AUTHOR:
Cobus Jooste[1]

AFFILIATION:
[1]Consolidoc Postdoctoral Fellow, Department of Mercantile Law, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Cobus Jooste

EMAIL:
cobusjooste@gmail.com

HOW TO CITE:
Jooste C. Driving openness – the myths about data and software access in the Data and Cloud Policy. S Afr J Sci. 2022;118(1/2), Art. #12124. https://doi.org/10.17159/sajs.2022/12124

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
law, data protection, policy, data, software, intellectual property, copyright

PUBLISHED:
27 January 2022

Driving openness – the myths about data and software access in the Data and Cloud Policy

Significance:

- This article responds to the highly controversial recently published Draft National Policy on Data and Cloud and the related stakeholder commentary from a cyberlaw perspective. Selected issues regarding the ownership of scientific data, databases and processing software, within the intellectual property law framework, are summarised. The law on issues regarding computer science and software engineering, specifically decompilation and digital rights management, are discussed insofar as these matters are implicated in the Data and Cloud Policy. A call to conduct a rights audit is issued to those whose scientific research activity relies on controlled data access and/or those who have a proprietary interest in the economic value of data.

Introduction

There is no doubt that *access* to information is fundamental to scientific, cultural and economic progress, but it is not, as some would have it, a recent symptom of the Fourth Industrial Revolution.[1,2] The opportunity to rely on the work of others has been a cornerstone of legal development since, at least, as early as 1710 when the first Copyright Act in the world was published. Since then, a substantial body of intellectual property and neighbouring laws has developed to address the tension between the public interest in access to information and the private interests in controlling the dissemination of their work or the use of their ideas.[3]

Therefore, it is worrying to note that, when dealing with the two most valuable manifestations of knowledge (namely data and software), the South African government seems unaware of the existence of private rights. In the recently published Draft National Policy on Data and Cloud (the Data Policy), the Ministry of Communications and Digital Technologies states 'it remains unclear how data generated through intellectual activities of varying degrees and types would be correctly categorised in terms of intellectual property rights'[1] and suggests, incorrectly, that there is no policy applicable to the ownership of data.[1]

Labouring under this misapprehension, the Data Policy proposes to provide universal access to data by, inter alia, consolidating all publicly funded data centres into a single, state-owned entity, nationalising all data generated in the country and establishing an open data platform to control access to all the data held by this entity.[1] The Data Policy's primary goal is to concentrate power over data in the state and seeks to localise data processing by creating widespread reliance on state-owned data and cloud infrastructure.

The Data Policy is indeed an alarming document and has met with substantial opposition. For the purpose of this article, the focus is on selected issues regarding the future of private ownership of software and data and the ability to control access thereto. This is done by addressing a number of myths espoused in the Data Policy, from a cyberlaw perspective.

Myth 1: Data is inherently valuable

The Data Policy waxes lyrical about data as the 'new oil' and declares data to be 'the infrastructure for the digital economy' because 'the greatest advantage of data is the value it generates after it is processed into information and knowledge.'[1] This view is problematic on two fronts. First, it implies that data need only be processed in order to extract economic value. Second, it implies that those who hold data collections are in a position to exploit the data at will. In other words, the Data Policy relies on the myth that data is a mineable commodity with inherent potential value.

Value is in the application of knowledge

In reality, data are simply digitally represented information and have no more inherent value than the contents of a dictionary or a list of phone numbers. The value of data per se is nil, and so too is the *potential* value of data. The value of data lies in the *application* of knowledge obtained from processed data – often described as a data value chain.[4] This is the reason why data are used as a commodity. Those who invest in data determine the value according to their possible application and the commercial benefit that may be derived from using the knowledge they have gained.[2]

Furthermore, data cannot be imbued with inherent value. Even when great skill and advanced analytical methodology are applied to data, the value is created in, and derived from, the results of those efforts. In these cases, the commercial or other value of the data exists only because of the results facilitated.[2] Consider, for example, two very different databases: a database containing the shopping habits of customers and a database containing the blood glucose level of patients. In isolation, each database may be used to create value through application. The first may aid advertisers in their targeted marketing efforts, assist retailers to make informed decisions about stock purchases or product placement, or help manufacturers to decide on product design or factory retooling. In all of these cases, the economic value lies in the cost savings or increase in turnover made possible as a result of the knowledge gained from the processed and interpreted data. The data itself, even after processing, are not responsible for the value. The second database is the same. Consisting of protected and sensitive information, its utility is restricted but it may be useful to medical research for any number of applications. In these cases,

the value depends on the results of the research. It may be directly economic, such as subsidy for published research output, or indirect, in the form of attracting future grants. Once again, the value is derived from the knowledge that was created from the data, not from the knowledge contained in the data.

Into this picture one may add so-called big data – the massification and concentration of data to extract value from the combination of a variety of data points. Thus, for example, by combining the two databases, one may determine a possible correlation between eating habits and blood glucose level. This statistic might be useful to medical aid providers when designing healthy living incentives or to pharmaceutical manufacturers or foodstuff producers when making decisions on product development. In all these cases, the value of the data is second hand – it is derived from the amendments made to current business strategy.

The law is purpose and outcome focused

Once this is understood, the fallacy in the Data Policy's approach is laid bare. From a cyberlaw perspective, data are treated as information – they have no inherent value. For this reason, the range of legal measures applicable to data depend on what the data may be (or may not be) used for or, in other cases, how the investment in processing the data, or creating knowledge from the data, should be repaid. The law recognises that the value of data is downstream, as are the threats to society.[5] As scholars have pointed out when considering the status of intellectual property, 'the constitutional conception of property, according to which the focus falls on the function that the alleged property has in society rather than the traditional, pre-constitutional conceptions of property'[6] is the basis for the manner in which the law, specifically intellectual property law, regulates the protection of information.

That is why the law operates as a data steward – it facilitates the value chain by balancing competing interests *in context* when it restricts human activity. The context of each data regulation will differ (e-commerce, evidence, crime, taxation, property, etc.) and, therefore, the law seeks to restrict everything that is necessary, but nothing more, in a delicate weighing exercise.[5] A plethora of statutes, judgements and common law provisions govern individual use cases. Some data law scholars criticise the sheer volume of legal measures applicable to ownership of information for being contrary to the sharing norms of scientific research.[7] For this reason, a global data policy which seeks to inculcate value into data itself will never survive legal critique unscathed.

The law restricts the application

The same holds true for the ability to extract value from data at will. The Data Policy is under the impression that those who amass data are free to manipulate it. Once again, this is at odds with the legal reality. Even those who hold data are seldom free to apply it for any purpose or at their own discretion, because data stewardship is a matter of combined legal measures.[4] Depending on the nature of the information, a volume of legal (including ethical, environmental, security and regulatory) *approvals* must be sought before information, whether in data format or not, may be processed. This administrative burden is so complex that it often accounts for a large portion of the resource investment in data processing. The Data Policy seems to neglect this fact and appears to convey that data sovereignty may be achieved merely through physical, local control.[4,5]

Myth 2: Data ownership is unclear

Of major concern is the Data Policy's manifest ignorance of the law regarding ownership of information. It assumes that possession is really nine tenths of the law, and that data are somehow a unique asset that exists independently from the information represented. Therefore, the Data Policy enviously refers to a strange creature it calls 'mega technological digital companies […] operational in selected countries' which, in their opinion, unfairly monopolise the (fantastical) inherent value in data. This myth is wrong-footed on both counts.

Most data are owned by someone

First, data ownership is not a question the law is concerned about – ownership vests in the information as it is expressed, whether it is data or not. As some notable legal scholars point out: 'few people have information; instead, what most people actually have is data'[8].

Therefore, the ownership of information is addressed by a variety of statutes, key among which is the Copyright Act. There are good reasons why copyright law is implicated directly where data are concerned. First, a large volume of the information contained in data is subject to copyright protection. Second, electronic communications operate by way of copying data, which means that the exclusivity of the right to make reproductions is always at stake.[8]

If the data consist of individual expressions of human intellectual endeavour, such as literary, musical or artistic works, films, broadcasts or computer programs, each data set (or even each data point), will automatically be owned by someone.[9] Because copyright law does not impose a merit test, it does not matter if the data point is meaningless.[9,10] Similarly, copyright protection has been held to vest in minute works (such as individual phrases or lines of code). Therefore, provided that a data point qualifies as a fit type of work, it may be individually protected. In many cases, data samples will be considered a literary work because it is a form of 'writing'[9] consisting of numerical values or symbolic language[11], but the same data may also qualify for protection as another type of work.

If the data points do not meet the threshold for protection as individual works (with individual owners), they may nevertheless be protected as a compilation[11] (with a single or joint owners) in a database. Ownership of the database vests in the person(s) who applied their mind to the selection and arrangement of information, regardless of where that information came from or whether it was lawfully obtained.[8] That is why, in South Africa, even the most mundane database may qualify for copyright protection.[9,11] All that is required to vest ownership is a sufficient, qualitatively small, effort in selecting and arranging the data. This effort does not have to be unique, novel, creative or in any way distinguishable from other databases, as long as it is clear that the selection and arrangement is the result of the author's own judgement, skill and labour.[10]

In practice, many databases consist of data collected by sensors, monitors, gauges or other electronic or technological means. In other cases, large volumes of data are produced automatically during the operation of a computer program. This has led some commentators to suggest that most databases will not qualify for copyright protection because they lack a human author.[12] This view is incorrect. The Copyright Act makes it clear that ownership in a computer-generated work, including individual works, compilations or databases,[9] will vest in the person who made the arrangements necessary to create the work.[13] The Supreme Court of Appeal[10] confirmed this and held that if the work has no human author, it is protected as a computer-generated work. If a human was involved in the creative process, it is protected as a computer-assisted work.[10]

The identity of the person who is vested with the rights in the work (data or database), is a factual question. It is determined, primarily, with reference to the first ownership rule.[9] Accordingly, the person who is identified, by law, as the author of the work will be the first owner of all the rights in that work.

There are, however, some cases where copyright will not vest in the information and it is, therefore, not subject to ownership. This would be the case where the work is not substantial enough, or is too commonplace, to qualify as a fit type of work. The most common example is statistical personal information. This is the reason why social media service providers insist that they do not own a user's data, because the user does not have a property right in their personal information. This fact does not prevent ownership from vesting in the database itself, but the use of that data will be restricted by the scope of authorisation granted by the data subject.

Data are transferrable property

Second, the law deals with data ownership as a matter of *intellectual property*. This means the law recognises the immaterial/intangible nature of the rights but has, nevertheless, incorporated it under the existing legal framework applicable to tangible property. Contrary to the view of some commentators on the Data Policy, copyright is movable property[14] and is protected against expropriation[14] in terms of section 25 (the property clause) of the Constitution[6,15].

Furthermore, the law acknowledges that this kind of property exists simultaneously everywhere in the world and also nowhere in the physical world. Internationally, copyright protection operates by way of the principle of national treatment.[9] This means the nature and scope of the owner's entitlement is determined according to the rules applicable in the country where the data are exploited[3] and will exist independently and simultaneously in all relevant countries of the world[7]. That means, in legal terms, data cannot be moved from one country to another.[4] Its storage location may change, but even if the data are entirely deleted in one country, the legal ownership in that country remains valid. However, to exploit the taxation potential of intellectual property, the South African Treasury has declared that any change in owner (by assignment) from a national to a foreigner, will amount to an export of capital and will therefore trigger foreign exchange control clearance.[9] This means that, if the Data Policy succeeds in forcing localisation of data, it will also trap that information in South Africa. This will likely have a devastating impact on foreign investment in research and development. This is one example of what commentators call the 'unintended and perverse consequences' of using the 'blunt tool' of data localisation.[5] Other commentators add that 'measures which introduce policy uncertainty or otherwise disincentivise investment must be avoided'[16] or else it will imperil South Africa's ability to compete on the African continent.

Myth 3: Controlling infrastructure allows control of data

The Data Policy is devoted to establishing local data processing 'infrastructure' as a combination of physical data storage/processing facilities and cloud services. Nowhere does it consider the fact that infrastructure consists of hardware *and software*. Without the appropriate computer programs in place to use the data, it is impossible to derive any downstream value from the data. This means the policy fails to recognise a number of substantial legal barriers to achieving its goal of promoting access.

Local access requires local software maintenance and development

A significant spin-off industry from the data economy is the creation and maintenance of database-related software. To be effective, these computer programs must handle large volumes of data and do so securely and reliably. All three factors pose substantial risks to the integrity of the data, the service, the data subject and an array of third parties. Therefore, the stability of data processing infrastructure has become big, lucrative, business[4] and it is a software-based issue. Very few database controllers will assume the cost and risks associated with in-house software and prefer to rely on proven solutions. In Africa, that means imported solutions, i.e. software copyright licences.

For many new entrants to the data market, the software licensing costs are prohibitive. Providing access to the data along with access to the software will not solve the problem of 'opportunity costs.'[4] On the contrary, it will exponentially increase the licence cost, because the programs will have to be sub-licensable, and will transfer that cost to the taxpayer. Commentary on the policy suggests that 'this policy will add unnecessary operational burdens on the shoulders of smaller businesses and make their growth yet more uncertain and difficult'[17] while others demonstrate that the policy will increase economic risks by, inter alia, impacting negatively on enterprise productivity, global market access and manufacturing[18].

Home grown infrastructure is unlikely

The only alternative is locally produced database software. To design such programs afresh is simply not feasible and the local software industry will have to learn from existing programs. However, the law currently prohibits access to the knowledge contained in computer program code. It is impossible to legally decompile a computer program in order to understand how it works.[19] A proposal to permit decompilation, by way of section 19B of the Copyright Amendment Bill 2017, is currently under re-consideration. However, even if this problematic provision becomes law, it will only permit decompilation in order to create an interoperable program. The proposed exception expressly prohibits creating a functionally equivalent program. Furthermore, any knowledge gained from decompilation may not be shared with anyone for any purpose. That means each program developer must incur the same substantial cost associated with decompilation before they may begin developing any database-applicable program and, even then, the result of their efforts may not be substantially similar to the program on which it was based.

Digital rights management criminalises access

While copyright law provides wide and strong protection for data and databases, it does not address the cybersecurity concerns associated with digital information. Therefore, the vast majority of data and databases is protected by layers of electronic security measures, ranging from the most basic (such as passwords, biometric user authentication, watermarking and usage logging) to the extreme (an array of data and communication encryption and real-time monitoring). However, these measures are of lesser value to the data controller if there is no punishment for a transgression.[20]

This aspect is dealt with by another area of law, namely anti-circumvention protection or digital rights management (DRM) provisions[19] contained in section 86 of the *Electronic Communications and Transactions Act 25 of 2002* and set to be replaced by the more extensive provisions in chapter 2 part 1 of the *Cybercrimes Act 19 of 2020*. The DRM provision *criminalises* the act of accessing a computer program or the medium on which it is stored, if permission has not been granted or the scope of use exceeds the extent of authorisation. Using, or being in possession of, a computer program designed or adapted primarily to access data or a computer program is also a crime.[8] The Copyright Amendment Bill creates additional offences when the electronic means which identify the owner or prevent access or copying of the work are tampered with. This double layer of protection over software, coupled with the decompilation limitation, impedes the development of local database software.

The majority of DRM provisions are attempts, based on foreign law, to address piracy and an array of cybercrimes and are not per se unwelcome.[21] However, the manner in which these provisions are drafted is often draconian and pose a barrier to participation in the digital economy.[20] DRM law simply means that *prior* authorisation to access data must be obtained, and this permission may be withheld unless a fee is paid. This cost is in addition to the copyright licensing fee but, unlike copyright infringement, DRM infringement is a crime.

Consequently, to make government's dream of open access to data possible, it will have to amend DRM law to decriminalise certain use cases (creating significant risk) or incur the cost of obtaining permission on behalf of its intended users. What government will do if permission is refused is a worrying, but not unlikely, eventuality.

Conclusion

The relationship between the law and data is a complex one. Unfortunately, the Data and Cloud Policy's ignorance of the most basic legal principles indicates an alarming disregard for private ownership and postulates a future where very little data will be safeguarded by the law to the extent it is now. Fortunately, government policy is not law and, as this article shows, many things will have to change in order to implement the policy goals. That is why every data stakeholder, in particular those who rely on the downstream value in research and development, should undertake an intellectual property rights audit as a matter of urgency. It may well be wise to divert from current strategy to safeguard their interests.

Acknowledgements

This article was prepared with financial support from the Vice-Rector: Research, Innovation and Postgraduate Studies, Stellenbosch University.

Competing interests

I declare that I have no competing interests. I provided research input to the author of one of the reports, published by the Mandela Institute of the University of the Witwatersrand and cited in this article, in my personal capacity.

References

1. Draft National Policy on Data and Cloud. Government Gazette No 44389 Notice 306 (1 April 2021).
2. Beyleveld A. Data protection in Kenya, Nigeria and South Africa in the 2020s and Beyond. Policy Brief 01. Johannesburg: Mandela Institute, University of the Witwatersrand; 2021.
3. World Intellectual Property Organisation (WIPO). Overview of the international protection of copyright and related rights: From the Berne Convention for the Protection of Literary and Artistic Works to the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty WIPO/CR/DAM/05/8. Proceedings of the WIPO National Seminar on Copyright and Related Rights for Lawyers and Judges; 2005 April 27–28; Damascus, Syria. Geneva: International Bureau of WIPO; 2005. p. 1–19.
4. De la Chapelle B, Porciuncula L. We need to talk about data: Framing the debate around free flow of data and data sovereignty. Paris: Internet and Jurisdiction Policy Network; 2021.
5. Van der Berg S. Data protection in South Africa: The potential impact of data localisation on South Africa's Project of Sustainable Development. Policy Brief 02. Johannesburg: Mandela Institute, University of the Witwatersrand; 2021.
6. Shay RM, Van der Walt A. Constitutional analysis of intellectual property. PER. 2014;17:52–85. https://doi.org/10.4314/pelj.v17i1.02
7. Reichman JH, Okediji RL. When copyright law and science collide: empowering digitally integrated research methods on a global scale. Minn Law Rev. 2012;96:1362–1480.
8. Van der Merwe DP, Roos A, Pistorius T, Eiselen GTS, Nel SS. Information and Communications Technology Law. 2nd ed. Johannesburg: LexisNexis; 2016. p. 1–49, p. 1362–1480.
9. Dean OH. Handbook of South African Copyright Law. 14th ed. Johannesburg: Juta Law Publishers; 2012.
10. Haupt t/a Softcopy v Brewers Marketing Intelligence (Pty) Ltd and Others 4 SA 458 (SCA). 2006.
11. Blignaut H. Copyright. In: Dean OH, Dyer A, editors. Dean & Dyer introduction to intellectual property law. Cape Town: Oxford University Press; 2014. p. 1–76.
12. Research ICT Africa. Written submission in response to the proposed National Data and Cloud Policy [document on the Internet]. c2021 [cited 2021 Sep 03]. Available from: https://researchictafrica.net/wp/wp-content/uploads/2021/06/RIA_Submission_DATA_and_Cloud_Policy.pdf
13. Tong L-A. Case Comment: Copyright and computer programs, computer-generated works and databases in South Africa. Eur Intellect Property Rev. 2006;28:625–628.
14. Du Bois M. Intellectual property rights and the Constitution. In: Dean OH, Dyer A, editors. Dean & Dyer introduction to intellectual property law. Cape Town: Oxford University Press; 2014. p. 466–491.
15. Dean OH. The case for the recognition of intellectual property in the Bill of Rights. J Contemp Roman Dutch Law. 1997;60:105–119.
16. Internet Service Providers' Association. Submissions on the Proposed National Data and Cloud Policy. [document on the Internet]. c2021 [cited 2021 Sep 03]. Available from: https://ispa.org.za/wp-content/uploads/2021/06/ISPA-Submission-Proposed-Data-and-Cloud-Policy-20210601.pdf
17. Free Market Foundation Rule of Law Project. Submission to the Department of Communications and Digital Technologies on the National Data and Cloud Policy [document on the Internet]. c2021 [cited 2021 Sep 03]. Available from: https://www.freemarketfoundation.com/dynamicdata/documents/20210611-submission-on-national-data-and-cloud-policy-2021.pdf
18. Global Data Alliance. Comments to the Republic of South Africa on The Proposed Data and Cloud Policy [document on the Internet]. c2021 [cited 2021 Sep 03]. Available from: https://globaldataalliance.org/wp-content/uploads/2021/07/05122021gdasafrdatacloud.pdf
19. Jooste C. Copyright law and the Internet. In: Papadopoulos S, Snail S, editors. Cyberlaw@SA. 4th ed. Pretoria: Van Schaik; 2021. p. 181–258.
20. Samuelson P. Intellectual property and the digital economy: Why the anti-circumvention regulations need to be revised. Berkeley Technol Law J. 1999;14:5–10. https://doi.org/10.1145/318536.318538
21. Jooste C, Karjiker S. Intellectual property law in the digital environment (EIP law). In: Dean OH, Dyer A, editors. Introduction to intellectual property law. Cape Town: Oxford University Press; 2014. p. 390–465.