Chapter 12. The digitalisation of science and innovation policy

By

Dmitry Plekhanov, Michael Keenan, Fernando Galindo-Rueda and Daniel Ker.

Digitalisation will profoundly affect the public sector and the evidence base on which it formulates, implements, monitors and evaluates public policy. The science, technology and innovation (STI) policy field is no exception. In recent years, many countries have begun to develop digital science and innovation policy (DSIP) initiatives, to help build a picture of the incidence and impact of their science and innovation activities. This chapter provides an introductory overview of DSIP systems in OECD member and partner countries. Drawing on the findings of a recent OECD survey of DSIP initiatives, it first outlines the main characteristics of the DSIP systems currently in use and under development. It then describes the promises and challenges of DSIP systems. It shows that much can be gained from this digital transformation by leveraging the untapped potential of data about STI. However, obstacles and risks also exist. These relate to privacy and confidentiality, interoperability standards, and potential misalignment of incentives between policy objectives and STI actors, including the private sector. If DSIP initiatives are to fulfil their future potential, STI policy needs to address these opportunities and challenges, sometimes at the international level. The chapter concludes by considering the outlook for DSIP systems and possible avenues for policy action.

Introduction

Publicly funded research systems generate, in addition to key research data outputs, considerable amounts of information about the operation of those systems. Policy makers can use this information to monitor the performance and improve the efficiency of their research systems. Emerging digital science and innovation policy (DSIP) initiatives increasingly interconnect various information sources, and apply new technologies and applications that allow policy makers to exploit them more extensively. These systems can help build a picture of the incidence and impact of science and innovation activities, providing potentially valuable tools to facilitate decision-making across the broad spectrum of STI policy and administration. For instance, ministries can use DSIP systems to design, implement, monitor and evaluate policies. Funding agencies can use them to plan, co-ordinate, monitor and evaluate their activities. With the growing wealth of data about research and innovation, DSIP systems could transform the ways in which STI policy is defined and public services are delivered.

Several drivers of change are influencing these developments. First, government itself is undergoing a digital transformation. Digital technologies offer opportunities to increase the access, reach and quality of public services, as well as improve policy making and service design (OECD, 2018, 2014; Ubaldi, 2013). The STI policy field is no exception, although it is not as far along the digitalisation road as other policy areas. Second, the growing adoption of open science (OECD, 2015a; Dai et al., 2018) has created various infrastructures – such as data repositories and interoperability standards – which DSIP systems can readily re-use. Open science has also raised expectations that STI policy should also be open. Third, the emerging interdisciplinary field of science-of-science and innovation policy (Lane, 2010; Husbands Fealing et al., 2011) strongly emphasises developing data and metrics that STI policy makers can apply to their decision-making. Several DSIP initiatives originated in this field, and several others are influenced by it.

This chapter provides an introductory overview of DSIP systems in OECD member and partner countries. Drawing on the findings of a recent OECD survey of DSIP initiatives, it outlines the main characteristics of the DSIP systems currently in use and under development. It then describes the promises and challenges of DSIP systems. Finally, it considers the outlook for DSIP systems and possible avenues for policy action.

The DSIP landscape: A brief overview

"DSIP initiatives" refer to the adoption or implementation by public administrations of new or re-used procedures and infrastructures relying on an intensive use of digital technologies and data resources, to support the formulation and delivery of science and innovation policy. The primary goal of DSIP initiatives is to support certain aspects of the public-policy process, although any actor in the system – including in the private sector – can provide functionalities.

The OECD DSIP project is a first attempt at mapping the landscape of DSIP initiatives in OECD member and partner countries. It addresses the highly specific nature of digital government in the area of science and innovation policy. It includes a survey of 39 DSIP initiatives from 29 OECD and partner countries, which provides much of the evidence used to prepare this chapter.
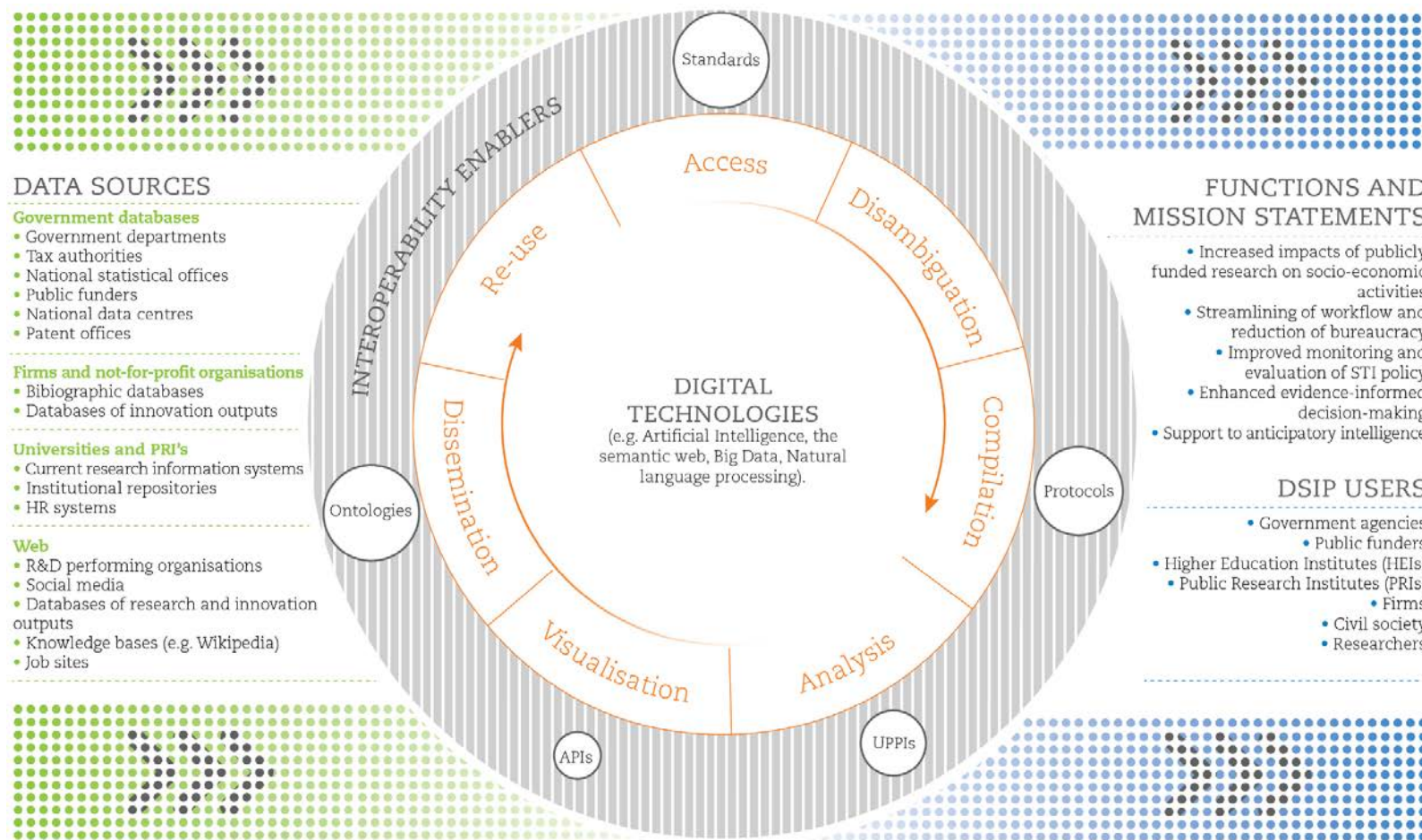
The results of the survey show that DSIP systems come in many shapes and sizes, making it difficult to classify them neatly. Broadly speaking, one group comprises systems that

build on a funding ministry or agency's administrative databases, linking them to other (typically external) data, e.g. to gain insights on funding outputs and impacts. Examples include Argentina's Sistema de Información de Ciencia y Tecnología Argentino (SICYTAR); South Africa's Research Information Management System; Poland's POL-on; and Federal RePORTER in the United States. Another group of DSIP systems consists of analytical solutions (often using machine learning, big data and semantic analysis) that collect and combine data from multiple data sources to provide insights for policy making. Examples include Corpus Viewer in Spain; Arloesiadur in the United Kingdom; SciREX Policymaking Intelligent Assistant System (SPIAS) in Japan; and iFORA in Russia.

While a few DSIP initiatives (e.g. Corpus Viewer) began as part of broader open government/big-data initiatives, most have originated in the STI policy domain. The main operators of DSIP systems captured by the OECD survey are STI ministries and funding bodies. Public research organisations (PROs) that provide governments with strategic policy intelligence services (e.g. evaluation and foresight) also operate DSIP systems in several countries (e.g. Japan and Korea). National statistical offices (NSOs) sometimes play a supporting role, shaped by their core statistical mandate and legislative framework, and the resources available to provide an enhanced range of digital services (Chapter 14).

Figure 12.1 provides a stylised conceptual view of a DSIP initiative and its main components. All of these elements interact in nationally specific ways, reflecting different histories and institutional set-ups. The main elements consist of various input data sources, which feed into a data cycle that is enabled by interoperability standards, including unique, persistent and pervasive identifiers (UPPIs). DSIP systems perform a number of functions and are often used by a mix of users. Box 12.1 highlights several examples of DSIP initiatives.

Figure 12.1. A stylised conceptual view of a DSIP initiative and its possible main components

> **Box 12.1. Examples of DSIP systems**
>
> In Argentina, the Ministry of Science, Technology and Productive Innovation uses SICYTAR to evaluate and assess STI policy initiatives, project teams and individual researchers. The system aggregates several databases, covering researchers' curriculum vitae; funded research and development (R&D) projects; information on public and private institutions performing R&D activities in Argentina; and information on large research equipment.
>
> In the Netherlands, the National Academic Research and Collaborations Information System (NARCIS) collects data from multiple sources, including funder databases, current research information systems (CRIS), institutional repositories of research performers and the Internet (Dijk et al., 2006). Data on research outputs, projects, funding, human resources and policy documents collected by NARCIS are used to inform policy makers on research activities undertaken in the Netherlands and to monitor open access. Funders also use the system to identify "white spots" in research to improve resource planning. NARCIS also serves as an important directory of research, providing researchers, journalists, and the domestic and international public with information on the status and outputs of Dutch science.
>
> In Norway, the research-reporting tool Cristin collects information from research institutions, the Norwegian Centre for Research Data and ethics committees. Cristin serves as a foundation for the performance-based funding model of the Ministry of Research and Education. It provides numerous users from government, industry, academia and civil society with verified information on the current status of Norwegian research.
>
> In Japan, the National Graduate Institute for Policy Studies designed the SPIAS system to strengthen national evidence-informed STI policy making. SPIAS uses big data and semantic technologies to process data on research outputs and impacts, funding, R&D-performing organisations and research projects, with a view to mapping the socio-economic impacts of research. SPIAS has been used to analyse leading Japanese scientists' performance before and after receiving grants from the Japan Science and Technology Agency; assess the impact of regenerative medicine research in Japan; and map emerging technologies.
>
> In Spain, Corpus Viewer, developed by the State Secretariat for Information Society and Digital Agenda, processes and analyses large volumes of textual information using natural-language processing techniques. Policy makers use the results of these analyses to monitor and evaluate public programmes, and formulate science and innovation policy initiatives. The system is currently restricted to government officials.

The promises of DSIP

Governments are increasingly launching DSIP initiatives, often with the following objectives:

- *Optimise administrative workflows*: digital tools can help streamline potentially burdensome administrative procedures and deliver significant efficiency gains within agencies. These benefits can also extend to those using public agencies' services, including researchers or organisations applying for (or reporting on) the

use of research grants; for example, they can use interoperability identifiers to link their research profiles to grant applications.

- *Support better policy formulation and design*: digitalisation offers new opportunities for more granular and timely data analysis to support STI policy; this should improve the allocation of research and innovation funding. Furthermore, DSIP systems often link data collected by different agencies, providing greater context to policy problems and interventions, and offering possibilities for a more integrated interagency policy design at the research or innovation system level.

- *Support performance monitoring and management*: DSIP systems offer the possibility of collating real-time policy output data. This can allow more agile short-term policy adjustments. It can improve insights into the policy process for accountability and learning in the medium to long term, so that evaluation becomes an open and continuous process. Policy makers and delivery agencies can consider the circumstances that make it possible and meaningful to use other digitally enabled data resources, such as altmetrics of research outputs and impacts (Priem et al., 2010; Sugimoto and Larivière, 2016). They can also rely on other data-collection approaches (e.g. web scraping) to complement and enhance existing approaches to assessing research.

- *Provide anticipatory intelligence*: technologies like big-data analytics can help detect patterns, e.g. emerging research areas, technologies, industries and policy issues. They can support short-term forecasting of policy issues and contribute to strategic policy planning (Peng et al., 2017; Choi et al., 2011; Zhang et al., 2016; Yoo and Won, 2018). For example, DSIP systems could identify job-market demand for specific STI fields and address potential mismatches on the supply side.

- *Help in general information discovery*: DSIP systems often include data on a wide range of inputs, outputs and activities. Policy makers and funders can use these data to identify leading experts in a given field (e.g. identify reviewers for project proposals), as well as centres of excellence (Sateli et al., 2016; Guo et al., 2012). This kind of information also helps researchers and entrepreneurs to identify new partners for collaboration and commercialisation.

- *Promote inclusiveness in science and innovation agenda-setting*: DSIP systems can contribute to the debate with stakeholders on policy options by providing detailed information about the policy problem in an accessible medium, e.g. through interactive data visualisation. The increased transparency provided by DSIP systems can empower citizens by providing them with knowledge about the nature and impacts of ongoing research and innovation. Thus, DSIP may be instrumental in building trust and securing long-term sustainable funding for research and innovation.

Fulfilling these promises will depend on policy makers' readiness to embrace the digital revolution (Box 12.2). It will also depend on meeting several challenges, discussed in the following section.

Box 12.2. In my view: Are science policy makers ready to embrace the digital revolution?

Clinton Watson, Principal Policy Advisor, New Zealand Ministry of Business, Innovation and Employment

Policy makers in science and innovation are charged with designing and overseeing funding mechanisms that funnel billions of dollars of public money into universities, PROs, businesses and not-for-profit entities. Yet despite the huge investments, the science-of-science policy has received almost no funding. Oftentimes, policy makers struggle to demonstrate real societal impacts from investments. Arguably, they have paid more attention to ensuring science systems continue to receive adequate funding and respond to domestic demands. Assessing and demonstrating performance has often played a secondary role to setting high-level objectives and getting money out the door.

Politicians and senior public-sector leaders are increasingly demanding hard evidence of what works and what does not. Science and innovation-related spending is no longer exempt from pressures to provide quantitative evidence of impact. In some countries, the storyline of good science delivering societal outcomes many years down the track is wearing thin. At the same time, hard-to-answer questions on optimal institutional settings, design of funding pots and efficient allocation systems persist. Policy makers need to focus more on supporting monitoring systems, evaluation frameworks and data infrastructures, working with the very researchers and academics they fund.

Advances in information technologies and data-linking techniques are now presenting policy makers with the tools to start answering the hard questions. A handful of countries have developed national research information systems that harvest data from multiple sources. If these systems can be linked to other national data infrastructures (e.g. housing economic, environmental and social data), science policy makers will be in a unique position to demonstrate quantitative relationships between science and innovation, and real-world outcomes. Researchers could also use these linked data infrastructures to prove, for example, that firms collaborating with universities become more productive, or that certain types of research lead to improved environmental outcomes over time. They could also produce useful descriptive statistics, such as the value and growth of spin-out companies.

For several years, I led efforts in New Zealand to improve data holdings on research, science and innovation. Through collaboration between government agencies and key sector bodies, we identified the enduring questions to answer, our data needs, our current data holdings and a high-level roadmap for action. Key challenges were securing trust in data use, developing communication channels within institutions and identifying best practice globally. Implementation has centred on securing sustainable funding, providing detailed communication of benefits, and establishing legal and governance frameworks.

The New Zealand experience and other similar initiatives all point to the social and cultural challenges in building data infrastructures for science and innovation policy. The idea of "social licence", or community acceptance and trust of data use, is in the spotlight. Institutions and researchers need to have assurance that data about their funding, activities and results will be handled appropriately and protected when needed. Many universities and research organisations are also not used to automatic data transfer to a central hub. The funding of national level systems also presents challenges. The optimal cost sharing
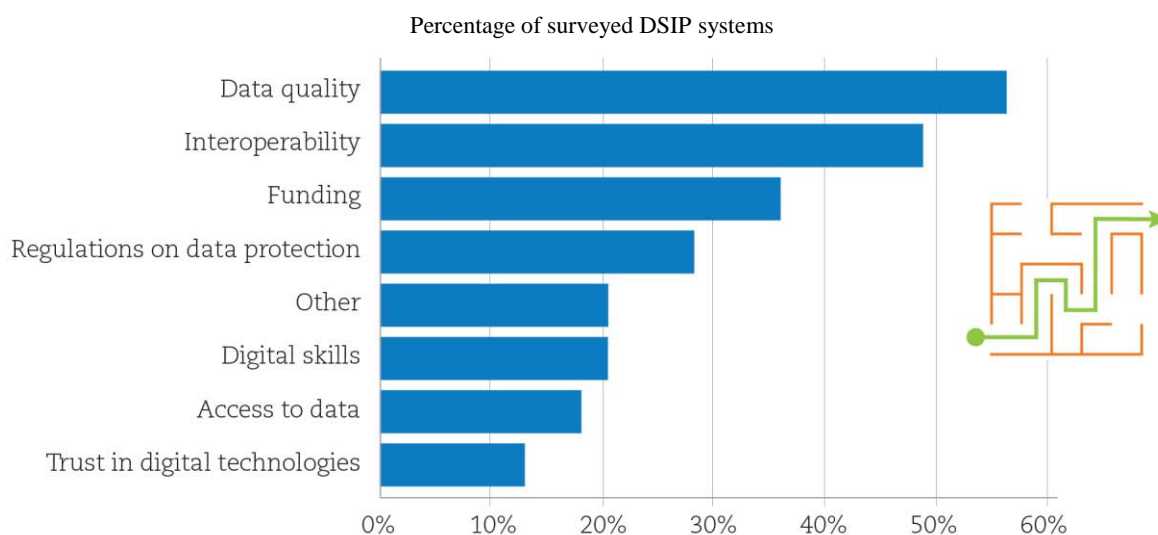
between the central research and innovation ministries, science funders and research providers will differ depending on institutional responsibilities and funding flows.

Science policy cannot afford to be immune to the digital transformation we are witnessing across economies. We need to embrace digitalisation if we are to prove the ongoing worth of science and innovation, and raise the effectiveness of public spending. Policy makers need to support digital tools and their social licence, creating long-term plans for establishing linked data infrastructures, establishing effective governance and funding structures, and building capacity for the science-of-science policy.

Main policy challenges

Realising the potential of DSIP involves overcoming several possible barriers. In their responses to the OECD questionnaire, DSIP administrators identified data quality, interoperability, sustainable funding and data-protection regulations as the biggest challenges facing their initiatives (Figure 12.2). Access to data, the availability of digital skills and trust in digital technologies were somewhat less often cited as challenges.

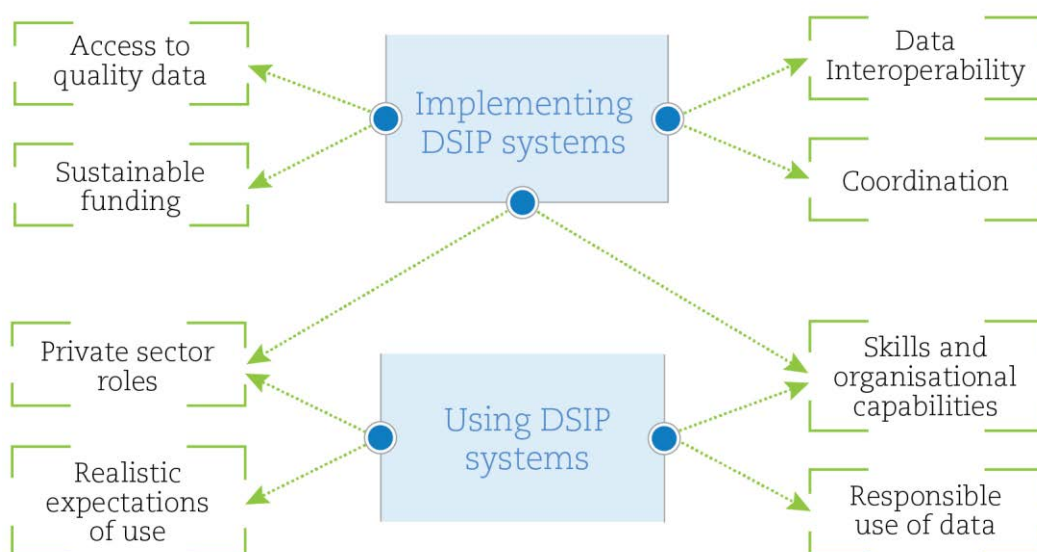**Figure 12.2. Main challenges facing DSIP initiatives**

Percentage of surveyed DSIP systems

Note: Questionnaire respondents could select more than one challenge facing their DSIP initiatives.
Source: OECD survey of administrators of 39 DSIP systems in OECD member countries and partner economies.
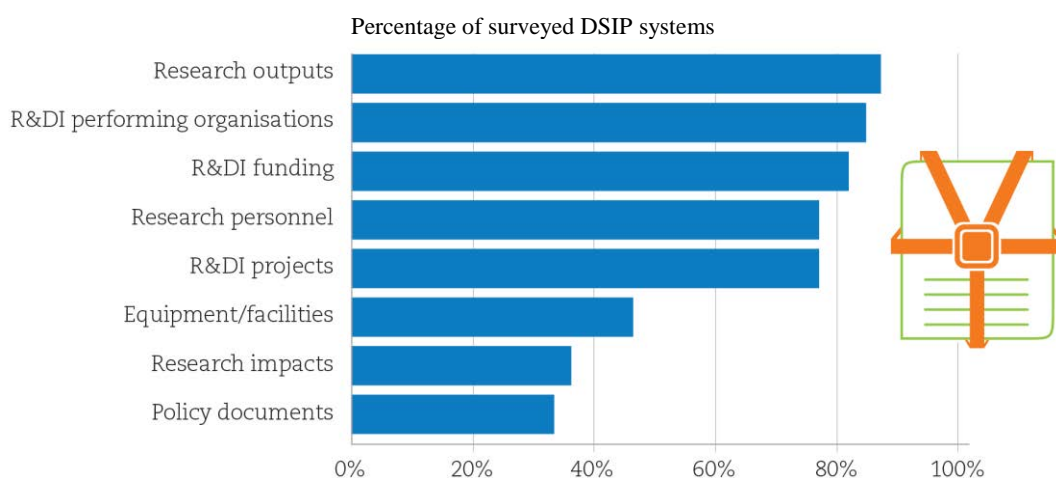
StatLink ⟋⟍ https://doi.org/10.1787/888933858335

Policy makers wishing to promote DSIP in their countries face further systemic challenges, including overseeing fragmented DSIP efforts and multiple (often weakly co-ordinated) initiatives; ensuring the responsible use of data generated for other purposes; and balancing the benefits and risks of private-sector involvement in providing DSIP data, components and services. Figure 12.3 summarises and organises the main challenges in implementing or using DSIP systems. The section below elaborates on each challenge.

Figure 12.3. Challenges in implementing and using DSIP systems

Access to quality data

Most DSIP systems draw upon different data sources to provide new insights that cannot be obtained through working with each data source separately. For example, they link data on inputs and outputs to provide insights on the impacts and efficiency of public research funding. Most of the DSIP systems surveyed incorporate data on research outputs (typically academic publications), research organisations, research funding (i.e. project and grant awards), research personnel and research projects (Figure 12.4). Some DSIP systems include data on research equipment and facilities, as well as research impacts (including citations and media mentions).

Figure 12.4. Types of information harnessed for DSIP systems

Percentage of surveyed DSIP systems

Note: Questionnaire respondents could select more than one type of information harnessed by their DSIP initiatives.
Source: OECD survey of administrators of 39 DSIP systems in OECD member countries and partner economies.

StatLink 🔗 https://doi.org/10.1787/888933858354

While data reusability is a major source of efficiency promised by DSIP systems ("enter once, re-use often"), respondents to the OECD survey of DSIP administrators cited data quality as a major challenge (Figure 12.2). Data quality is a multi-dimensional concept, encompassing relevance, accuracy, credibility, timeliness, accessibility, interpretability, coherence and cost efficiency (OECD, 2011). It ultimately defines whether data serve a given purpose. Data used in DSIP systems may have been generated for different or related purposes, meaning that users must assess quality factors for each intended application. Data are predominantly sourced from a mix of funding agencies (typically their administrative data, e.g. databases of grant awards) and research performers (e.g. university CRIS), as well as proprietary bibliometric and patent databases. However, available data may not capture precisely what is needed for the DSIP system (need for relevance/interpretability); alternatively, they may be presented in an unstructured format that is complicated to process (need for accessibility/coherence). Fixing this may require further complementary resources, including additional metadata; algorithms for data processing; and secure digital infrastructures for (shared) data storage, processing and access. The costs involved may discourage more widespread data sharing, particularly when its benefits are not always obvious to those providing the data (OECD, 2017).

Other potential barriers exist to open-data sharing. These include bureaucratic competition and conflicting interests among government organisations and individual departments, and notions that any value to be extracted from administrative data should be initially – and primarily – the preserve of the data owners. Several systems provide tiered access to their data, whereby policy officials inside the host organisation can access more granular data. A lack of trust in the manner in which shared data will be used may also hinder sharing. For example, organisations may be legitimately concerned that their data will be misused or poorly interpreted by users with an inadequate understanding of its meaning and limitations. As semi-autonomous agents, organisations may also fear the unwanted scrutiny of their operations that open administrative data might invite. Privacy and confidentiality are also major concerns when re-using data collected for other purposes (Lane et al., 2015.
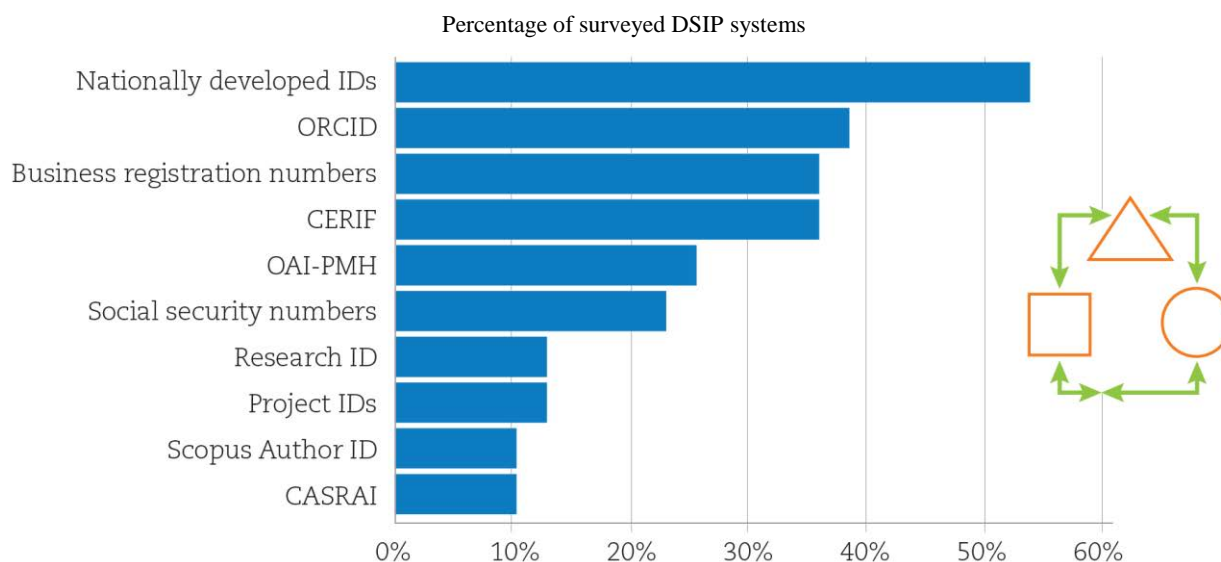
Data interoperability

The databases used in DSIP systems have often been locally designed, without adherence to common standards. Hence, many of the data relevant to STI policy are stored in inaccessible silos, complicating data re-use. Ensuring data compatibility is not only potentially beneficial to policy makers and other stakeholders managing national research and innovation systems, it can yield considerable benefits for individuals and organisations doing (or reporting on) research. If an individual data item is made interoperable, it can be re-used across multiple systems, meaning it can be provided to authorities only once. Interoperability also allows diffusing updates across systems more easily and automatically comparing information from multiple sources (e.g. checking the consistency of project-funding reports submitted by researchers and employers). An integrated and interoperable system leads to a considerable reduction in the reporting and compliance burden, freeing up more time and money for research itself.

Research organisations, funders and non-profit organisations have started designing standards, vocabularies and protocols that connect and disambiguate research data and metadata to improve interoperability between silos. Some DSIP systems use existing national identifications (IDs) – e.g. business registration and social security numbers – as well country-specific IDs for researchers (Figure 12.5). In recent years, attempts have been made to establish international standards and vocabularies to improve the international interoperability of DSIP infrastructures. These include UPPIs, which assign a standardised

code that is unique to each research entity, persistent over time and pervasive across various datasets. One example is Open Researcher and Contributor ID (ORCID), which aims to resolve name ambiguity in scientific research by developing unique identifiers for individual researchers. Figure 12.1 sets out several prominent examples.

Figure 12.5. Use of interoperability enablers in DSIP systems

Percentage of surveyed DSIP systems

Note: Questionnaire respondents could select more than one type of interoperability enabler used in their DSIP initiatives.
Source: OECD survey of administrators of 39 DSIP systems in OECD member countries and partner economies

StatLink 🔗 https://doi.org/10.1787/888933858373

Table 12.1. Examples of interoperability enablers in DSIP and related systems

Type	Examples
UPPIs for STI actors	ORCID Digital Object Identifier (DOI) Global Research Identifier Database (GRID) International Standard Name Identifier (ISNI) Ringgold ID
Author IDs generated by publishers/indexers	Researcher ID Scopus Author ID
Management standards for data about STI	Common European Research Information Format (CERIF) Consortia Advancing Standards in Research Administration Information (CASRAI) Dictionary VIVO ontology
Protocols	Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

In addition to the reduced administrative burden, interoperability allows quicker, cheaper and more accurate data matching, making existing analyses less costly and more robust, and facilitating new analyses. Interoperability can produce more timely and detailed insights, enabling more responsive and tailored policy design. Furthermore, the gradual emergence of internationally recognised UPPIs makes it easier to track the impacts of research and innovation activities across borders, and map international partnerships.

As Figure 12.5 shows, several DSIP systems also use the CERIF data-management standard to promote uniform management and exchange of research information.[1] A smaller number use the standard dictionary of research administration developed by CASRAI. Semantic ontologies can also help improve interoperability among DSIP systems. For example, the VIVO project has developed an ontology for research information that enables a federated search of organisations, researchers and activities, and their relationships. However, because few DSIP systems currently use semantic technologies, none of the systems included in the OECD survey have deployed it yet.

Thus, although identifiers, standards and protocols have proliferated, interoperability remains a major challenge (Figure 12.2). In the absence of national strategies to ensure the architectural coherence and interoperability of public databases, there is a risk that data sources will become fragmented, undermining the functionality of DSIP systems and raising the costs of data integration.

Domestic and international co-ordination

As individual ministries and agencies in the policy system increasingly rely on digital tools to exploit their administrative data, DSIP initiatives are proliferating. Some are ambitious in scope and seek to become national data hubs; many are tentative experiments, limited in scope and with little cross-departmental co-ordination. They lack the mandate (and resources) to expand, and prefer to remain manageable and modest.

This fragmented landscape presents likely drawbacks, including inefficient, overlapping efforts; missed opportunities to establish interoperability standards that would improve data exchange; and greater funding uncertainties. However, simultaneously running multiple small-scale experiments may present benefits, by providing more space for innovation and promoting a more agile DSIP ecosystem. Moreover, if these distributed initiatives adopted common data-management frameworks (including interoperability standards), then a modular landscape of interconnected DSIP initiatives could emerge. This interconnectedness could have some advantages over a single dominant system, by achieving greater service scalability and flexibility, as well as more seamless integration with external digital infrastructures. It could also support data integration from a wider range of ministries and agencies funding R&D, and thus provide a more complete picture of national policy and funding for research and innovation.

For these benefits to be realised, however, there needs to be co-ordination of data-management frameworks. Most countries already have ministries and agencies that formulate high-level national digital strategies, establish technological architectures and promote good practices in public-data management. However, although these measures can provide the necessary conditions for DSIP to flourish, they are insufficient on their own. For instance, specificities to the STI domain around data sources, interoperability standards and the intended (and unintended) uses of DSIP data require further – but still underdeveloped – co-ordination and support at the STI policy level. Most countries still lack dedicated plans for DSIP, and only a few (e.g. Norway and New Zealand) have appointed lead agencies to formulate and co-ordinate common frameworks for STI policy-related data management. In the absence of national co-ordination mechanisms, the wide adoption of international (or private-sector) interoperability standards could provide an "invisible hand" for co-ordination, but DSIP owners would still need to prepare and agree to share their own administrative data. Thus, beyond finding technical solutions for system interoperability, DSIP is ruled by political considerations and compromise.

Sustainable funding

Most of the DSIP systems surveyed are funded by their host organisations' operating budgets, which could be a positive sign of their long-term survival. However, many DSIP systems are relatively new, and it is difficult to estimate their sustainability. More than one-third of the surveyed DSIP administrators pointed to funding as a challenge for their systems, the third highest ranked challenge after data quality and interoperability (Figure 12.2). The individuals charged with building and maintaining DSIP systems often underestimate the magnitude of the task, particularly with respect to data access, disambiguation and linking; this can lead to significant project delays and cost overruns. As with any infrastructure, maintenance and use costs may also be higher than the initial investment costs.

Skills and organisational capabilities

A distinction should be made between the skills and organisational capabilities needed to use DSIP systems in policy making and analysis, and those needed to build and maintain digital infrastructures, which DSIP administrators ranked as a low-level challenge (Figure 12.2) – perhaps because many systems use well-established digital tools and techniques easily mastered by existing digital teams. When faced with more challenging problems, the individuals implementing DSIP infrastructures can readily buy the necessary technical expertise on the market.

A few DSIP initiatives are experimenting with more advanced digital tools, such as semantic technologies to link datasets, algorithms to support big-data analytics, and interactive visualisations and dashboards to promote data use in the policy process. For example, Spain's Corpus Viewer uses natural-language processing techniques to process and analyse large volumes of textual data on Spanish research funding (Box 13.1). In the United Kingdom, the Arloesiadur project[2] – a partnership between the Welsh government and NESTA, with inputs from a company specialising in data visualisation – combines traditional indicators with data from social networks, company websites and collaboration platforms, to provide interactive visualisations of research and innovation networks in Wales (Mateos-Garcia et al., 2017). Together with IBM, the long-established Flanders Research Information Space is exploring ways to use web scraping to capture Flemish research outputs scattered across the web.

To date, the public sector has rarely used advanced digital tools in its DSIP initiatives. This reluctance may stem from the costs of hiring digital-technology professionals with expertise in big data, machine learning and natural-language processing, which can be prohibitive, given competition from the private sector for these skills. It may also reflect policy data needs, which remain quite straightforward (c.f. many policy makers eschew advanced econometric studies in favour of simpler indicators). Approaches such as semantic analysis tend to be quite technical; interpreting the information they provide requires certain skills.

Considering the skills and organisational capabilities needed to utilise the data and functions of DSIP systems, the STI policy-making community is increasingly attracting quantitatively literate officials with backgrounds in various analytical disciplines (Chapter 14 on next-generation data and indicators). A striking number of DSIP initiatives seem to target this specialised audience of analysts: many DSIP users are evaluators and analysts who act as intermediaries, processing the data before feeding it to decision-makers. This situation could change in the future thanks to advances in visual analytics, which could

open up DSIP systems to a wider range of non-analyst users, both in government and beyond.

At the level of the policy organisation, a mix of capabilities is required, including technical staff with specialised skills in data curation and stewardship, to manage the use of necessary standards and metadata. Policy analysts and decision-makers would find it useful to possess statistical skills, i.e. knowledge of key concepts and statistical software. Existing staff can accumulate some of these capabilities gradually, by upskilling through massive open online courses; this is a more cost-effective option than hiring expensive data scientists. In this way, DSIP initiatives could benefit from a process of cumulative organisational learning and deploy increasingly ambitious technologies.

Private sector roles

The private sector plays an increasingly important role in DSIP systems. For example, various academic publishers, web service companies and data-management systems provide access to proprietary databases, digital analytical tools and unique identifiers. Beyond the simple provision of services, these relations encompass different levels of public-private co-operation, such as joint development of methods and tools to analyse research impact, and collaboration on the design and implementation of digital platforms for policy-making purposes.

Three companies with long-standing ties with the academic research community – Elsevier (the world's largest academic publisher and owner of the Scopus index), Holtzbrinck Publishing Group (owner of Springer Nature and Digital Science) and Clarivate Analytics (formerly part of Thomson Reuters and owner of the Web of Science index) – are developing digital solutions for workflow management and research analytics on top of in-house databases of research outputs. By acquiring and developing digital tools that complement their product portfolios, and building interoperability linkages between in-house and external solutions, they are creating digital platforms of interconnected digital products with similar functionalities to publicly owned DSIP systems. They are using machine learning, natural-language processing and big-data analytics to exploit in-house databases. They are also designing new add-on analytical services to monitor and assess research and innovation activities.

Although they do not formally provide DSIP solutions to governments, large information and communication technology firms provide some of the building blocks for DSIP. Some, like Google Scholar, Microsoft Academic, Baidu Scholar and Naver Academic, have already transformed scientific and technical discovery with their search engines. DSIP systems only minimally rely on these solutions, but this could change as they become more sophisticated (e.g. by deploying artificial intelligence and semantic-search tools) and provide wider coverage of research outputs.

Private-sector involvement in DSIP initiatives offers several benefits. Private firms can often provide off-the-shelf, well-developed solutions and building blocks for DSIP. These can be implemented quickly and at an agreed cost, sparing the public sector the need to develop the necessary in-house skills beforehand. As highlighted earlier, private companies can also promote interoperability through their standards and products; moreover, the largest firms operate across national borders, and can therefore promote international interoperability. This can expand the scope and scale of data within a DSIP system utilising these products and standards; for example, policy makers can compare the features of their own research systems with others. At the same time, governments often expect their open public data to spur innovation (e.g. new products and services) in the private sector.

Potential risks also exist when the public sector relies on the private sector for DSIP systems and components. For example, outsourcing data-management activities to the private sector may result in a loss of control over the future development trajectory of DSIP systems; reliance on proprietary products and services may lead to discriminatory access to data, even if these concern research activities funded by the public sector; and the public sector's adoption of commercial standards for metrics may drive the emergence of private platforms exhibiting network effects that are difficult to contest. Furthermore, while methods and algorithms are sources of competitive advantages, the secrecy surrounding them can undermine trust in such systems, particularly when they are used to assess research performance.

Realistic expectations of use

By re-using and combining data from a variety of sources, DSIP can provide policy makers with a broader view of the research and innovation landscape, and consequently furnish evidence to help them allocate funding. However, expectations around the uses of DSIP should avoid a "naïve rationalism" that ignores the inherent messiness of policy making. DSIP can inform policy judgement, but it cannot and should not provide a "technical fix" to what are ultimately political judgements, shaped by competing values and uncertainty. If they were "open by design", DSIP systems could promote inclusiveness in science and innovation agenda-setting, making it less technocratic and more democratic. Whatever the policy setting, an embedded and routine use of DSIP will depend not just on digital technologies, but also on favourable social and administrative conditions promoting their adoption.

Responsible use of data

Private and confidential data make up a considerable portion of the data processed by the public sector, and can be potentially useful in DSIP systems. However, these data must be used responsibly (OECD, 2013). This often means anonymising or aggregating them – e.g. when the identity of individual companies would become apparent in more granular data. More than one-quarter of the DSIP administrators surveyed highlighted data-protection regulations as a challenge (Figure 12.2).

More than half of the DSIP systems surveyed play a role in research assessment. Some, like the Cristin system in Norway, the Lattes Platform in Brazil, and the METIS system in the Netherlands, are the primary sources of data for national research assessments. However, some evidence exists that non-policy actors are also using DSIP data – e.g. to assess the performance of individuals – raising concerns over the responsible use of linked open data generated for other purposes. DSIP could reinforce some existing misuses of data (e.g. reliance on journal-impact factors in various types of assessments), which could further distort the incentives and behaviour of individuals and organisations (Edwards and Siddhartha, 2017; Hicks et al., 2015).

Over-reliance on data is dangerous when its interpretation is problematic – hence the need to improve our understanding of STI processes, to make sense of the data. DSIP systems offer a great opportunity to develop such an understanding, as the data can be made available to a broad community of researchers, who could further develop the emerging field of science-of-science and innovation policy (Lane, 2010; Husbands Fealing et al., 2011). AI-based tools can also be mobilised to promote such an understanding. Barring that, DSIP systems will simply result in even more data being interpreted – and often misinterpreted – in many ways.

Although few surveyed DSIP administrators reported a lack of trust in digital technologies as a major challenge (Figure 12.2), this could change with the introduction of newer and more advanced technologies and processes, e.g. machine learning and big-data analytics. These technologies rely on notoriously opaque algorithms, which could undermine trust in DSIP-based solutions. The use of data sources with questionable provenance – e.g. data derived through web scraping of company websites – is another potential source of mistrust, which should be treated with care.

Future outlook

Digital content and processes will play an important role in the future policy design, operational delivery and governance arrangements of research and innovation. Governments cannot continue to work in analogue mode when society and the economy are increasingly working in digital mode. The rapid and broad uptake of digital technologies and data across the public sector will place increasing pressure on governments to rethink the management of core policy processes and activities, including with regard to STI policy.

The digital transformation of STI policy and its evidence base is still in its early stages. As digitalisation becomes increasingly pervasive, uncertainties remain as to what it will cover, who will take the lead, and what roles existing actors (including NSOs, as data clearinghouses for statistical purposes) will play. The consequences on the relations (including governance arrangements) between STI actors are also uncertain. Moreover, international co-operation could take different forms and perform different functions in the future DSIP landscape.

STI policy makers could assume a relatively passive stance in the face of these developments: their activities – including the evidence base they use to inform their decision-making – will inevitably become increasingly digitalised. Alternatively, they could adopt a more active stance, shaping the DSIP ecosystems to fit their needs. This will require strategic co-operation, through significant interagency co-ordination and sharing of resources (such as standard digital identifiers), and a coherent policy framework for data sharing and re-use in the public sector. Since several government ministries and agencies formulate science and innovation policy, DSIP ecosystems should be founded on the principles of co-design, co-creation and co-governance.

In a desirable future scenario, DSIP infrastructures will provide multiple actors in STI systems with up-to-date linked microdata to help inform their decision-making. They will erode information asymmetries, and empower a broad group of stakeholders to participate more actively in the formulation and delivery of science and innovation policy. Policy frameworks will have resolved privacy and security concerns, and national and international co-operation on metadata standards will have addressed interoperability issues. Best practices in the responsible use of DSIP systems will have taken hold, informed by widely accepted norms of acceptable use. While the private sector will provide supporting infrastructures and services, the public sector will own its data, ensuring they remain outside of "walled gardens", for others to readily access and re-use.

Considerable scope also exists for international mutual learning and co-operation in developing digital data infrastructures for STI policy. Given the global nature of science and innovation activities, there could be particular benefits in establishing further international standards – including strengthening existing OECD legal and informal guidance instruments – to take full stock of the potential and challenges of DSIP.

Notes

[1] Administrative standards relate closely to, but do not fully align with, OECD statistical standards (e.g. Frascati Manual [OECD, 2015b]) or definitions contained in OECD legal instruments. As noted in Chapter 14 (on next generation data and indicators), it is important to ensure closer correspondence between these different international standards and the standards used by countries and supranational organisations, such as the European Union.

[2] "Innovation Directory" in Welsh.

References

Choi, S. et al. (2011), "SAO network analysis of patents for technology trends identification: A case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells", *Scientometrics*, Vol. 88/3, pp. 863-883, Springer International Publishing, Cham, Switzerland, https://doi.org/10.1007/s11192-011-0420-z.

Dai, Q., E. Shin and C. Smith (2018), "Open and inclusive collaboration in science: A framework", *OECD Science, Technology and Industry Working Papers*, No. 2018/07, OECD Publishing, Paris, https://doi.org/10.1787/2dbff737-en.

Dijk, E. et al. (2006), "NARCIS: The Gateway to Dutch Scientific Information", in *Digital Spectrum: Integrating Technology and Culture – Supplement to the Proceedings of the 10th International Conference on Electronic Publishing*, pp. 49-58, ELPUB, Bansko, Bulgaria, https://elpub.architexturez.net/doc/oai-elpub-id-233-elpub2006.

Edwards, M. and R. Siddhartha (2017), "Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition", *Environmental Engineering Science*, Vol. 34/1, pp. 51-61, Mary Ann Liebert, Inc. Publishers, New Rochelle, NY, http://online.liebertpub.com/doi/pdf/10.1089/ees.2016.0223.

Guo, Y et al. (2012), "Text mining of information resources to inform forecasting innovation pathways", *Technology Analysis & Strategic Management*, Vol. 24/8, pp. 843-861, Routledge, London, https://doi.org/10.1080/09537325.2012.715491.

Hicks, D. et al (2015), "Bibliometrics: The Leiden Manifesto for research metrics", *Nature*, Vol. 520/7548, pp. 429-31, Macmillan Publishers, London, https://doi.org/10.1038/520429a.

Husbands-Fealing, K. et al. (eds.) (2011), *The Science of Science Policy: A Handbook*, Stanford Business Books, Stanford University Press, Stanford, CA, http://www.sup.org/books/title/?id=18746.

Lane, J. et al. (2015), "New Linked Data on Research Investments: Scientific Workforce, Productivity, and Public Value", *NBER Working Paper*, No. 20683, National Bureau of Economic Research, Cambridge, MA, http://www.nber.org/papers/w20683.

Lane, J. (2010), "Let's make science metrics more scientific", *Nature*, Vol. 464, pp. 488-489, Macmillan Publishers, London, https://doi.org/doi:10.1038/464488a.

Mateos-Garcia, J., K. Stathoulopoulos and S. Bashir Mohamed (2017), "An (increasingly) visible college: Mapping and strengthening research and innovation networks with open data", *SocArXiv Papers*, University of Maryland, College Park, MD, https://doi.org/10.17605/OSF.IO/3CU67.

OECD (2018), "Going Digital in a Multilateral World", Interim Report of the OECD Going Digital Project, Meeting of the OECD Council at Ministerial Level, Paris, 30-31 May 2018, OECD, Paris, http://www.oecd.org/going-digital/C-MIN-2018-6-EN.pdf.

OECD (2017), "Key Issues for Digital Transformation in the G20, report prepared for a joint G20 German Presidency/OECD conference, Berlin, 12 January 2017, OECD, Paris.

OECD (2015a), "Making Open Science a Reality", *Science, Technology and Industry Policy Papers*, No. 25, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jrs2f963zs1-en.

OECD (2015b), *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris, https://doi.org/10.1787/9789264239012-en.

OECD (2014), "Recommendation of the Council on Digital Government Strategies", Adopted by the OECD Council on 15 July 2014, OECD, Paris, http://www.oecd.org/gov/digital-government/Recommendation-digital-government-strategies.pdf.

OECD (2013), *The OECD Privacy Framework*, OECD, Paris, http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf.

OECD (2011), "Quality Dimensions, Core Values for OECD Statistics and Procedures for Planning and Evaluating Statistical Activities", OECD, Paris, http://www.oecd.org/sdd/21687665.pdf.

Peng, H. et al. (2017), "Forecasting potential sensor applications of triboelectric nanogenerators through tech mining", *Nano energy*, Vol. 35, pp. 358-369, Elsevier, Amsterdam, https://doi.org/10.1016/j.nanoen.2017.04.006.

Priem, J. et al. (2010), Altmetrics: A manifesto, 26 October 2010, http://altmetrics.org/manifesto (accessed 5 February 2017).

Sateli, B et al. (2016), "Semantic User Profiles: Learning Scholars' Competences by Analyzing Their Publications", in González-Beltrán A., F. Osborne and S. Peroni (eds.), *Semantics, Analytics, Visualization. Enhancing Scholarly Data,* SAVE-SD 2016, Lecture Notes in Computer Science, Vol. 9792, Springer, Cham, Switzerland, https://doi.org/10.1007/978-3-319-53637-8_12.

Sugimoto, C. and V. Larivière (2016), "Social media indicators as indicators of broader impact", Presentation given at OECD Blue Sky Forum, Ghent, www.slideshare.net/innovationoecd/sugimoto-social-media-metrics-as-indicators-of-broader-impact.

Ubaldi, B. (2013), "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives", *OECD Working Papers on Public Governance*, No. 22, OECD Publishing, Paris, https://doi.org/10.1787/5k46bj4f03s7-en.

Yoo, S.H. and D. Won (2018), "Simulation of Weak Signals of Nanotechnology Innovation in Complex System", *Sustainability*, Vol. 10/2/486, MDPI, Basel, https://doi.org/10.3390/su10020486.

Zhang, Y. et al. (2016), "Technology roadmapping for competitive technical intelligence", *Technological Forecasting and Social Change*, Vol. 110, pp. 175-186, Elsevier, Amsterdam, https://doi.org/10.1016/j.techfore.2015.11.029.

From:

OECD Science, Technology and Innovation Outlook 2018

Adapting to Technological and Societal Disruption

Access the complete publication at:
https://doi.org/10.1787/sti_in_outlook-2018-en