OECD PRINCIPLES AND GUIDELINES FOR ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

Dirk Pilat[1]* and Yukiko Fukasaku[2]

*[1] Organisation for Economic Co-operation and Development, 2, Rue André-Pascal, 75775 Paris, CEDEX 16, France
Email: dirk.pilat@oecd.org
[2] Innovmond, Paris, France
Email: yukiko.fukasaku@innovmond.fr

ABSTRACT

Ministers of science and technology asked the OECD in January 2004 to develop international guidelines on access to research data from public funding. The resulting Principles and Guidelines for Access to Research Data from Public Funding were recently approved by OECD governments and are discussed below. They are intended to promote data access and sharing among researchers, research institutions, and national research agencies. OECD member countries have committed to taking these principles and guidelines into account in developing their own national laws and research policies, taking account of differences in their respective national context.

Keywords: Open access, Access to research data, Science policy, International guidelines

1   INTRODUCTION

Throughout OECD member countries, 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 knowledge to help address the myriad challenges facing humanity. To assist OECD member countries in their efforts to provide access to these research data, the OECD has developed *Principles and Guidelines for Access to Research Data from Public Funding*. These principles and guidelines provide a broad framework that is intended to assist governments and research funding bodies of Member countries in formulating their policies. They are intended to promote data access and sharing among researchers, research institutions, and national research agencies, while recognising and taking into account, the various national laws, research policies and organisational structures of Member countries.

The ultimate goal of the *Principles and Guidelines* is to improve the efficiency and effectiveness of the global science system. New information and communication technologies (ICTs) are opening up new avenues for collaboration and sharing of research data. 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. Sharing and increased access to publicly funded research data not only helps to increase the potential of new digital technologies and networks in advancing research, but also increases the returns from public investment in research. Improved access to, and sharing of, data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, enables 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, and permits the creation of new data sets when data from multiple sources are combined.

The power of computers and the Internet have created new fields of application for not only the results of research, but also for the *sources* of research: the base material of *research data*. 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. 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. 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; and the enhancement of value from international co-operation.

To promote greater scientific and social returns from 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. The recently adopted *OECD Recommendation* on Principles and Guidelines concerning Access to Research Data from Public Funding is intended to support the development of good practices in this area and aims to foster the global exchange and use of research data.

2   DEVELOPING THE OECD PRINCIPLES AND GUIDELINES

In January 2004, Ministers of science and technology of OECD countries met in Paris and discussed the need for international guidelines on access to research data. At the meeting, the governments of the 30 OECD countries and of China, Israel, Russia and South Africa adopted a *Declaration on Access to Research Data from Public Funding*. In this declaration, they recognised the importance of access to research data and invited the OECD "to develop a set of OECD guidelines based on commonly agreed principles to facilitate optimal cost-effective access to digital research data from public funding to be endorsed by the OECD Council at a later stage".[1]

This request was taken up by OECD's Committee for Scientific and Technological Policy, which launched a project by asking a group of experts to develop a set of principles and guidelines. The experts drafted a first set of principles and guidelines and engaged in several rounds of consultation with research institutions and policy making bodies in the OECD member countries to achieve a consensus. A workshop involving key stakeholders was held in Paris in February 2006, which also contributed to this process. The work leading up to the final draft revealed that international frameworks to facilitate access were still lacking in the Member countries, but also that improved access was generally seen as benefiting the advancement of research, boosting its quality and facilitating cross-disciplinary research co-operation. Stakeholders considered that international guidelines would be useful in giving guidance to institutions in need of policies and in enhancing international co-operation in research. The principles and guidelines that resulted from this extensive consultation process were approved by the OECD's Committee for Scientific and Technological Policy in October 2006. The Principles and Guidelines were attached to an OECD Recommendation and endorsed by the OECD Council on 14 December 2006.[2]

A "Recommendation" is a legal instrument of the OECD that is not legally binding, but through a long standing practice of the Member countries, is considered to have a great moral force. Recommendations of the OECD are adopted when Member governments are prepared to make a political commitment to implement the principles (and/or guidelines) set out therein. This type of instrument is often referred to as "soft law".

OECD Recommendations set out collective and precise standards or objectives which the Member countries are expected to implement. Although some flexibility in the drafting of the standard is possible, the flexibility should not be so wide as to allow individual Member countries to modify the standard or object, consequently defeating the idea that there is a commitment on their part. That said, OECD Recommendations often leave significant flexibility to Member countries as regarding the means through which they implement them to take into account differences in legal, cultural, economic and social contexts. Thus, in some countries, implementation could be achieved through regulatory measures while others will choose, for example, to resort to concerted action with national stakeholders.

Recommendations are considered to be vehicles for change, and OECD Member countries need not, on the day of adoption, already be in conformity. What is expected is that they will work seriously towards attaining the standard or objective within a reasonable time frame, while taking account of the extent of difficulty in closing the gap in each Member country.

---

[1] The text of the Ministerial declaration on access to research data can be found on the OECD Internet site, at: http://webdomino1.oecd.org/horizontal/oecdacts.nsf .

[2] The formal text of the Recommendation can be found at: http://webdomino1.oecd.org/horizontal/oecdacts.nsf. The associated Principles and Guidelines have been published by the OECD and can be found at: http://www.oecd.org/document/55/0,2340,en_2649_34269_38500791_1_1_1_1,00.html.

The *OECD Recommendation* states that Member countries should take into consideration the Principles and Guidelines on Access to Research Data from Public Funding set out in the Annex to the Recommendation, as appropriate for each Member country, to develop policies and good practices related to the accessibility, use and management of research data. It also instructs OECD's Committee for Scientific and Technological Policy to review the implementation of this Recommendation as necessary; as well as to review the Principles and Guidelines on Access to Research Data from Public Funding when appropriate, to take into account advances in technology and research practices, with the intention of further fostering international co-operation.

The resulting *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 in developing data access arrangements. 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.

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.*

The specific aims and objectives of the *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; and

- Offer recommendations to Member countries on how to improve the international research data sharing and distribution environment.

The *Principles and Guidelines* 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 the following major issues inherent in providing data access:

- Technological issues: access to research data, and their optimum exploitation, requires appropriately designed technological infrastructure, interoperability at the international level, and effective data quality controls;

- Institutional and managerial issues: while increased accessibility benefits research 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.

- Legal and policy issues: national laws and international agreements, particularly in areas such as intellectual property rights and the protection of privacy must be fully taken into account in the design of data access arrangements; and,

- 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.

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.

3   PRINCIPLES AND GUIDELINES FOR ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

3.1   Scope and definitions

What precisely are these principles and guidelines? As discussed above, the *Principles and Guidelines for Access to Research Data from Public Funding* aim to 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 the *Principles and Guidelines* is to improve the efficiency and effectiveness of the global science system.

The *Principles and Guidelines* 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 the *Principles and Guidelines*. 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 the *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).

The *Principles and Guidelines* are aimed principally at research data in digital, computer-readable format. It is in this format of course 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*, however, 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 data obtained from research conducted by government agencies or departments, or conducted using public funds provided by any level of government. As the nature of "public funding" of research varies significantly from one country to the next, these *Principles and Guidelines* recognise that a flexible approach to improved access to research data is required.

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.

3.2 Principles

The following principles are considered important in developing access arrangements for research data from public funding.

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 should aim at being easy, timely, user-friendly and preferably Internet-based.

B – Flexibility

Flexibility means 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. This implies that 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*, specific national, social, economic and regulatory contexts should be taken into account.

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. Research organisations and government research agencies should actively disseminate information on research data policies. Standards for cataloguing data should be established and agreed by the various research communities. 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

D – Legal conformity

Data access arrangements should respect the legal rights and legitimate interests of all stakeholders in the public research enterprise. This means that research data access and use 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.

- 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.

- 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. As public/private partnerships in the funding of research and related data production are increasing, balanced public/private arrangements should be worked out to facilitate broad access to research data. 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. 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 should be developed in consultation with representatives of all directly affected parties. In collaborative research programmes, international scientific co-operation or public / private partnership based projects the parties involved should negotiate research data sharing arrangements as early as possible in the life of the research project. 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.

With data management becoming ever more complex in certain areas of research, formal and explicit institutional agreements on the terms of access and use are needed. 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. As long-term sustainability of the data access infrastructure is 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. Several factors need to be considered in this context. First, the use of codes of conduct for professional scientists and their communities could help simplify and reduce the regulatory burden placed on access. Second, 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. Moreover, 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

Although science is becoming a highly global endeavour, incompatibility of technical and procedural standards can be a serious barrier to multiple uses of data sets. Technological and semantic interoperability is therefore 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.

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. Also, the work of organisations engaged in setting more general information and communication technology standards should 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. As 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. Also, research institutions and professional associations should develop appropriate practices with respect to the citations of data and the recording of citations in indexes.

- As developing metadata is important for enabling scientists to understand the exact implications of the data sets, the origin of sources should be documented in a verifiable way and should be readily available.

- 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.

J – Security

In developing access regimes, special attention should also be devoted to guaranteeing the integrity and security of research data, notably by supporting the use of appropriate techniques and instruments. Every effort should be made to ensure the completeness of data and absence of errors. The data, along with relevant meta-data and descriptions, should be protected against intentional or un-intentional 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. Data access arrangements should promote further cost effectiveness 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.

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. Also, researchers or database producers should be adequately recognised and rewarded for their data management activities in tenure and promotion review.

L – Accountability

Data access arrangements should be subject to periodic evaluation by user groups, responsible institutions and research funding agencies. Such evaluations should help to increase the support for open access among the scientific community and society at large. The cost, benefit and performance of data access arrangements may not be easy to establish, but efforts should be made to demonstrate the benefits of open data access to justify sustained support from all levels of government

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.

M – Sustainability

Due consideration should also 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. As most research projects, and the public funding for them, are limited in duration, research funding agencies and research institutions, therefore, should consider the long-term preservation of data at the outset of each new project, including the most appropriate archival facilities for the data.

4   FOLLOWING UP ON THE PRINCIPLES AND GUIDELINES

The OECD Principles and Guidelines do not necessarily mark the end of the OECD's work on data access. Further work may be required in the near future. One element of this work is monitoring how countries are implementing the OECD Recommendation on data access, as required by the Recommendation itself. Another is returning to the principles and guidelines at a later stage, to ensure that they remain relevant for policy development.

Other work might be needed, however. For example, good practices in implementing sustainable access arrangements are still being developed in many countries and could benefit from further exchange among the countries and institutions involved. Moreover, questions remain on the costs and benefits of providing access to data from research. These, and other, issues might be addressed by the OECD in future work, providing this is considered a priority by OECD member countries.