

5

The future of access to data for science, technology and innovation

This chapter draws conclusions from the preceding work and proposes possible ways forward for the future of access to data for science, technology and innovation. It starts by summarising the policy issues identified and infers implications for policy makers. It concludes by providing potential scenarios and a view of potential future developments in this policy field.

Policy issues and implications

Issues identified in enhancing access to data for STI

The analytical work in the previous chapters has identified the following main policy issues:

- i) **Balancing the benefits of data sharing with the risks.** “As open as possible, as closed as necessary” is gradually replacing the “open-by-default” mantra associated with the early days of the open-access movement. Opening data can provide benefits in advancing the science, technology and innovation (STI) agenda, but these need to be balanced against the issues of cost, privacy, security and malevolent uses. A staged approach should be used to enhance access to data, including sharing within communities of certified users, adapting the degree of certification of users to the sensitivity of data and creating safe environments where certified users can access sensitive datasets in controlled environments.
- ii) **Technical standards and practices** – keeping up with the pace of technological progress. Applying the findability, accessibility, interoperability and reuse (FAIR) principles depends on developing and adopting common technical frameworks. The challenge is that technology development is now far outpacing standard-setting, creating regulatory gaps. Implementation of FAIR principles is an important initiative to close this policy gap.
- iii) **Defining responsibility and ownership.** Intellectual property right (IPR) protection is a basic condition for incentivising innovation. However, advances in technology can provide opportunities for new methodologies, such as text and data mining (TDM). Copyright regulation which excludes temporary copies of text for the sole purpose of TDM can represent an impediment to research and innovation. In the case of public-private partnerships, policy objectives should be clearly defined to expressly allow or forbid private ownership over the data derived from publicly funded research.
- iv) **Incentives and rewards.** Recognition and rewards are needed to encourage researchers to share data. Current academic reward systems mostly encourage the publication of scientific results and do not sufficiently value data sharing. More remains to be done to raise awareness of open-government data among researchers and enhance the appeal of sharing access to data.
- v) **Business models and funding for provision of enhanced access.** Costs are most often borne by data providers, while benefits accrue to users. Although enhanced access does not necessarily mean free data, in most cases, data should be free at the point of use. What are the financing options?
- vi) **Building human capital and institutional capabilities to manage, create, curate and reuse data.** Better data skills for researchers, data stewards and users are a prerequisite for advancing data sharing.
- vii) **Exchange of sensitive data across borders.** Sensitive datasets can be shared on a more restricted basis with trusted and certified users. Significant barriers currently exist to providing such services across borders, owing to a lack of international legal frameworks ensuring the same levels of legal protection against misuse.

Implications for policymakers

Building on current experience and looking forward, some policy implications can be drawn for governments in each of these domains:

- i) **Balancing the benefits of data sharing with the risks**
 - Public data for STI should be “as open as possible, as closed as necessary”. Governance arrangements are critical when accessing sensitive data.
 - Institutions should develop realistic consent frameworks and set up ethics review boards, with a mandate to arbitrate in cases where obtaining consent is impossible or impractical.

- Governments should strive to enhance trust among different stakeholders, and create consensus around data sharing and reuse. The risk of privacy breaches cannot be completely avoided, but should be managed through clear and transparent procedures. Creating safe environments that are open to certified researchers in controlled environments is a step in this direction.
 - Specific initiatives can be launched to support data integration, exploring ways in which data from different sources can be combined transparently across different institutions. These initiatives should explore important issues related to sensitive data, such as anonymisation and informed consent.
 - Socio-economic assessments should be undertaken to monitor the impact of open research data, with specific attention to where – and to whom – benefits accrue. Steps should be taken to ensure broad access to research data, to avoid a new “digital divide”.
 - Solutions should be explored for data integration across sectors and disciplines (e.g. providing cross-disciplinary access to register data in Sweden).
 - Governments should implement proactive campaigns to foster systemic trust in data initiatives. Measures include informing and actively engaging with stakeholders, performing risk management and mitigation, and enforcing mandatory breach notification.
- ii) **Technical standards and practices – keeping up with the pace of technological progress**
- Developing and adopting community-agreed standards is critical to FAIR data. Individuals and bodies (such as the Research Data Alliance) that work in this area should be supported accordingly.
 - Good metadata are critical to data interoperability and reuse. Data controllers should be encouraged to comply with standardised reference models (e.g. the Open Archival Information System).
- iii) **Defining responsibility and ownership**
- Information about ownership and licensing should be contained within the metadata, and specified for all prospective data products in research-data management plans. Open-use licences should be used wherever appropriate (OECD, 2015).
 - OECD member countries, including their intellectual property (IP) protection expert agencies, should carefully consider the implications of any amendments to copyright legislation and IPR regimes as they relate to access to publicly funded data for research. They should implement well-established international IP norms and promote (rather than inhibit) research and innovation in new areas, such as text and data mining, and deep learning.
 - Specific arrangements could promote data sharing in public-private partnerships, while respecting the legal rights and legitimate interests of all stakeholders in the public-research enterprise.
- iv) **Incentives and rewards**

Specific policies should be promoted to incentivise data sharing among researchers. Such measures could include:

- developing new indicators and metrics for data sharing, and incorporating them into institutional-assessment and individual researcher-evaluation processes
 - promoting the use of unique persistent digital identifiers for individual researchers and datasets, to enable citation and accreditation
 - developing attractive career paths for data professionals within the research-data ecosystem – they are necessary to the long-term stewardship of research data and service provision, but are very often attracted by private-sector careers over public-sector ones.
- v) **Business models and funding**

Public data for STI is a public good. Sound policies are needed to optimise data sharing while ensuring rational use of public monies. Governments could consider:

- developing strategies and roadmaps, including long-term funding plans and business models, to build a sustainable research-data infrastructure (i.e. data repositories and services)
- exploring how public investment in research data and infrastructure can be used to leverage private investment (as well as skills and data resources), while ensuring openness and accountability.

vi) **Building human capital**

Human capital is needed to manage, create, curate and reuse data. Depending on the discipline, scientists may or may not have adequate data-management skills. Users of the data do not always have the appropriate skills for interpretation and analysis. The data stewards sometimes lack skills to apply the relevant standards. Possible measures include:

- developing a national data-skill strategy for STI, identifying specific skill gaps, as well as the education and training requirements needed to fill them
- facilitating co-operation across different education and research actors, to ensure coherence and complementarity in capacity-building activities for data skills.

vii) **Exchange of sensitive data across borders**

- International legal frameworks should be developed to ensure the same level of legal protection against misuse. Such frameworks should adopt common approaches and rules for sharing data (especially personal and other confidential data) in safe environments, facilitating exchanges across borders.

Potential future developments in this policy field

The significance of data for STI will undoubtedly continue to increase over the next decade. The volume of data produced globally amounted to 16 zettabytes (ZB) in 2016 and is expected to grow to 163 ZB by 2025 (Reinsel, Gantz and Rydning, 2017). The importance of artificial intelligence in assisting scientific discovery is also expected to grow significantly. Access to well-managed data is a key enabler of this development (Kitano, 2016).

Enhanced access to research data holds considerable promise for increasing research productivity and innovation, and developing solutions to complex societal challenges. However, realising this potential – and minimising the potential risks – will require strategic planning and policy interventions. The *OECD Recommendation of the Council concerning Access to Research Data from Public Funding* (OECD, 2006) and the more recent FAIR principles for data access provide a broad framework for policy development and co-operation across communities. Many countries have already taken up the challenge, and adopted open-science policies and/or strategies on open access to research data. At the European level, the European Commission has taken the lead in ensuring policy coherence across countries.

Beyond data as such, access sharing will need to be increasingly applied to a broader category of scientific information, including software and publications. Clearly, data are linked to publications for reasons of reproducibility, and a growing number of publishers and funders require the publication of supporting data. Over time, this trend may be extended to raw data, beyond the subset needed for immediate reproduction of results (for example, the European Commission's Data Management Pilot encourages sharing data beyond the bare minimum on a voluntary basis). Conversely, access to datasets should be linked to shared access to publications featuring the results obtained from these data.

Access to the appropriate software and algorithms is gaining importance in ensuring the correct usage and interpretation of data. This interdependence means that large volumes of data can only be analysed with appropriate algorithms, and vice versa: algorithms can only be trained through exposure to vast quantities of data.

Blockchain technology is a potential tool that could improve the traceability of inventions, providing a way of tracing the source of innovation back into the network of public collaborative science and innovation.

Successful implementation of open-data policies and strategies crucially requires establishing governance systems and processes that ensure transparency and foster trust across the research community and society at large. Mandates and incentives will need to be used judiciously to support and facilitate changes in research behaviour, without stifling creativity and innovation. Long-term investment in technical infrastructure and human capital will be required. Technical standards need to be developed, and legal and ethical concerns addressed.

Several ways forward are possible, of which Box 5.1 presents two extreme cases. It is up to policymakers to decide which scenario would best suit the national interests and to activate the levers to promote the preferred scenario.

Much needs to be done, but much is already being done. Understandably, policy intervention focuses on exploiting the exciting opportunities created by enhanced access to research data. Enhanced access to data can help address issues related to the reproducibility and accountability of scientific research, provide solutions to pressing socio-economic challenges and unite the global scientific community around these issues. Looking to the future, however, it is also important to consider and mitigate the potential risks.

The advent of data-driven science coincides with a crisis of confidence in science and the advent of the “post-truth” era. Opening up public-research data means that new actors will be able to analyse and interpret the data from their own perspectives, and not necessarily with the critical objectivity expected from scientists. The old adage “if you have enough data, you can prove anything” is not unfounded.

In the new world of open science, the scientific community will need to work rigorously, clearly communicating the scientific method and limitations of its analyses, and engaging in honest discourse and dialogue with the public and policymakers. In a hyper-competitive research enterprise characterised by enormous pressure to succeed and growing hype around scientific breakthroughs, it is vital to ensure that open science and data can be trusted. Technological developments (such as blockchain) can assist in this regard. Ultimately, however, trust is a social construct, which needs to be carefully nurtured over time.

Box 5.1. Possible future ways forward for enhanced access

In a possible “best-case” scenario, trust would be earned across society, thanks to strong governance initiatives ensuring strong risk management and mitigation, elaborated in transparent consultation with stakeholders. Ethics review boards would credibly represent individual interests and arbitrage consent issues. On the technical side, strong global standards would emerge, akin to Transmission Control Protocol/Internet Protocol for Internet communication, complemented by more specialised standards for specific applications. IPR and licensing provisions would promote responsible data access and reuse and comprise a standard part of machine-readable metadata. Data citation would be ubiquitous and could become an integral part of researcher evaluation. Financing of repositories would be based on long-term infrastructure strategies and sustainable models. Finally, digital skills would be addressed through a strategic approach encompassing initial education and lifelong learning for data producers, stewards and users.

A “worst-case” scenario is also possible, in which repeated security and privacy breaches would be inadequately managed, fostering a general level of mistrust. Standards would continuously lag behind technology development, while IPRs would be insufficiently defined to support widespread data reuse. Incentives for researchers to publish their data would remain weak, and initiatives to develop data skills would be poorly designed.



From:

Enhanced Access to Publicly Funded Data for Science, Technology and Innovation

Access the complete publication at:

<https://doi.org/10.1787/947717bc-en>

Please cite this chapter as:

OECD (2020), “The future of access to data for science, technology and innovation”, in *Enhanced Access to Publicly Funded Data for Science, Technology and Innovation*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9bc5b630-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.