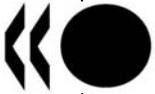


For Official Use

C(2006)184



Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

29-Nov-2006

English - Or. English

COUNCIL

**C(2006)184
For Official Use**

Council

**DRAFT RECOMMENDATION OF THE COUNCIL CONCERNING ACCESS TO RESEARCH DATA
FROM PUBLIC FUNDING**

(Note by the Secretary-General)

JT03218772

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

English - Or. English

Increasing the return on public investments in scientific research

1. The public science systems of OECD Member countries are based on the principle of openness and the free exchange of ideas, information and knowledge. New information and communication technologies (ICTs), now in widespread use throughout all research disciplines, have greatly aided this system of free exchange and have opened up new avenues for collaboration and sharing. The progress of science, however, depends on more than just technologies. Research policies, practices, support systems and cultural values all affect the nature of new discoveries, the rate at which they are made, and the degree to which they are made accessible and used.

2. The power of computers and the Internet has created new fields of application for not only the results of research, but the *sources* of research: the base material of *research data*. Moreover, research data, in digital form, are increasingly being used in research endeavours beyond the original project for which they were gathered, in other research fields and in industry. Administrative data from the institutions of OECD Member countries, such as employment information, are now used extensively in the social sciences, as well as in policy making. Data from public health organisations play a growing role in the advancement of life sciences. Similarly, geo-spatial data collected by many different government organisations are essential for environmental and other types of research. The list goes on.

3. Scientific databases are rapidly becoming a crucial part of the infrastructure of the global science system. The international *Human Genome Project* is but one good example of a large-scale research endeavour in which an openly accessible data repository is being used successfully by many different researchers, all over the world, for different purposes and in different contexts. Many other examples, involving research undertakings both large and small, are readily available.

4. Effective access to research data, in a responsible and efficient manner, is required to take full advantage of the new opportunities and benefits offered by ICTs. Accessibility to research data has become an important condition in:

- the good stewardship of the public investment in factual information;
- the creation of strong value chains of innovation;
- the enhancement of value from international co-operation.

5. more specifically, improved access to, and sharing of, data:

- reinforces open scientific inquiry;
- encourages diversity of analysis and opinion;
- promotes new research;
- makes possible the testing of new or alternative hypotheses and methods of analysis;
- supports studies on data collection methods and measurement;
- facilitates the education of new researchers;

- enables the exploration of topics not envisioned by the initial investigators;
- permits the creation of new data sets when data from multiple sources are combined.

6. Sharing and open access to publicly funded research data not only helps to maximise the research potential of new digital technologies and networks, but provides greater returns from the public investment in research.

7. Throughout OECD Member countries, continuously growing quantities of data are collected by publicly-funded researchers and research institutions. This rapidly expanding body of research data represents both a massive investment of public funds and a potential source of the knowledge needed to address the myriad challenges facing humanity.

8. To promote improved scientific and social return on the public investments in research data, OECD member countries have established a variety of laws, policies and practices concerning access to research data at the national level. In this context, international guidelines would be an important contribution to fostering the global exchange and use of research data.

9. These *Principles and Guidelines* are meant to apply to research data that are gathered using public funds for the purposes of producing publicly-accessible knowledge. The nature of “public funding” of research varies significantly from one country to the next, as do existing data access policies and practices at the national, disciplinary and institutional levels. These differences call for a flexible approach to data access and recognition that one size does not fit all. Moreover, the balance between the costs of improved access to research data and the benefits that result from such access will need to be judged by individual national governments and their research communities.

10. Whatever differences there may be between practices of, and policies on, data sharing, and whatever legitimate restrictions may be put on data access, practically all research could benefit from more systematic sharing. As the authors of the US National Research Council study, *Bits of Power*, pointed out:

The value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research.

11. The specific aims and objectives of these *Principles and Guidelines* are to:

- promote a culture of openness and sharing of research data among the public research communities within Member countries and beyond;
- stimulate the exchange of good practices in data access and sharing;
- raise awareness about the potential costs and benefits of restrictions and limitations on access to and the sharing of research data from public funding;
- highlight the need to consider data access and sharing regulations and practices in the formation of Member countries’ science policies and programmes;
- provide a commonly agreed upon framework of operational principles for the establishment of research data access arrangements in Member countries;
- offer recommendations to Member countries on how to improve the international research data sharing and distribution environment.

12. The *Principles and Guidelines* contained in this document should assist governments, research support and funding organisations, research institutions and researchers themselves in dealing with the barriers and challenges in improving the international sharing of, and access to, research data. These *Principles and Guidelines* should be considered in light of, and applied to, the following major issues inherent in providing data access:

- Technological issues: access to research data, and their optimum exploitation, requires appropriately designed technological infrastructure, broad international agreement on interoperability, and effective data quality controls.
- Institutional and managerial issues: while increased accessibility is important to all science communities, the diversity of the scientific enterprise suggests that a variety of institutional models and tailored data management approaches are most effective in meeting the needs of researchers.
- Financial and budgetary issues: scientific data infrastructure requires continued and dedicated budgetary planning and appropriate financial support. The use of research data will not be maximised if access, management, and preservation costs are an add-on or after-thought in research projects. It is important to note, however, that the cost of storing and managing data has decreased dramatically in recent years, and lack of knowledge about such changes can, in itself, be a barrier to advancement.
- Legal and policy issues: national laws and international agreements, particularly in areas such as intellectual property rights and the protection of privacy, directly affect data access and sharing practices, and must be fully taken into account in the design of data access arrangements.
- Cultural and behavioural issues: appropriate educational and reward structures are a necessary component for promoting data access and sharing practices. These considerations apply to those who fund, produce, manage, and use research data.

13. In working towards better access to research data in the context of these *Principles and Guidelines*, Member countries will need to determine the appropriate balance between the costs of improved access to this data and the benefits that result from such access. The efforts to improve access, of course, need to be carried out within existing financial limitations.

14. In light of the preceding, the Secretary-General invites the Council to adopt the following draft conclusions:

THE COUNCIL:

- a) noted document C(2006)184;
- b) adopted the draft Recommendation of the Council concerning Access to Research Data from Public Funding set out in the Appendix to document C(2006)184, and agreed to its declassification.

APPENDIX

DRAFT RECOMMENDATION OF THE COUNCIL CONCERNING ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

THE COUNCIL,

Having regard to Article 5b) of the Convention on the Organisation for Economic Co-operation and Development of 14 December 1960;

Having regard to Rule 18b) of the OECD Rules of Procedure;

Recognising the wide range of benefits that arise from improving international access to, and use of, publicly funded research data, as expressed in the *Ministerial Declaration on Access to Research Data from Public Funding* of 30 March 2004 [C(2004)31/REV1];

Recognising that international exchange of data, information and knowledge is essential to the advancement of research and innovation in all Member countries;

Recognising the substantial benefits that science, the economy and society at large could gain from the opportunities that expanded use of research data have to offer;

Recognising that improved access to research data will increase the value derived from public investments in data collection, management and preservation;

Recognising that undue restrictions on access to, and use of, research data from public funding diminish the quality and efficiency of research and innovation;

Recognising that enhanced availability of research data from public funding for developing economies will enhance their participation in the global research system, thereby contributing to their social and economic development;

Recognising that rapid changes in information technology will require a careful and flexible approach to developing data access policies and that access arrangements must take into account differences in information infrastructures at the national and international levels and between research disciplines;

Recognising that the access to and use of research data from public funding may be constrained by domestic laws such as those on national security, the protection of privacy and confidentiality and intellectual property rights;

Recognising that delayed access to research data may be legitimate to allow for the right of first-use by researchers and for commercial exploitation;

Recognising that the design of specific policies on access to research data will be shaped by individual Member countries taking into account their own political, administrative and legal context;

Recognising that the balance between the costs of improved access to research data and the benefits that result from such access will need to be judged, taking into account the prevailing financial limitations, by individual Member countries;

Recognising the Member countries' declared commitment to working towards the establishment of access arrangements for research data from public funding, balancing the interests of open access to publicly funded research data with the need to restrict access in specific instances to protect social, scientific, and economic interests;

Having regard to the above-mentioned *Ministerial Declaration* which highlights the need to take the necessary steps to strengthen existing instruments and – where appropriate – to create within the framework of international and national law, new instruments and practices supporting international collaboration in access to research data;

On the proposal of the Committee for Scientific and Technological Policy;

RECOMMENDS that Member countries take into consideration the Principles and Guidelines on Access to Research Data from Public Funding set out in the annex to this Recommendation and which form an integral part thereof and apply them, as appropriate for each Member country, to develop policies and good practices related to the accessibility, use and management of research data;

INSTRUCTS the Committee for Scientific and Technological Policy to review the implementation of this Recommendation as necessary;

INSTRUCTS the Committee for Scientific and Technological Policy to review the Principles and Guidelines on Access to Research Data from Public Funding and when appropriate, to take into account advances in technology and research practices, with the intention of further fostering international co-operation.

ANNEX

PRINCIPLES AND GUIDELINES FOR ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

I - Objectives

These *Principles and Guidelines for Access to Research Data from Public Funding* (hereafter the “*Principles and Guidelines*”) provide broad policy recommendations to the governmental science policy and funding bodies of Member countries on access to research data from public funding. They are intended to promote data access and sharing among researchers, research institutions, and national research agencies, while at the same time, recognising and taking into account, the various national laws, research policies and organisational structures of Member countries.

The ultimate goal of these *Principles and Guidelines* is to improve the efficiency and effectiveness of the global science system. They are not intended to hinder its development with onerous obligations and regulations or impose new costs on national science systems.

II - Scope and definitions

These *Principles and Guidelines* are meant to apply to research data, whether already in existence or yet to be produced, that are supported by public funds for the purposes of developing publicly-accessible scientific research and knowledge. The *Principles and Guidelines* are not intended to apply to research data gathered for the purpose of commercialisation of research outcomes, or to research data that are the property of a private sector entity. Access to such data is subject to a range of considerations that are beyond the scope of this document. Moreover, in some instances, access to or use of data may be restricted to safeguard the privacy of individuals, protect confidentiality, proprietary results or national security.

Research data

In the context of these *Principles and Guidelines*, “research data” are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.

This term does not cover the following: laboratory notebooks, preliminary analyses, and drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (e.g., laboratory samples, strains of bacteria and test animals such as mice). Access to all of these products or outcomes of research is governed by different considerations than those dealt with here.

These *Principles and Guidelines* are principally aimed at research data in digital, computer-readable format. It is indeed in this format that the greatest potential lies for improvements in the efficient distribution of data and their application to research because the marginal costs of transmitting data through the Internet are close to zero. These *Principles and Guidelines* could also apply to analogue

research data in situations where the marginal costs of giving access to such data can be kept reasonably low.

Research data from public funding

Research data from public funding is defined as the research data obtained from research conducted by government agencies or departments, or conducted using public funds provided by any level of government. Given that the nature of “public funding” of research varies significantly from one country to the next, these *Principles and Guidelines* recognise that such differences call for a flexible approach to improved access to research data.

Access arrangements

Access arrangements are defined as the regulatory, policy and procedural framework established by research institutions, research funding agencies and other partners involved, to determine the conditions of access to and use of research data.

III - Principles

A – Openness

Openness means access on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination. Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.

B – Flexibility

Flexibility requires taking into account the rapid and often unpredictable changes in information technologies, the characteristics of each research field and the diversity of research systems, legal systems and cultures of each Member country. Specific national, social, economic and regulatory implications should be considered when organisations develop research data access arrangements, and when governments develop policies to promote data access and review the implementation of these *Principles and Guidelines*.

C – Transparency

Information on research data and data-producing organisations, documentation on the data and specifications of conditions attached to the use of these data should be internationally available in a transparent way, ideally through the Internet. Lack of visibility of existing research data resources and future data collection poses serious obstacles to access.

Factors to consider in ensuring transparency include:

- Information on data-producing organisations and their holdings, documentation on available data sets and conditions of use should be easy to find on the Internet.
- Research organisations and government research agencies should actively disseminate information on research data policies to individual researchers, academic associations, universities and other stakeholders in the publicly-funded research process.
- Whenever relevant, all members of the various research communities should assist in establishing agreements on standards for cataloguing data. The application of existing standards should be

considered, whenever appropriate, in order to avoid placing additional burdens on research resources and work loads of researchers and their institutions.

- Information on data management and access conditions should be communicated among data archives and data producing institutions, so that best practices can be shared.

D – Legal conformity

Data access arrangements should respect the legal rights and legitimate interests of all stakeholders in the public research enterprise.

Access to, and use of, certain research data will necessarily be limited by various types of legal requirements, which may include restrictions for reasons of:

- National security: data pertaining to intelligence, military activities, or political decision making may be classified and therefore subject to restricted access.
- Privacy and confidentiality: data on human subjects and other personal data are subject to restricted access under national laws and policies to protect confidentiality and privacy. However, anonymisation or confidentiality procedures that ensure a satisfactory level of confidentiality should be considered by custodians of such data to preserve as much data utility as possible for researchers.
- Trade secrets and intellectual property rights: data on, or from, businesses or other parties that contain confidential information may not be accessible for research.
- Protection of rare, threatened or endangered species: in certain instances there may be legitimate reasons to restrict access to data on the location of biological resources for the sake of conservation.
- Legal process: data under consideration in legal actions (*sub judice*) may not be accessible.

Subscribing to professional codes of conduct may facilitate meeting legal requirements.

E – Protection of intellectual property

Data access arrangements should consider the applicability of copyright or of other intellectual property laws that may be relevant to publicly-funded research databases. Factors to consider include:

- As public/private partnerships in the funding of research and related data production are increasing, balanced public/private arrangements should facilitate broad access to research data where appropriate. The fact that there is private sector involvement in the data collection should not, in itself, be used as a reason to restrict access to the data. Consideration should be given to measures that promote non-commercial access and use while protecting commercial interests, such as delayed or partial release of such data, or the voluntary adoption of licensing mechanisms. Such measures can allow the primary participants to fully exploit the research data without unnecessarily shutting off access.
- In those jurisdictions in which government research data and information are protected by intellectual property rights, the holders of these rights should nevertheless facilitate access to such data particularly for public research or other public-interest purposes.

F – Formal responsibility

Access arrangements should promote explicit, formal institutional practices, such as the development of rules and regulations, regarding the responsibilities of the various parties involved in data-related activities. These practices should pertain to authorship, producer credits, ownership, dissemination, usage restrictions, financial arrangements, ethical rules, licensing terms, liability, and sustainable archiving.

Access arrangements, whether at the governmental or institutional levels, should be developed in consultation with representatives of all directly affected parties. In collaborative research programmes or projects, and especially in international scientific co-operation or in research projects based on public / private partnerships where there are differences in regulatory frameworks, the parties involved should negotiate research data sharing arrangements as early as possible in the life of the research project, ideally at the initial proposal stage. This will help ensure that adequate and timely consideration will be given to issues such as the allocation of resources for sharing and sustainable preservation of research data, differences in national intellectual property laws, limitations due to national security, and the protection of privacy and confidentiality.

Access arrangements also should be responsive to factors such as the characteristics of the data, their potential value for research purposes, the level of data processing (raw versus partially processed versus final), whether they are homogeneous data from a facility instrument or sensor versus heterogeneous field data collected by single researchers, data on human subjects or physical parameters, and whether the data are generated directly by a government entity or as a result of government funding. These variations in the origin or type of data should be taken into consideration when establishing data access arrangements.

Further, consideration should be given to the following:

- Many of the problems related to access, dissemination and sharing of data result from the lack of explicit institutional agreements on the terms of access and use. With data management becoming ever more complex in certain areas of research, traditional informal arrangements between researchers may no longer be adequate and may need to be complemented by formally agreed practices and procedures.
- Responsibility for the various aspects of data access and management should be established in relevant documents, such as descriptions of the formal tasks of institutions, grant applications, research contracts, publication agreements, and licenses.
- Long-term sustainability of the infrastructure required for data access is particularly important. Research institutions and government organisations should take formal responsibility for ensuring that research data are effectively preserved, managed and made accessible in order that they can be put to efficient and appropriate use over the long term.

G – Professionalism

Institutional arrangements for the management of research data should be based on the relevant professional standards and values embodied in the codes of conduct of the scientific communities involved.

Factors to consider include:

- The use of codes of conduct for professional scientists and their communities could help simplify and reduce the regulatory burden placed on access.

- Mutual trust between researchers, and trust between researchers, their institutions and other organisations plays an important role in the establishment and maintenance of such codes of conduct.
- In current research practice, the initial data-producing researcher or institution is sometimes rewarded with temporary exclusive use of the data. The rules for such incentive arrangements should be developed and explicitly stated by the funding sources in co-operation with the affected research communities.

In certain areas of science, a lack of planning for and execution of the proper documentation and archiving of data sets is one of the key impediments to realising maximum value from the investment in research data. Project and program planning activities, at all levels, should expressly acknowledge data issues at the earliest stages to take into consideration funding and technical assistance for the essential organisation and curation of those data sets. Attention should be paid to incentives and the development of professional expertise in all areas of research data management.

H – Interoperability

Technological and semantic interoperability is a key consideration in enabling and promoting international and interdisciplinary access to and use of research data. Access arrangements, should pay due attention to the relevant international data documentation standards. Member countries and research institutions should co-operate with international organisations charged with developing new standards.

Although science is becoming a highly globalised endeavour, incompatibility of technical and procedural standards can be the most serious barrier to multiple uses of data sets.

Factors that should be considered include:

- The standards employed should be explicitly mentioned as this is the first requirement for interoperability.
- Adoption of the practices of disciplines most advanced in this respect should be promoted, in particular by the international professional organisations dealing with science and the collection and preservation of data for research and technological purposes.
- The work of organisations engaged in setting more general information and communication technology standards should also be considered.

I – Quality

The value and utility of research data depends, to a large extent, on the quality of the data itself. Data managers, and data collection organisations, should pay particular attention to ensuring compliance with explicit quality standards. Where such standards do not yet exist, institutions and research associations should engage with their research community on their development. Although all areas of research can benefit from improved data quality, some require much more stringent standards than others. For this reason alone, universal data quality standards are not practical. Standards should be developed in consultation with researchers to ensure that the level of quality and precision meets the needs of the various disciplines.

More specifically,

- Data access arrangements should describe good practices for methods, techniques and instruments employed in the collection, dissemination and accessible archiving of data to enable quality control by peer review and other means of safeguarding quality and authenticity.
- The origin of sources should be documented and specified in a verifiable way. Such documentation should be readily available to all who intend to use the data and incorporated into the metadata accompanying the data sets. Developing such metadata is important for enabling scientists to understand the exact implications of the data sets.
- Whenever possible, access to data sets should be linked with access to the original research materials, and copied data sets should be linked with originals, as this facilitates validation of the data and identification of errors within data sets.
- Research institutions and professional associations should develop appropriate practices with respect to the citations of data and the recording of citations in indexes, as these are important indicators of data quality.

J – Security

Specific attention should be devoted to supporting the use of techniques and instruments to guarantee the integrity and security of research data. With regard to guaranteeing the integrity of a data set, every effort should be made to ensure the completeness of data and absence of errors. With regard to security, the data, along with relevant meta-data and descriptions, should be protected against intentional or unintentional loss, destruction, modification and unauthorised access in conformity with explicit security protocols. Data sets and the equipment on which they are stored should be protected as well from environmental hazards such as heat, dust, electrical surges, magnetism, and electrostatic discharges.

K – Efficiency

One of the central goals of promoting data access and sharing is to improve the overall efficiency of publicly-funded scientific research to avoid the expensive and unnecessary duplication of data collection efforts.

Consideration should be given to the following:

- Data access arrangements should promote further cost effectiveness within the global science system by describing good practices in data management and specialised support services.
- While publicly funded research data are subject to the default rule of openness under Principle A, this does not mean that all such data should be preserved permanently. The data archiving community should carry out cost-benefit assessments periodically and constantly develop and refine retention protocols to ensure that those data sets with the greatest potential utility are preserved and made accessible. Use of accepted retention protocols and thorough documentation of data should help to reduce unnecessary duplication of effort as well as to establish the necessary selectivity in preservation.
- Specialised support services, for example through collaboration with non-academic specialists on specific research projects or the engagement of data management specialist organisations, should be considered as a means to ensure the cost-effective production, use, management and archiving of research data.

- Insufficient incentives for researchers or database producers may lessen their efforts on data-related activities. The development of new reward structures and the adaptation of existing ones, including recognition of data management activities in tenure and promotion review, should be considered as a way to address this problem.

L – Accountability

The performance of data access arrangements should be subject to periodic evaluation by user groups, responsible institutions and research funding agencies. Although each party is likely to use somewhat different evaluation criteria, the sum total of the results should provide a comprehensive picture of the value of data and of data access regimes. Such evaluations should help to increase the support for open access among the scientific community and society at large.

The following should be considered in establishing evaluation criteria:

- Overall public investments in the production and management of research data.
- Management performance of data collection and archival agencies.
- Extent of re-use of existing data sets.
- Knowledge generated from the re-use of existing data.
- The use of targeted foresight exercises to determine the nature and scope of data preservation activities and the types of data most likely to be needed in the future.

Even if gaining clear insight into the cost, benefit and performance of data access arrangements will not be an easy task, those in charge of data access arrangements should put effort into showing the benefits of open data access to justify and help ensure sustained support from all levels of government.

M – Sustainability

Due consideration should be given to the sustainability of access to publicly funded research data as a key element of the research infrastructure. This means taking administrative responsibility for the measures to guarantee permanent access to data that have been determined to require long-term retention. This can be a difficult task, given that most research projects, and the public funding provided, have a limited duration, whereas ensuring access to the data produced is a long-term undertaking. Research funding agencies and research institutions, therefore, should consider the long-term preservation of data at the outset of each new project, and in particular, determine the most appropriate archival facilities for the data.