issue. For example in Sweden and Austria VAT is paid back to libraries. Likewise, in order to offset this issue, in Lithuania and the Czech Republic specific consortia have been set up by libraries to get VAT refunded⁴⁴.

VAT may also impact on the financing of European research. In fact, VAT on R&D costs incurred by taxable persons is deductible where the goods and services in question are used for VAT-taxable transactions. The current rules of Council Directive 2006/112/EC of 28 November 2006 concerning exemptions, public authorities and subsidies may affect this deductibility and hence place an unnecessary burden on certain research activities because deduction is only allowed for VAT costs incurred by taxable persons. The Commission has committed itself to reviewing these three issues, which need to be modernised. When reviewing the legislation, the Commission will critically examine the restrictions on the recovery of VAT on R&D expenditure. It will also consider to what extent the current rules on public authorities and subsidies hamper the creation of public-private partnerships and cost-sharing arrangements, including in the research field where such structures are increasingly being used to conduct R&D efforts requiring the pooling of resources from public and private entities or the outsourcing of research by private entities to public ones (contract research).

⁴⁰ Specialist periodicals subject to certain conditions

⁴¹ Newspapers, magazines, periodicals, excluding: publications with no less than 67% of space dedicated to free of charge or paid commercial announcements or advertisements

⁴² Newspapers and periodicals subscribed to for a period of at least one month

⁴³ Staff periodicals and periodicals issued by non-profit organisations are exempted

⁴⁴ <http://www.sub.uni-goettingen.de/frankfurtgroup/vat/EndberichtVAT210906.pdf>

The application of current VAT rules to public entities is complex and leads to inconsistent results across the Community. Furthermore, the difference in VAT treatment between public and private entities causes distortions of competition, produces economic inefficiencies and encourages tax avoidance schemes. The Commission will examine ways of simplifying these rules and facilitating their more uniform application throughout the Community in order to secure a level playing field in those sectors of activity where both public and private entities intervene, e.g. the provision of contract research. The closely related problems of exemptions, public authorities and subsidies will be treated as a package and the social and economic impacts of any possible legislative proposal will be assessed before a proposal is presented in 2008.

2.2.7. *Intellectual property rights issues*

As the creators of scientific articles scientists hold the copyright in their work unless it is contractually or statutorily assigned to others. The assignment by authors of their copyright has been standard practice in their relationship with publishers in recent years, and this has important implications for the flow of scientific information. Most publishers allow authors to archive copies of their articles on the web; they allow the author to share the article with colleagues and the article to be used for educational and research purposes.

The assignment of copyright by authors to scientific publishers has facilitated the digitisation and preservation efforts undertaken by publishers and protected authors from plagiarism. However, as suggested by the EC-commissioned “Study on the economic and technical evolution of the scientific publication markets in Europe”, it is necessary to further investigate issues surrounding copyright provisions in order to find “precise legal solutions that would provide legal certainty to authors, but also potentially to other parties, in terms of dissemination of published material”⁴⁵.

Digital rights management technologies enable publishers to control online access in order to prevent unauthorised use. At present, these technologies are not widely used in relation to scientific publications. Possible future widespread use of DRMs may have a negative impact on services associated with new technologies (e.g. data mining of research articles or the linking of research articles to their underlying data). A balance will have to be found between preventing unauthorised use and the aim of using new information technologies to maximise the usefulness of scientific literature and underlying data for the benefit of research.

A further important issue to consider is the use of data in databases. Scientists are both users and producers of databases. Much of the knowledge produced by scientists is collected and distributed through databases, and access to the data held in such databases is critical to the advancement of science.

Copyright law, while protecting original selection, sequence and arrangement, has treated data as such as lacking originality and has therefore classified it as non-protectable. By adopting Directive 96/9/EC on the legal protection of databases, the European Union has, however, introduced specific protection rights — copyright and a sui generis right — for databases against unauthorised reproduction.

⁴⁵ http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf (see p.13 of the study).

The 2005 evaluation report⁴⁶ on Directive 96/9 by the Commission's Internal Market and Services Directorate-General indicates that certain members of the academic and scientific community have expressed concerns that the exceptions to the "sui generis" right were too restrictive with regard to the access to and use of data and information for scientific and educational purposes. The report notes that the protection granted by the Directive comes "precariously close to protecting basic information". At the same time it points to the case law of the European Court of Justice which has restricted the scope of protection for "non-original" databases. This case law limits the risk of undesirable effects on downstream, incremental innovation and access to scientific data.

2.3. Access and dissemination: initiatives and developments

2.3.1. Publishers' new business models

In recent years, publishers have offered libraries so-called "bundle deals" (big deals), in which libraries subscribe to packages of electronic journal titles from publishers at lower costs than the actual combined subscription prices. This model involves institutional and other subscribers paying for access to bundles of online journals through consortia or site licensing arrangements. Consequently, libraries have formed consortia to strengthen their bargaining power vis-à-vis publishers and to share resources.

A similar business model implemented by big publishers in medium and small countries is the so-called "country deal" or "national deal". These countries negotiate the terms with a few of the main scientific publishers to allow researchers, teachers and students of those particular countries general access to a wide selection of electronic journals. Such deals have been concluded with Iceland, Ireland, Finland, Denmark, and more recently Luxembourg⁴⁷.

These types of deal are partly responsible for lowering the average cost per title of current serial subscriptions by 23 percent over the five-year period to 2003-2004⁴⁸. Data from the Association of Research Libraries in the US shows similar trends, with serial unit costs falling since 2000 in tandem with an increase over the same period in the total number of serials purchased by member libraries⁴⁹.

In addition to these generally positive effects, bundle deals and national deals can raise new challenges. The EC-commissioned "Study on the economic and technical evolution of the scientific publication markets in Europe" indicates that bundle deals may cause lock-in effects and create an entry barrier to new entrants since they are usually based on relatively multi-annual contracts. They may thus impede competition and the development of new innovative and more efficient ways of disseminating information. Similarly, bundle deals tend to utilise most of the libraries budgets, thus limiting the purchase of monographs and journals from smaller publishers.

⁴⁶ http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf

⁴⁷ http://www.gouvernement.lu/salle_presse/communiqués/2006/11/14bnl/index.html.

⁴⁸ Creaser, C., Maynard, S. and White, S. (2005). LISU Annual Library Statistics 2005

⁴⁹ <http://www.arl.org/stats/arlstat/05pub/05intro.html>

2.3.2. *The Open Access movement and resulting trends*

An important recent trend has been the development of the Open Access movement, based on the viewpoint that access to publications and data can be improved in the internet age. Material that is “Open Access” is immediately accessible free of charge on the internet and may include peer-reviewed journal articles, conference papers, theses, working papers, etc.

Over the last few years, many initiatives have been driven by and associated with the Open Access movement. Individuals and organisations have sought alternative ways to the traditional subscription journal model for disseminating research results. The history of this movement is described in Peter Suber’s very complete “Timeline of the Open Access Movement”⁵⁰. Key milestones in the Open Access movement include the following:

- Public Library of Science (PloS) Open Letter (September 2001)⁵¹. The members of this non-profit association of scientists and physicians committed to publishing in, editing, reviewing for and personally subscribing to only those journals that allow authors to deposit their published articles in publicly accessible public resources within six months of publication. More than 30 000 signatories have endorsed this letter.
- Budapest Open Access initiative statement (February 2002)⁵². This initiative arises from a meeting convened in Budapest by the Open Society Institute (OSI). The agreed statement introduced the term open access and identified two complementary strategies to achieve open access: self-archiving and open access journals, both facilitated by internet technology and researchers’ willingness to share their findings. More than 300 organisations and 4 000 individuals have signed the initiative.
- Bethesda Statement on Open Access Publishing (June 2003)⁵³. This statement of principles was endorsed by stakeholders working in the biomedicine research community at a conference hosted by the Howard Hughes Medical Institute. It identified the necessary steps that each stakeholder, within this particular research community, should undertake to promote open access to the primary scientific literature.
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (October 2003)⁵⁴. This Declaration was signed at a conference hosted by the Max Planck Society, and thereafter endorsed by 196 international research organisations and remains open for signature. According to the declaration,

⁵⁰ <http://www.earlham.edu/~peters/fos/timeline.htm>.

⁵¹ <http://www.plos.org/support/openletter.shtml>.

⁵² <http://www.soros.org/openaccess/read.shtml>.

⁵³ <http://www.biomedcentral.com/openaccess/bethesda/>.

⁵⁴ <http://OpenAccess.mpg.de/openaccess-berlin/berlindeclaration.html>.

“author(s) of Open Access contributions must grant to all users a free, irrevocable, worldwide right of access to, and a license to copy, use, distribute, transmit and display the work publicly ... in any digital medium”. In addition, “a complete version of the work and all supplemental materials, ... [should be] deposited ... in at least one online repository”⁵⁵.

The Declaration also stressed the importance of finding solutions to legal and financial problems in order to help the transition to open access.

The Open Access movement has led to experiments with so-called open access journals. Open access journals are peer-reviewed journals whose articles may be accessed online without charge for the reader. Publishers of open access journals charge authors for publication costs, thus reversing the traditional model in which a library pays for access to the contents of a journal through a subscription. This charge is usually paid by a research grant or through institutional funds. This arrangement is problematic if the author is not affiliated with any institution and would therefore have to pay the relatively high fee for publishing "open access" out-of-pocket. Open access publishers supplement their funding with other sources of income such as advertising and donations.

The number of open access journals has grown steadily over the past years. At present, the Directory of Open Access Journals records over 2 400⁵⁶ full-text, quality-controlled scientific and scholarly journals (accounting for about 10% of total journal output). In 2004, it was estimated that 55% of open access journals relied fully on public funding, 28% on print subscription revenues and 17% on the author-pays model⁵⁷. The subjects covered are mainly health sciences, social sciences, and biology and life sciences⁵⁸.

BioMed Central is a well-known open access publisher with over 100 journals in its portfolio, including many with high impact factors. Other examples are the journals from the Public Library of Science, such as PLoS Medicine, PLoS Biology. In cases of authors' financial need, BioMed Central, PLoS and other open access journal publishers may waive the publication fee. Fees levied by open access journals vary but, as a guideline, BioMed Central charges €1 000⁵⁹ per article for most of its journals, and PLoS charges €1 200.

Another form of open access is found in “hybrid” journals: these are subscription-based publications that will make specific articles generally accessible online without charge if authors opt to pay for this service. An example of a hybrid journal is the Springer Open Choice programme, and Blackwell Online Open service.

The following are some examples of the results resulting from open access experimenting:

⁵⁵ <http://Open.Access.mpg.de/openaccess-berlin/berlindeclaration.html>.

⁵⁶ <http://www.dOpen.Accessj.org/>.

⁵⁷ Regazzi, J. 2004, The Shifting Sands of Open Access Publishing: A Publisher's View, *Serials Review*.

⁵⁸ EPIC, 2004. URL: http://www.jisc.ac.uk/uplOpen.Accessded_documents/ACF1E88.pdf.

⁵⁹ <http://www.biomedcentral.com/info/authors/apccomparison/>.

- Nucleic Acids Research (NAR), one of OUP’s flagship journals, was converted to full open access from January 2005. Previously it had been a delayed open access journal, with content freely available after six months. OUP reported in June 2006 that NAR’s income per article had dropped from USD 4 647 in 2004 to USD 3 622 in 2005. Income from subscriptions (which include a print copy) has declined steadily.
- The M. Ware Consulting study on scientific publishing⁶⁰ mentions an article published in Nature⁶¹ in June 2006 which stated that “the early break-even hoped for by PLoS is some way off”. The article mentioned that PLoS lost USD 1 million in 2005, and its author fees and advertising revenues covered only 35 percent of total costs. In response to the Nature article, PLoS pointed out that its journals are at an early stage of their lives and also said that it envisaged relying on grant support for the foreseeable future, despite the increase in author charges.
- Hindawi Publishing is a commercial STM publisher based in Egypt, whose open access journal collection has increased recently to 52 journals, including titles in a wide range of subjects from engineering and mathematics, to biomedicine, chemistry, and materials science, and has reported that its OA publishing is profitable⁶², despite charging fairly low author pays charges (typically only €350-400 per article.)

The level of uptake of open access in hybrid journals is relatively low. The Oxford University Press⁶³(OUP) reported some data from the first full year of its Oxford Open scheme. A total of 360 open access papers were published, or 7.9% of the total of 4 575. The data show greatest uptake in the life sciences, followed by medicine, with limited uptake in other disciplines.

Experiments with open access options are very recent and it is therefore too early to tell whether open access is a desirable and viable publishing business model⁶⁴. Author awareness and knowledge of open access are still quite low, but growing. Most scientific journal publishers are experimenting with open access. A debate is currently taking place on whether open access journals are more likely to be cited than subscription-based journals and, as a result, on whether impact factors vary between traditional and open access journals.

Available data on open access must be interpreted with caution as they apply to a relatively small number of journals and because open access journals are still very young. This is an area in which research has recently been carried out, but most of it has been on specific subject areas or titles, making it difficult to generalise, as stated in the “UK scholarly journals: 2006 baseline report”⁶⁵. This report pointed out that “Any study of variance in article impact

⁶⁰ Scientific publishing in transition: an overview of current developments. Mark Ware Consulting.

⁶¹ Butler, D. Open-access journal hits rocky times, in Nature, 20 June 2006.

⁶² http://blogs.nature.com/news/blog/2006/06/openaccess_journal_hits_rocky.html#comment-17436

⁶³ <http://www.oxfordjournals.org/oxfordopen/>.

⁶⁴ http://www.alp.org/publications/FAOpen_AccesscompleteREV.pdf.

⁶⁵ UK scholarly journals: 2006 baseline report. An evidence-based analysis of data concerning scholarly journal publishing.

in Open Access journals versus subscription journals faces a key methodological challenge in that a given article cannot be Open Access and non-Open Access at the same time and, therefore, an exact like-for-like comparison of research impact over the same time period is not possible”.

The report also states that “there is some consistency in results that show more citations for articles self archived in repositories as distinct from the same or similar articles available in a subscription journal. Overall, deposit of articles in open access repositories seems to be associated with both a larger number of citations, and earlier citations for the items deposited”.

The report concludes that “the reasons for this, however, have not been clearly established — there are many factors that influence citation rates, including the reputation of the author, the subject-matter of the article, the self-citation rate, and, of course, how important or influential the repository is in its own right”⁶⁶.

2.3.3. *Open repositories*

A development running parallel to the development of open access journals is self-archiving, through the deposit of articles in open access archives or repositories⁶⁷. Repositories can be either institutional or subject-based. They are collections of research articles and the so-called “grey literature” (conference papers, theses and working papers) that have been placed there by their authors. In the case of journal articles this may be done either before (pre-prints) or after peer review (post-prints).

arXiv. Hosted by Cornell University, arXiv has become the most prominent repository used in the physics community, holding more than 300 000 pre/post-prints for the physics community to read and review. Authors in the field of physics continue to value the quality control function that journals ensure but also the rapid and wide dissemination that arXiv provides.

An analysis of the situation of institutional repositories in a set of EU Member States undertaken by Westrienen⁶⁸ in 2005 shows a wide variety of results depending on the country. For example, it concludes that on average 35% of universities have repositories. The estimate shows a spread from around 5% of universities in Finland, where repositories are just getting started, to essentially 100% deployment in countries like Germany and the Netherlands, where repositories have already achieved importance as a common infrastructure across the relevant national higher education sector.

When looking into the details of the size of repositories, the study witnessed wide differences both between and within countries. The average number of materials per repository typically amounts to a few hundred (± 300), with the exception of the Netherlands, with an average of 12 500 records per institutional repository.

⁶⁶ UK scholarly journals: 2006 baseline report. An evidence-based analysis of data concerning scholarly journal publishing.

⁶⁷ http://www.jisc.ac.uk/publications/publications/pub_openaccess.aspx.

⁶⁸ <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>.

As regards the type of material hosted by repositories, what comes across clearly from the available data is a focus on textual material. However, within this type of material we see strong differences according to country: e.g. in Norway 90% of the current records are for books and theses, while in France it is estimated that 80% of the current records are for articles. The findings suggest that repositories currently hold traditional (print-oriented) research literature and grey literature: non-published journal articles, books, theses and dissertations, and research reports.

At present, the depositing of materials is usually completed by “resource managers” (e.g. librarians). Research⁶⁹ suggests that greater persuasion efforts would be needed to convince the scientific community of the value of archiving and to encourage authors to deposit materials. Whilst some thematic communities are self-motivated, the more common picture is that long-term deposit of research results has a relatively low priority for scientists. Research indicates that voluntary deposit schemes have had limited uptake, whereas with mandatory policies deposit rates rise towards 90%⁷⁰.

In the UK and the Netherlands, there are large national programmes (JISC Digital Repositories programme & SURF Programme Digital Academic Repositories (DARE)) that are advancing the deployment of repositories as well as standards and best practices surrounding their implementation⁷¹. A special situation exists in Germany (DINI-German Initiative of Networked Information)⁷², where a national body was set up to certify (DINI Certificate) repositories according to certain standards

DARE. In the Netherlands the Digital Academic Repositories programme (DARE) is implementing a national scientific information infrastructure. The objective of DARE was to provide networked free access to academic research output from all universities. Partners in this programme include the Dutch National Library (KB), the Dutch Research Council (KNAW), the Dutch Organisation for Scientific Research (NWO) and the SURF foundation. The basic DARE infrastructure currently holds almost 100 000 scientific/technical reports and research articles, and will in a later phase include experimental or observational data, rich media and other digital objects like video and audio. All articles are automatically stored and preserved for permanent access in the e-Depot system of the National Library (KB). At the same time, a National Data Archiving and Networked Service (DANS) is being set up to store, archive, preserve, curate, certify and provide permanent access to distributed trusted scientific repositories of data, the primary information source on which research draws.

There are difficulties in establishing, filling and maintaining repositories. These include lack of institutional support, lack of awareness and information on intellectual property issues (getting copyright permissions to deposit, concerns about who will use material that has been deposited, how it will be used, and whether it will be appropriately attributed), and concerns about long term maintenance costs and value for money.

⁶⁹ Swan, Alma and Brown, Sheridan (2005) Open Access self-archiving: An author study, pp. 1-104.

⁷⁰ Sale, A (2006b). The impact of mandatory policies on ETD acquisition. D-Lib Magazine 12 (4). <http://www.dlib.org/dlib/april06/sale/04sale.html> (April 2006).

⁷¹ <http://www.surf.nl/eng/download/country-update2005.pdf>

⁷² <http://www.dini.de>.

Publishers feel concerned by the trend towards “self-archiving”. From their perspective, the availability of pre-prints (articles before peer-review) and/or post-prints (articles after peer-review) in a repository could lead librarians to cancel journal subscriptions. With a view to protecting subscriptions, publishers are therefore applying so-called “embargo” periods after publication during which they do not allow self-archiving. In this regard, the question of whether embargo periods should exist and how long they should be are key questions. The embargo periods implemented by publishers vary by discipline and journal and are shortest for science and technical journals (6-12 months) and longer for the social sciences and humanities (around 18 months).

At the same time, in order to ensure wide dissemination and a good return on investment, research funding bodies are discussing embargo periods for research results that they have funded, i.e. maximum periods of time after publication after which research grant recipients must provide the publications resulting from the funded research for free and public use. Some bodies, such as the UK Wellcome Trust, already mandate deposit of research results on a repository after a certain period of time (see further details below). While publishers argue that too limited embargoes could put journal subscriptions at risk, proponents of open access put forward that self-archiving has not led to journal cancellations by librarians.

Institute of Physics (IOP). The Institute of Physics is a renowned scientific publisher in the field of physics. IOP has managed to co-exist with arXiv with hardly any effect on its readership base. However, IOP has indicated that in the case of journals specialised in areas in which papers appear as pre-prints in arXiv (e.g. high-energy physics journals), the number of article downloads is less than would expect in comparison with other journals. IOP is concerned that this might eventually reduce subscriptions and threaten the viability of journals.

A database of publisher policies is maintained by the SHERPA/RoMEO⁷³ project, with the following results:

- 40% allow archiving of both pre- and post-print articles;
- 26% allow archiving of post-print articles;
- 9% allow archiving of pre-print articles;
- 25% do not formally support archiving.

Some publishers also allow authors to archive the publisher’s final version of an article (post-print version as published), although this is less common. Some also require full bibliographic details plus a link from the pre-print or post-print version to the published version of the article.

⁷³ <http://www.sherpa.ac.uk/romeo.php>

2.3.4. *Activities of legislatures, funding bodies and international organisations*

In the UK, the 2004 House of Commons Science and Technology Committee Report⁷⁴ on Scientific Publications was the first major attempt to analyse the dynamics of the scientific publishing system, setting the scene for many of the developments that have taken place since. The report's first recommendation was that the UK Research Councils make it a condition of grant that researchers deposit a copy of their articles in a local institutional repository within one month of publication. The second recommendation was that the funding bodies should make funds available for authors to pay open access publication charges. These recommendations have been partially taken up by the UK Research Councils and the Wellcome Trust, with longer embargo periods (see below).

In the United States, a US Senate bill⁷⁵ was introduced on 2 May 2006 by Senators John Cornyn and Joseph Lieberman to facilitate access to publicly funded research output. If passed, the Federal Research Access Act (FRAA) would establish that every federal agency with an annual extramural research budget of USD 100 million or more must implement a public access policy that is consistent with and advances the federal purpose of the respective agency. According to the bill, each agency must:

- i. Require each researcher – funded totally or partially by the agency – to submit an electronic copy of the final manuscript that has been accepted for publication in a peer-reviewed journal;
- ii. Ensure that the manuscript is preserved in a stable digital repository maintained by that agency or in another suitable repository that permits free public access, interoperability, and long-term preservation;
- iii. Require that free, online access to each taxpayer-funded manuscript be available as soon as possible, and no later than six months after the article has been published in a peer-reviewed journal.

The Wellcome Trust⁷⁶ is the world's largest medical research charity funding research into human and animal health. It has recently endorsed an assertive open access policy requiring electronic copies of any research papers that have been accepted for publication in a peer-reviewed journal and that have been supported by Wellcome Trust funding to be deposited in an open repository (PubMed Central) and to be made freely available within six months of the journal publishers' official date of final publication. The Trust is contributing financially to the establishment and operation of the UK version of PubMed Central. In addition, the Wellcome Trust meets the author's fees should a researcher choose to publish in an open access journal, making a research paper immediately accessible.

The UK Research Councils⁷⁷ are the umbrella organisations for the UK's public research

⁷⁴ <http://www.publications.parliament.uk/pa/cm200304/cmsselect/cmsstech/399/39902.htm>.

⁷⁵ <http://cornyn.senate.gov/index.asp?f=record&lid=1&rid=237171>.

⁷⁶ http://www.wellcome.ac.uk/doc_WTD002766.html.

⁷⁷ <http://www.rcuk.ac.uk/access/default.htm>.

institutions. They have recognised that access to research outputs is a highly complex issue on which there are a broad range of opinions. In developing their positions, the Research Councils have consulted with all concerned stakeholders. Most of the Research Councils have recently produced specific guidance to the research communities they fund on access to outputs in each field of research. At present five of the eight Research Councils have endorsed open access policies, with embargo periods ranging from six months (Medical Research Council) to “at the earliest opportunity”. The Research Councils are prepared to fund authors’ fees as part of the research grants.

The US National Institutes of Health (NIH)⁷⁸ are the largest funder of medical research in the world, and the largest funder of non-classified research in the US federal government. Their budget for 2005 was USD 28 billion. The general policy of the NIH is to provide free online access to full-text, peer-reviewed journal articles arising from their funded research. They ask every scientist who receives an NIH research grant, and who publishes the results in a peer-reviewed journal, to deposit a digital copy of the article in PubMed Central, the online digital library maintained by the NIH. PubMed Central then provides free online access some time after the article is published in a journal; the length of the delay is to be determined by the author but may not be longer than 12 months. Under this voluntary scheme, only 5% of eligible research papers were deposited in PubMed Central. The NIH is currently in the process of reviewing this voluntary scheme and is considering a mandatory deposit policy.

Established in 2006, the European Research Council (ERC) is Europe's funding body for frontier research under the European Commission's Seventh Framework Programme for Research and Development. In December 2006, the European Research Council's Scientific Council published a Statement on Open Access. This statement "stresses the fundamental importance of peer-reviewed journals in ensuring the certification and dissemination of high-quality scientific research", but notes that the "high prices of some journals" [...] raise significant worries concerning the ability of the system to deliver wide access and therefore efficient dissemination of research results, with the resulting risk of stifling further scientific progress. The ERC's Scientific Council therefore displays "the firm intention [...] to issue specific guidelines for the mandatory deposit in open access repositories of research results – that is, publications, data and primary materials – obtained thanks to ERC grants, as soon as pertinent repositories become operational."

A further impulse at the European Level was given by EURAB, the European Research Advisory Board. In a [report on "Scientific Publication: policy on open access" of December 2006, EURAB recommends that the Commission "consider mandating all researchers funded under FP7 to lodge their publications resulting from EC-funded research in an open access repository as soon as possible after publication, to be made openly accessible within 6 months at the latest."](#)

The OECD Working Party on the Information Economy prepared a comprehensive report on scientific publishing in 2004⁷⁹, but within the OECD, lately discussions have focused on the specific issue of access to research data. The OECD Member States have endorsed the view that the increasing online availability of research data is changing research practices and the

⁷⁸ <http://www.nih.gov>.

⁷⁹ <http://www.oecd.org/dataoecd/42/12/35393145.pdf>.

growing trend of making primary data sources directly accessible is changing the business models of the scientific publishing industry. This principle was underlined in the 2004 OECD Science Ministerial Declaration on Access to Research Data from Public Funding⁸⁰, which recognised that open access to, and unrestricted use of, data promotes scientific progress and facilitates the training of researchers. The OECD Committee for Scientific and Technological Policy (CSTP) at Ministerial Level recently approved a draft Recommendation on Access to Research Data and decided to forward it to the OECD Council for endorsement. If approved, the CSTP will have to monitor the implementation of the Recommendation.

The open access approach for science

Two longstanding successful open access experiments are the GenBank and the Protein Data Bank. The success of the genome project, which is generally considered to be one of the great scientific achievements of recent times, is due to the fact that the world's entire library of published DNA sequences has been an open-access public resource for the past 20 years.

CERN is the birthplace of the most successful internet application, the World Wide Web. The new LHC (Large Hadron Collider) will be delivering the first experimental results in 2007 and CERN is working to ensure the widest possible dissemination of research results through an experimental publishing initiative that aims to move a critical mass of top-quality journals towards an open access model. A task force was set up for this purpose. After extensive consultation with stakeholders, in June 2006 it published a report⁸¹ that estimates that the seven top-level journals, accounting for 50% of research papers in the field of particle physics, would require an effort of €5 to 6 million over a two-year period to publish under an "author pays" open access business model (estimation based on data for the 2003-2005 period). These costs would correspond to average costs for submission of €400-1 800 or by published article in the range of €755-2 500. CERN is currently working towards the implementation phase of this initiative in the form of a Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP³).

3. PRESERVATION ISSUES

3.1. The importance of digital preservation

The organisational, legal, technical and financial challenges for digital preservation briefly addressed in this section, including the Member States' legal framework for legal deposit, have been described in detail in two recent Commission staff working papers⁸² related to the Communication "i2010: digital libraries" and the Commission Recommendation on the digitisation and online accessibility of cultural material and digital preservation.

⁸⁰ http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html.

⁸¹ http://library.cern.ch/OATaskForce_public.pdf.

⁸² http://europa.eu.int/information_society/activities/digital_libraries/doc/communication/annex1_en.pdf
and
http://europa.eu.int/information_society/activities/digital_libraries/doc/recommendation/impact_assessment/en.pdf.

Long-term preservation of “born-digital” journals is a major concern for stakeholders; unlike print journals, they cannot assume that the existence of copies in numerous different libraries will be sufficient to ensure archival preservation. A survey conducted in 2003 among the research community in the US found that 78% of respondents considered electronic scholarly journals to be “invaluable research tools”⁸³. This dependence on the convenience and enhanced accessibility of electronic scholarly resources has raised concern regarding their long-term preservation and future accessibility. The same survey revealed that 83% of respondents rated as “very important” the preservation of electronic scholarly resources for future use. This concern is not without reason: recent studies have shown that, three years after publication, half of the internet references quoted in three prestigious journals were no longer retrievable⁸⁴. Access to scientific output is increasingly electronic — and fragile — and its future availability is a growing concern. Digital files may become inaccessible in future as software and hardware changes. It is therefore essential that the content be kept in a “neutral” format from which it can be migrated as necessary.

Another important aspect of digital preservation of scientific information concerns the need to collect, safely store and make available results of the observations of unique physical phenomena or experiments that by their very nature are not easily reproducible. This goes well beyond a simple archiving service — the data is only useful if accompanied by information about the conditions under which the observation was made or the experiments carried out.

The climate change debate is now splashed across the front pages of newspapers, capturing the public’s attention. However, the analysis of possible remedies must rely on concrete evidence linking foreseeable causes to the observed changes. This requires intensive analysis of meteorological data collected over the years, comparing them with the predictions of various climate models, including observations made by satellites. The World Climate Research Programme makes available the Radiosonde Atmospheric Temperature data, collected since 1958, in the study of longer-term climate variability and change⁸⁵. The European Space Agency and the Royal Netherlands Meteorological Institute have also developed an animated model of the evolution of the ozone layer showing the development of this cyclical phenomenon⁸⁶ over the years. These are two concrete examples of research which relies on preservation of digital data.

While scientific fields more inclined to rely on quantitative methods, such as physics, chemistry, engineering, biology, astronomy or the medical sciences, constitute areas where concern for the long-term preservation of scientific data appears to be of particular importance, the social sciences and humanities are also concerned. For example, the increasingly multicultural nature of European societies, brought about by personal mobility and the scale of migratory flows, raises worry among public opinion about the possible

⁸³ <http://www.portico.org/about/why.html>.

⁸⁴ Bugeja and Dimitrova, “The Half-Life Phenomenon: Eroding Citations in Journals”, *The Series Librarian* 49, No 3 (2005). Concordant information also available from <http://www.ariadne.ac.uk/issue47/fenton/#1>.

⁸⁵ See <http://www.ncdc.noaa.gov/oa/cab/ratpac/index.php>.

⁸⁶ See <http://www.temis.nl/> and http://www.esa.int/esaCP/SEMQBOKKKSE_index_0.html.

relationship between this trend and the perception of increased insecurity. The UK Economic and Social Research Council launched a group of projects aimed at assessing to what extent individual and neighbourhood effects are able to account for the geographical variation in crime patterns. The analysis provides a meaningful contribution to the theoretical debate on the relative influence of individual, family, school and social ties and neighbourhood factors on crime patterns. The resulting model is also seen as a useful tool for urban planners, local authorities, youth services and the police. Once again, models of this type rely on the comparative analysis of data patterns over time, therefore requiring that data collected remains accessible and usable in the long term. This example also highlights the importance of preserving not only the data but also its evolving meaning (e.g. the classification of criminal offences).

The long-term preservation and availability of collections of scientific and scholarly publications and scientific data already constitutes a very important source of indirect economic value. The results of the survey of digital preservation needs conducted by the UK-based Digital Preservation Coalition in 2005 illustrate some of the ways in which this economic value may be apparent. Around 64% of the respondents consider that preservation of digital data is an important mechanism to protect intellectual property and 22% to support patent applications. This interpretation is consistent with other market studies. In the survey conducted in 2006 by the industry group AIIM⁸⁷, UK respondents also place “library and knowledge management” among the top five application domains of digital preservation.

In addition, market research has demonstrated that the provision of systems and services for digital preservation is also becoming an increasingly important economic activity in the ICT sector. The AIIM industry survey mentioned above, covering more than 1 200 organisations in industrialised countries (US, Canada, Germany, UK, Benelux, Australia and Brazil), found that 70% of respondents considered the management (and preservation) of electronic information to be of critical importance. For 61% of the respondents, improved efficiency and productivity appears to be the main driver for investment in digital preservation technologies. About 20% of the respondents are planning to invest more than €1 million in electronic content management and preservation technologies, though this percentage increased to 47% and 53% for the government and financial sectors respectively. The evolution of these figures over the last five years indicates that this is a persistent pattern, not simply the result of a sudden or occasional interest in the topic. The dynamism of the document electronic management service sector can be seen from the overall growth in revenues and profits (only 6.2% of respondents indicated a decrease in revenues while 46% reported net profits growth exceeding 10%). This also seems to be having a positive impact on employment, with 38% of companies anticipating an increase in staff numbers of more than 10% in the following year.

3.2. Preservation: initiatives and developments

An infrastructure of digital repositories, hosting this increasing volume of data, has been suggested as one of the answers when it comes to the organisation, governance and management of scientific information. Such an infrastructure would allow remote access and provide adequate levels of redundancy, which are essential for long-term preservation. As far as the organisation of the pan-European infrastructure is concerned, there has been debate

⁸⁷ Association for Information and Imaging Management - www.aiim.org.

with the participation of the scientific community and within groups like the e-IRG⁸⁸ (e-Infrastructure Reflection Group). The three-layer model (local, national, European) was widely supported. The structure must be inclusive and open to later additions, e.g. new countries. It was felt that this structure was the most robust for a pan-European infrastructure. Thematic repositories are also important, but it was felt that these, and wider international collaboration, could be successfully accommodated.

The vision for this new infrastructure of repositories is becoming a reality in certain well defined disciplines and from the results of particular projects. Alongside this, and because of it, there are wider and more geographically diverse collaborative networks of scientists working on the same problem. The result is a new-found ability to access, move, manipulate, analyse and visualise the results of such work.

The Commission is supporting this new way of conducting science and plans to dedicate substantial resources through the Capacities programme in FP7 (2007-2012), supporting the development of this e-infrastructure at all levels: connectivity (GÉANT), shared computer resources (GRID) and digital repositories.

National libraries such as the British Library and the Koninklijke Bibliotheek are investing large sums in systems for hosting and preservation of digital content; these are government-funded, in the former case in the context of developing legislation for legal deposit of non-print materials.

Among other initiatives dealing with long-term preservation of scientific information, the work done in the US by JSTOR and Portico deserves special mention.

JSTOR is a not-for-profit organisation with a dual mission to create and maintain a trusted archive of important scholarly journals, and to provide access to these journals as widely as possible.

A related initiative launched by JSTOR but more specifically focused on digital preservation issues is Portico. Its primary mission is to preserve scholarly literature published in electronic format, guaranteeing that these materials remain accessible in the future. The work being developed by Portico addresses the issues of economic sustainability, technological infrastructure and the mobilisation of stakeholders (notably libraries and publishers). As part of this process special attention is devoted to analysing the changing habits of the target users and anticipating future uses of scientific information. This work has resulted in valuable guidance for libraries and publishers on how to address the critical issue of preserving their online collections and guaranteeing optimal usability (formats, searchability, and persistence of links, versioning and longer-term accessibility). Today only a relatively small number of articles (about 200 000) have been preserved by Portico, which is an indication of the dimension of the challenge presented by digital preservation. Portico relies on contributions from publishers and libraries as the main source of revenue supporting their activities. Government agencies (Library of Congress and NDIIPP programme) and charities (Andrew W. Mellon Foundation) constitute the other sources of funding.

⁸⁸ <http://www.e-irg.org/>.

4. OVERVIEW OF RELEVANT COMMUNITY INITIATIVES

4.1. i2010 digital libraries

On 1 June 2005 the Commission presented the i2010 initiative, which seeks to optimise the benefits of the new information technologies for economic growth, job creation and the quality of life of European citizens. The Commission has made digital libraries a key aspect of i2010. In its Communication “i2010: digital libraries” of 30 September 2005, it set out its strategy for digitisation, online accessibility and digital preservation of Europe’s collective memory. The digital libraries initiative follows up on a letter of 28 April 2005 by the Heads of State and Government of France, Germany, Hungary, Italy, Poland, and Spain asking the Commission to take necessary steps to improve access to Europe's cultural and scientific heritage. The involvement of the Commission in the digital libraries field is not new. In recent years, the Commission has contributed to the coordination of digitisation actions in the Member States and has co-funded relevant R&D actions (see below).

The Communication was discussed by the Culture Council of 14 November 2005. At their meeting, Culture Ministers gave strong backing to the EU digital libraries initiative.

The Commission adopted on 24 August 2006 a Recommendation⁸⁹ on the digitisation and online accessibility of cultural material and digital preservation, calling on the Member States to take concrete measures. At its meeting on 13 November 2006 the Education, Youth and Culture Council welcomed the Recommendation and unanimously adopted very positive conclusions demonstrating the Member States’ commitment to improving the digitisation and online accessibility of cultural material and its digital preservation.

4.2. EU research policy

The Communication “Scientific information: access, dissemination and preservation” is part of the Community policy on research, which looks to maximise the socio-economic benefits of research and development efforts for the public good. It represents an initial step within a wider policy process addressing the connections and interactions between the scientific publication system and research excellence.

4.3. Investigation, public consultation and interaction with stakeholders

As mentioned above, the EC commissioned a “Study on the economic and technical evolution of the scientific publication markets in Europe” and held a public consultation on the Study from March to June 2006⁹⁰. An online consultation on the digital libraries initiative, held between September 2005 and January 2006, also yielded a number of comments in relation to scientific and scholarly information, although the consultation did not specifically address this area⁹¹.

⁸⁹ http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l_236/l_23620060831en00280030.pdf.

⁹⁰ http://ec.europa.eu/research/science-society/page_en.cfm?id=3185.

⁹¹ An overview of the replies to the consultation can be found at http://europa.eu.int/information_society/activities/digital_libraries/doc/results_online_consultation/en.pdf. Individual responses can be read at

The Commission also works with advisory groups, for example within a subgroup on scientific information of the High Level Expert Group on Digital Libraries and within the European Research Advisory Board (EURAB).

Stakeholder workshops on scientific information questions have been organised and bilateral contacts have been developed with publishers, the research community, libraries and funding agencies.

Examples of relevant research projects in FP5 and FP6

The Commission started exploring issues of access, dissemination and preservation of scientific information, through project funding. Research on digital preservation was initiated in FP5-IST but at that time work was still fairly embryonic.

Main contributions have come from the projects DELOS and ERPANET. DELOS⁹², a thematic network on digital libraries, co-organised a joint workshop with the National Science Foundation in 2003 which gave rise to the report “Invest to Save”. This report identified the main problems faced by libraries in particular and organisations in general with the transition from the analogue to the digital world in terms of long-term archiving and preservation. The associated research agenda identified a number of priorities later reflected in calls for proposals under FP6-IST and in the activities of the NDIIPP (National Digital Information Infrastructure and Preservation Program) in the United States. The ERPANET⁹³ project was an accompanying measure whose main objective was to raise awareness of digital preservation among the organisations concerned: memory institutions (museums, libraries and archives), the ICT and software industry, research institutions, government bodies, entertainment and creative industries, and commercial sectors (including for example pharmaceuticals, petro-chemicals and finance). This included the creation of a knowledge base on state-of-the-art developments in digital preservation and the transfer of that expertise among individuals and institutions. ERPANET organised, between 2003 and 2004, more than 20 workshops bringing together several hundred experts and practitioners in digital preservation across Europe.

The preparatory work done in FP5 was of critical importance in preparing the ground for the subsequent Framework Programme (FP6-IST). The work programme for 2003-2005 of FP6-IST included as one of the research topics “the access to and preservation of cultural and scientific resources”. The proposals submitted by the research community, unanimously considered by the evaluation panel as being of high quality, resulted in the selection of two integrated projects, launched in mid-2006, representing a total investment of more than €30 million.

The project PLANETS⁹⁴ brings together the national libraries and archives of several European countries (UK, NL, DK, CH, DE, AT) and focuses on the preservation of the assets

http://europa.eu.int/information_society/activities/digital_libraries/cultural/actions_on/consultations/online_consultation/index_en.htm.

⁹² <http://www.delos.info/>.

⁹³ <http://www.erpanet.org/>.

⁹⁴ www.planets-project.eu.

held by these institutions. The goal is to develop a coherent methodological approach and a set of technological tools that can be adopted by similar organisations across Europe.

The project CASPAR⁹⁵ involves major research organisations in Europe (CCLRC (UK), CNRS (FR), CNR (IT), European Space Agency) and broadens the scope of work to also cover the preservation of scientific data, audiovisual content and digitised cultural heritage. The research agenda of these projects includes the discussion and development of relevant technical standards, their adoption by standards bodies and their promotion among the industry and users. Both projects adopt the OAIS model as the main architectural reference for their implementation of digital preservation systems. These developments are expected to offer a solution to the most immediate problems faced by organisations having to deal with the long-term preservation of their digital resources.

Despite the high expectations placed in PLANETS and CASPAR, the magnitude of the problem faced requires further research work. In order to prepare more extensively for future work in this area, the Commission is also funding the coordination action DPE (Digital Preservation Europe) whose remit includes updating the research agenda of DELOS, reflecting the new challenges resulting from the evolution of the internet and web-publishing technologies and the wide adoption of ICT by the research community. The work being carried out by DPE is already visible in the preparation of FP7-IST. The work programme 2007-2008 in challenge 4 – “Digital Libraries and Content” establishes digital preservation as a priority and calls for research on “radically new approaches to digital preservation”, capable of addressing critical issues of high volume, dynamic and volatile web content, the evolving meaning and usage context of digital content and the need to safeguard integrity, authenticity and accessibility over time.

Other relevant projects are DILIGENT, DRIVER and EURO-VO-DCA launched in FP6.

The DILIGENT⁹⁶ project aims to build a knowledge layer on top of the existing GEANT and middleware layers. DILIGENT deals with many issues relevant for digital repositories (i.e. data management, common approach to standards, protocols and interfaces, interaction between national and international repositories, handling complex objects). DILIGENT will create a tested infrastructure on grid-enabled technology.

DRIVER⁹⁷ stands for Digital Repository Infrastructure Vision for European Research. It is intended to supplement GEANT2, the successful infrastructure for computing resources, data storage and data transport. DRIVER is to deliver the content resources, i.e. any form of scientific output, including scientific/technical reports, working papers, pre-prints, articles and original research data. The vision, to be accomplished in a second phase, is to establish the successful interoperation of both data network and knowledge repositories as integral parts of the e-infrastructure for research and education in Europe.

⁹⁵ <http://www.casparpreserves.eu/>.

⁹⁶ <http://www.diligentproject.org/>.

⁹⁷ <http://www.diligentproject.org/>.

EURO-VO-DCA⁹⁸ (the European Virtual Observatory Data Centre Alliance) coordinates the national and European agencies' VO initiatives, implements networking of European data centres, disseminates knowledge and good practice about the VObs technical framework, organises feedback from implementation of interoperability standards, prepares the inclusion of theoretical astronomy in the VObs framework, seeks coordination with national and international projects for computational grids, and helps data centres from beyond the partners' countries to participate in the VObs endeavour.

SEADATANET⁹⁹ is a project aiming to enable a pan-European infrastructure for Ocean and Marine Data Management. While access to marine data is very important for marine research and other vital research issues such as climate change prediction, the marine observing system is highly fragmented. SEADATANET therefore seeks to develop a standardized system for managing relevant data sets and to build a network of existing infrastructures (oceanographic data centres, satellite data centres of 35 countries).

⁹⁸ <http://www.euro-vo.org>.

⁹⁹ <http://www.seadatanet.org/>.