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

26-Nov-2013

English - Or. English

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY
COMMITTEE FOR SCIENTIFIC AND TECHNOLOGICAL POLICY

DSTI/STP/TIP(2013)13
For Official Use

Working Party on Innovation and Technology Policy

BACKGROUND PAPER FOR THE TIP WORKSHOP ON OPEN SCIENCE AND OPEN DATA

12 December 2013
OECD Conference Centre, 2 rue André-Pascal, 75016 Paris

Delegates will find a background paper for the TIP Thematic Workshop on Open Science and Open Data to be held at OECD headquarters, starting at 9:00 am. The thematic workshop will be held in English only.

Contacts: Giulia Ajmone Marsan; E-mail: giulia.ajmonemarsan@oecd.org;
Mario Cervantes; E-mail: mario.cervantes@oecd.org

JT03349247

Complete document available on OLIS in its original format

English - Or. English

1. Introduction

1. Information and communication technologies, new data storage infrastructure and large scale computing are modifying the way science and research are conducted, disseminated and diffused. On the one hand, the Internet and related online platforms are creating new opportunities to organise and publish the content of research projects, scientific publications and large data sets, so as to make it immediately available to other scientists, researchers as well as potential users in the business community and the society in general. On the other hand, ICTs allow the collection of large amount of data that can be at the basis of scientific experiments and research, contributing to make science increasingly data-driven. As a consequence, it is now possible to access, use and re-use research and scientific inputs and outputs (both articles and datasets) at free or extremely low marginal cost and speed the transfer of knowledge amongst researchers and across scientific fields, opening up new ways of collaboration and new research domains. This transformation of science into a more open and data-driven enterprise is often grouped under the term "Open Science" or "Science 2.0".

2. However while open science creates new possibilities for research as well as innovation it also creates new challenges for policy makers who need to better understand the implications for policy making and the broader social and economic impacts.

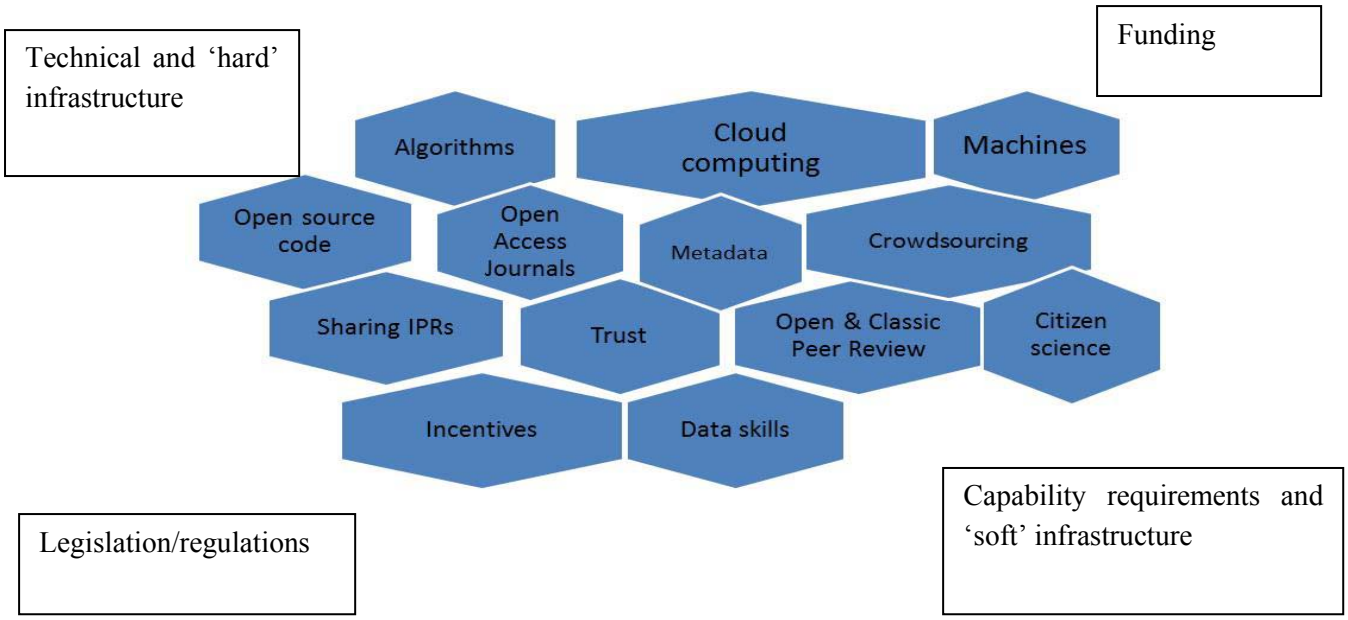
3. This background paper sets out the key issues for discussion at the TIP "*Thematic Workshop on Open Science and Open Data*", to be held on the 12th December 2013 at the OECD Headquarters in Paris. Each section of this note describes and highlights potential benefits but also risks associated to different aspects of open science and open data that will be discussed at the workshop.

2. The power of open and data-driven science for research and innovation

4. Open and data-driven science, especially big data, is changing the way science is carried out as it allows researchers to scale research through the use of algorithms, cloud computing and crowdsourcing of ideas and findings, including via "citizen science". In addition, as science becomes data driven its social and economic value is enhanced. Just as business or government data are increasingly being used to create new goods and services, scientific data is enhancing the quality and output of scientific research. Open research data allows scientists in one field to exploit data in other fields and reveal relationships or patterns that were not visible before. It allows them to formulate and test new hypotheses and improve predictability of scientific models.

5. With respect to research, the power of open and data driven science lies in paving the way for a fundamentally different approach to science, with broad implications for the way academic research is funded and carried out, the infrastructure that is needed and the skills required. The new way of doing science is based on several building blocks, notably an ICT infrastructure of hardware, algorithms for exploiting large data sets, open source software and interoperability of operating systems, metadata, appropriate IPR arrangements and the sourcing of knowledge and data from the crowd via the Internet. Incentives and trust however remain important in this new eco-system for science (Figure 1).

Figure 1. Building Blocks for Open and Data Driven-Science or Science 2.0



6. In many ways the open science is also affecting the pact between scientists and academia and between scientists and society. Science has essentially been self-regulated by academics with occasional incursions from government ministries and funding agencies. The social compact for public support to science has relied on trust with society. As science is funded and produced in a more open and transparent manner the "trust" between science and society will be subject to greater scrutiny from citizens.

7. Meanwhile open and data driven science has important effects on innovation. For example the links to innovation are enhanced as business themselves can use the tools of science to exploit data. Businesses can create new products or services by combining marketing data with scientific tools like hypotheses testing, advanced statistical analysis and benchmarking against control groups. In short the scientific method and approach, via data analytics, is increasing the knowledge intensity of business processes and the content of consumer products. This potentially allows firms to continue to specialise and gain competitive advantage even in mature markets. Box 1 below highlights some of the key policy rationales for efforts to promote open science and open data for research and innovation.

Box 1. Rationales for Open Science and Open Data for Research and Innovation

The following factors are often associated to openness in science and research:

- **Improved efficiency in science**: open science efforts can increase the effectiveness and the productivity of the research system by: 1) reducing duplication and costs of creating, transferring and re-using data; 2) allowing more research from the same data; 3) multiplying opportunities for domestic and global participation in the research process;

- **Increase transparency and quality in the research validation process,** by allowing to a greater extent replication and validation of scientific results;

- **Speeding the transfer of knowledge:** open science can reduce delays in the publication of articles and data sets and promote in a faster way the path from research to innovation;

- **Increasing knowledge spillovers to the economy:** increasing access to the results of publicly funded research can foster spillovers and boost innovation across the economy. The disclosure and release of public and scientific data may promote the development of innovative products and services in firms as well as increase awareness and conscious choices among consumers;

- **Addressing global challenges more effectively:** global challenges require co-ordinated international actions to be addressed. Open science and open data approaches may promote collaborative efforts and faster knowledge transfer for a better understanding and potentially identifying solutions of challenges such as climate change or ageing population;

- **Promoting citizens' engagement in science and research:** open science and open data initiatives may promote awareness and trust in science among citizens. In some cases, a greater citizens' engagement may lead to active participation in scientific experiments and data collection.

Source : adapted from DSTI/STP/TIP(2013)5

3. Evolving organisational and business models for diffusing results of public research

8. The changing nature of the scientific enterprises is reflected in the emergence of new organisation and business models for the diffusion and dissemination of scientific publications and data sets. In recent years the advent of the internet and the decreasing costs of publishing have transformed and modified the range of services offered by scientific publishers. Open access journals have emerged and traditional publishers have started offering new kinds of services that go beyond the distribution of articles to subscribers. Publishers collect large databases containing bibliometric information often used in the evaluation of science and research, store data sets and images related to scientific experiments, offer social networking services and platforms for collaborative science and research. Box 2 below provides a brief description of the different open access publishing schemes.

Box 2. Openness and scientific publications: multiple routes

The term *open access* refers to the possibility to access scientific articles published in academic journals online, free of charge to the reader and, at least in some cases, free of copyright and licensing restriction. There exists different path of open access publishing associated to different level of rights to use and re-use both the text and the data contained in the article. Open access paths are generally divided into green or gold schemes.

Green open access: the scientific article is archived by the author along or after the traditional publication in peer reviewed journals. Often the article is freely accessible after an embargo period that follows the traditional publication on scientific journals. Green open access is also called "self-archiving" or the green route.

Gold open access: the scientific article is immediately available free of charge to the reader at the time of publishing. The associated cost for publishing are shifted from the readers (namely, journals subscribers) to the author of the publication (via funding agencies, universities or in general the institution the author is affiliated to) who pays for having the article published following the gold route.

Increasingly, traditional (non-open access) scientific journals are offering the possibility to publish single articles through the gold route, as a consequence articles free of charge to the reader and article accessible only to paying subscribers can co-exists on the same journal.

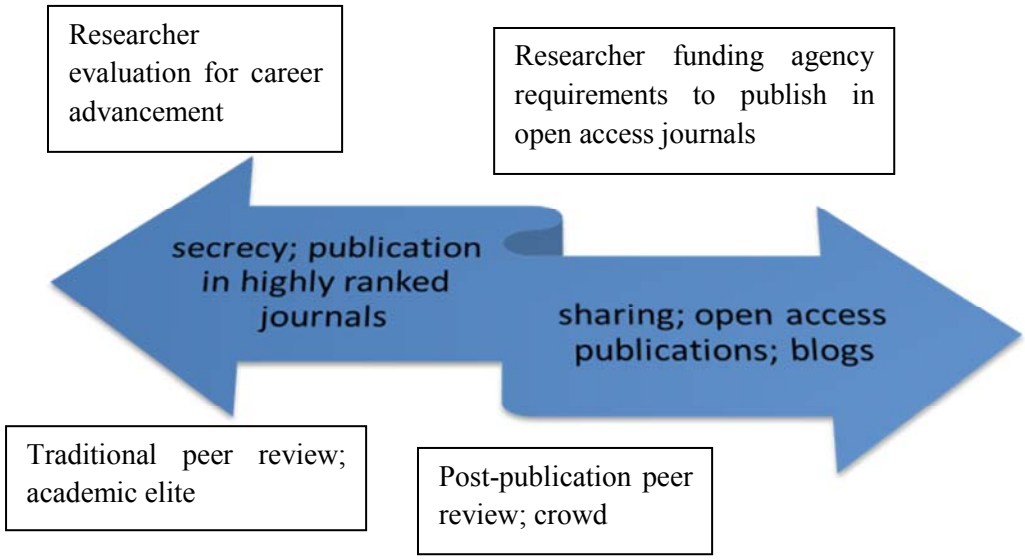
Source : adapted from DSTI/STP/TIP(2013)5

9. Some of the potential challenges associated to open access to scientific findings are related to maintaining high quality standards in scientific publications and to the incentive mechanisms for researchers and scientists to disclose the results of their research. At the centre of this challenge is the question of the ability of the peer review system to remain robust in the face of greater scientific output and the emergence of open access publishing. However, the recent mediatisation of weakness in peer review process of open access journals[1] should not detract policy makers' attention from the challenges to the peer review system in general (e.g. overly burdened reviewers, conflicts of interest, etc.) including in traditional peer review. The business models for scientific publishing will need to address potential issues related to the right level of quality control and peer-reviewing, especially if the costs associated to the publication are shifting from subscribers to authors (like in the gold open access scheme).

10. Other issues to consider are related to the incentives (or disincentives) for scientists to share knowledge and research results. If the traditional business models for publication are evolving towards more openness and knowledge sharing, incentives mechanisms for scientists and researchers may need to be revisited and adapted. Scientific excellence is associated to global competition among researchers, to discover and consequently publish first in top quality journals. The "*publish or perish"* ethos in research may discourage collaboration and prevent researchers from sharing and diffusing data sets or ideas, especially at the pre-publication stage. In addition, most evaluations of universities and researchers are almost entirely based on teaching and bibliometric indicators, attributing little value to the sharing pre-publication inputs, such as data and other relevant information (Figure 2).

[1] For example, J. Bohannon, Who's Afraid of Peer Review?, Science, 4 October 2013: Vol. 342, 6154, pp. 60-65, available at: http://www.sciencemag.org/content/342/6154/60.full

Figure 2. Conflicting incentives and channels for researchers to disclose results?



11. Finally, open science initiatives call for the definition of the "right" balance between openness and IP protection. For example, open access to scientific publication may not necessarily involve access to data and the associated right of use and re-use and vice-versa. Other challenges may be related to the definition of ownership of large scale datasets, potentially collected by machines or software providers and privacy, confidentiality or even security issues, in particular with respect to certain classes of data, such as medical and biological records.

Questions for discussion:

- *What are the main characteristics of the **evolving business models** for open access to scientific research and data (differentiated access schemes, new services offered by publishers and IT firms, emerging opportunities for open collaboration among researchers, etc.)?*
- *How to maintain high **quality** standards in scientific journals and data collection while at the same time providing open access and use?*
- *How to design **incentive mechanisms** for scientists, researchers and data collectors, to promote effective knowledge and data sharing (cost sharing schemes of open access, career advancement rules, etc.)?*
- *Open science and **IP**: what **mechanisms of protection** are the most effective to promote sharing and openness in science without offsetting incentives for discovery and innovation?*

4. Public funding, infrastructure and skills: how to make open science sustainable

12. Despite a broad consensus around the principle of open access, the issues related to the long-term sustainability of open access journals and open data efforts are still unsolved questions. In many OECD countries, open science initiatives ranging from open access schemes to large database infrastructure and collaboration platforms are largely financed through public subsidies. In the case of gold open access

publishing, for example, the cost of publishing are shifted from journal subscribers to the authors of scientific publications, universities or research institutes.

13. The increasingly data-driven nature of science and research calls for investing and building the IT infrastructure associated to data gathering, cleaning and storage. All these operations are costly, time-consuming and need to be performed constantly over time. Other barriers to openness and knowledge transfer are associated to the inter-operability of systems, the international standards in use, the development of clear and explanatory meta-data. The solutions to these issues are not technically insurmountable per se, but require a high degree of international co-ordination (such for example in the adoption of common standards), investments in the infrastructure itself and also in the development of the right set of skills for the personnel devoted to taking care of such datasets or platforms. Not all researchers may have the time, the willingness or the right skills to perform such operations. Possible solutions may involve a closer cooperation with the private sector, for example public-private partnerships for the delivery of services associated to the maintenance of the infrastructure necessary for open science platforms.

Questions for discussion:

- *Are public **subsidies** for open access sustainable? How to design self-sustainable and long-lasting open science initiatives? Are there conflicts of interest arising from the "gold" route to open access publishing?*

- *How to develop and maintain **infrastructure** for open science and open data? Issues of inter-operability, metadata, harmonised data cleaning and updating, etc.*

- *How to develop the necessary set of **skills** to build and maintain open science and open data archives, platforms and databases?*

- *What is the role of **public-private partnerships** in sharing risks and benefits?*

5. Measuring the impact of open access and open research data

14. Open access initiatives generally aim to increasing the speed of and broadening the scope for the diffusion of scientific results from publicly funded research. This may involve open access to scientific publications or data (both in terms of inputs and outputs of scientific research), open collaboration research platforms, greater dissemination of scientific research to citizens and associated use of the results by the society. In order to better understand the impact of these relatively new phenomena, new metrics and indicators will need to be defined. For instance, the traditional bibliometric indicators may be complemented by metrics that capture the level of the dissemination of scientific papers such as internet downloads, discussion in scientific blogs and non-technical journals, etc. However, how and whether further dissemination is associated to excellence and quality still needs to be carefully quantified. In addition the collaborative nature of open science increasingly raise the importance of the definition of indicators capturing knowledge flows and research interactions rather than volume of publications or citations only.

15. As regards indicators of the impact of open science and open research data, it is possible to distinguish between the impact on the efficiency of the research system (in terms of speed of diffusion,

reducing duplication of research efforts, reduction of costs for research conducted by larger teams, etc.) and the impact of open science initiatives on research, innovation and the economy more generally. There are a few avenues for developing indicators on the use of public science in innovation processes, such as innovation surveys, references to scientific literature in patents (science-patent links), and the exploitation of social network statistics. Such new indicators, through their impact on benchmarking and comparative analysis, could help promote greater disclosure of scientific results and data by the academic and scientific community, governmental organisations, but also by the business sector in order to fuel innovation.

Questions for discussion:

- *How to **define open science** for the purpose of measurement and how to measure it?*
- *How to measure **the impacts on the research system**?*
- *How to measure **the impacts on innovation** (e.g. patenting, creation of research data start-ups, creation of new business and services, etc.)?*
- *How to capture the impact of open access and open data on international research, both in terms of **scientific networks** and **cross-border open data flows**?*
- *What **new indicators** can be used or developed to measure and assess impacts? How to measure the degree of dissemination of scientific research not only amongst scientist and experts but also amongst **citizens**?*
- *What can be learned from **users** of open science and open data repositories?*