

# Notes on Data Sharing Policy in South Africa

---

*Wim Hugo and Johan Pauw*

*January 2013*

## International Trends

1. Data sharing and availability of publicly funded RDI<sup>1</sup> outputs at no cost, or at the cost of dissemination, is a widely adopted trend in the developed world<sup>2</sup>, and is the basis of engagement for a number of influential organisations<sup>3</sup>, including GEO<sup>4</sup> and ICSU<sup>5</sup>. This trend is expected to continue and penetrate into developing countries.
2. There is increasing support from the scientific community for the concept of peer-reviewed data publication. This is in response to the fact that many scientists are focused on the production of data sets, and do not receive adequate professional recognition for their efforts. It is expected that this will lead to the following consequences:
  - a. Improved accessibility to and description of data sets, driven by scientists' desire to publish;
  - b. The majority of these published data sets will be freely available in the public domain, supported by transparency demanded in scientific journals<sup>6</sup> and the emerging infrastructure allowing linkages between data sets and scholarly articles.
  - c. A need for the infrastructure to publish, curate, and disseminate published research outputs is implied – and not all institutions or relatively short-lived funded projects will be in a position to provide such infrastructure.

In broad terms, the trend can be summarised as 'Data Democracy' or 'Free and Open Access', which is based on the following two principles:

1. Free access: it is recognised that data should be free of charge if possible, but this does not mean that reasonable cost of availability and dissemination cannot be recovered. This is often funded as a public good.
2. Open access: the emphasis is on equal opportunity to discover, obtain, and use the data without prejudice.

---

<sup>1</sup> Research, Development and Innovation outputs

<sup>2</sup> OECD: <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

<sup>3</sup> List of organisations subscribing to a free and open data policy: [http://www.codata.org/data\\_access/policies.html](http://www.codata.org/data_access/policies.html)

<sup>4</sup> GEO: <http://www.codata.org/GEOSS/GEOdataPolicyBriefingMar07dist.pdf>

<sup>5</sup> <http://www.icsu.org/events/ICSU%20Events/international-symposium-the-case-for-international-sharing-of-scientific-data-a-focus-on-developing-countries>

<sup>6</sup> Brussels Declaration: <http://www.dlib.org/dlib/january11/smit/01smit.html>

Developing countries stand to benefit considerably from trends towards free and open access to data, and are already doing so.

## Motivations for Free and Open Data Access

There are potentially many motivations for promotion of Free and Open Access to data, but most can be distilled into just two broad lines of reasoning:

1. Governments invest large sums of money into science as a driver for data acquisition, knowledge creation, capacity building, and innovation. This is a **virtuous circle** that is fed by availability of data, information, knowledge, and capacity. At a basic level, reduced access<sup>7</sup> to the outputs of such funding diminishes the return on investment. What is often overlooked, though, is that reducing the efficiency of this feedback loop hurts the return on investment even more by limiting indirect returns and collaboration. It is the equivalent of removing compound interest from a financial investment.
2. Secondly, the public has funded the outputs from state department data collection and state funded research, and **owns it already**. Allowing selective access infringes basic rights of citizens in general, and the scientific community in particular. This right has to be balanced by reasonable measures to allow researchers to exploit the academic value of their work without undue competition, typically in the period leading up to publication of a paper or thesis.

## Counter-Arguments

Counter-arguments generally fall into the following categories:

1. Developing Countries often indicate that they regard some of their research data as sensitive, based on the (possible) **future commercial value** of such data, or on its **conservation implications** – especially in the case of natural resources. Each case has to be evaluated on merit, and is not compromised by a generally free and open access policy.
2. The data has **current commercial value**, and the state entity depends on the income derived from it to fund its operations.
3. It is obvious that some data held by the state is **private** to companies or individuals.
4. The users will **apply the data incorrectly**, will apply it to **challenge government**, or **gain financially** from it: this is the least defensible argument, in the sense that the onus is on the user to ensure the correct application of the data, that it is perfectly reasonable for the public to assess and challenge government in respect of its conclusions and performance, and that the public investment in research output, if applied for financial gain outside of government, creates jobs, opportunities, and growth.

Hence, while it is reasonable to limit access to some data, for the reasons outlined above, this should be the exception and not the norm: default policy should support free and open access.

---

<sup>7</sup> Typical reductions in access: financial constraints, embargoes and classification, poor dissemination, lack of standardisation, poorly described data.

## Legal Aspects

There is a tension between the drive towards free and open access, and current or future legislation.

1. Intellectual Property Rights from Publicly Financed Research and Development Act (Act No 51 of 2008): potentially restricts the access to research data and outputs which would conflict with the principles of data democracy. On the other hand, it protects the exploitation of intellectual property in cases where research outputs have a specific commercial application.
2. Spatial Data Infrastructure Act (Act No 54 of 2003): Improves discoverability but does not guarantee access. Could limit data availability because of the legal obligations of 'custodianship', which may make institutions unwilling to publish all available data sets.
3. South African Weather Service Amendment Bill: A benchmark example of a step in the wrong direction, it impedes climate change research and discriminates against local enterprise while foreign entities are free to continue with services and data in competition with the SA Weather Service.
4. Promotion of Access to Information Act, 2000 (Act No. 2 of 2000):
5. Protection of Information Bill: policy or regulation required to declassify and exempt the data.
6. National Archives and Records Service of South Africa Act, 1996 (Act No. 43 of 1996)

***The concern is, in almost all cases, not the intent of the Act but the potential to abuse the Act and unduly or unfairly restrict access thereby harming knowledge creation and socio-economic development.***

## Policy Recommendations

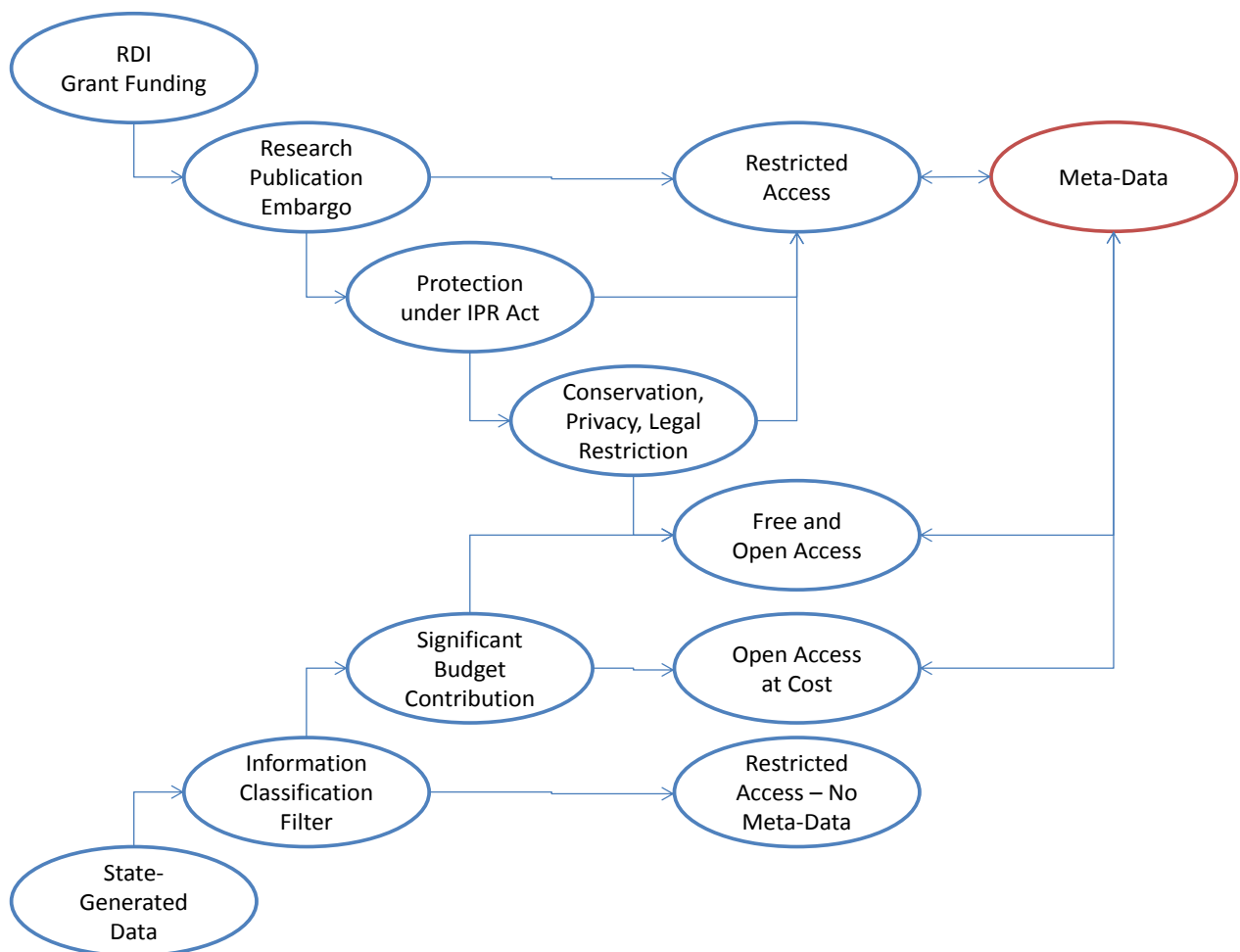
1. All data, information, and research outputs generated by state-funded means should ideally be included in a national policy. This includes grant-funded university research, and data generated by state departments. It specifically excludes contract research performed for private entities.
2. The data so described should be exempted or declassified in terms of the Protection of Information Bill, as provided for in the authority of each state department<sup>8</sup>.
3. The legal meaning of 'custodian' of a publicly funded database should include the provision of free and open access.
4. Conditions and qualifications are valid, based on
  - a. Reasonable embargoes to allow publication of research and papers;
  - b. Privacy afforded to legal entities and individuals;
  - c. Conservation considerations, mainly to protect endangered species and habitats;
  - d. Infringement of rights afforded by the Intellectual Property Rights Act;
  - e. Infringement of current or future legal rights to exploit natural resources.
5. Existing arrangements where state departments or agencies derive a significant proportion of their income from the sale of publicly funded data should be discontinued and mitigation measures should be put in place to replace the loss of income.

---

<sup>88</sup> Article 19, Authority to Declassify

6. If the state implements a policy of free and open access to data, the state should fund and support the electronic infrastructure whereby research outputs are preserved, discovered, and accessed. This infrastructure is a necessary expenditure<sup>9</sup> to protect the investment in Research, Development and Innovation, and should complement and underpin the arrangements put in place by individual state departments, and academic or research institutions. For general environmental data as required by the Risk and Vulnerability Atlas, the Department of Science and Technology has already commissioned the South African Environmental Observation Network to develop and host the South African Earth Observation System (SAEOS), a central data portal with the functionality to access geo-spatial data from distributed data holdings. It is therefore recommended state departments and agencies should be required to provide access to their data holdings to the SAEOS.

### Policy Process Model



The process model makes provision for four states of access:

1. RDI outputs for which access is restricted, due to three situations:
  - a. Grant-funded research that has not yet been published is available to a collaboration community but not generally;

<sup>9</sup> Proposals in this regard are in preparation, utilising the DST-funded infrastructure provided by CHPC and SANReN

- b. Grant-funded research that has a commercial exploitation potential that far outweighs its general public good, and is offered protection under the Intellectual Property Rights Act;
  - c. Grant-funded research that is embargoed because of conservation considerations, privacy issues, or legal rights of third parties.
  - d. In all of these cases, it remains good practice to publish the meta-data openly: this means that the research outputs are discoverable but not openly or freely accessible.
2. State-generated data that is classified: not openly and freely accessible, and no meta-data is openly published.
3. State-generated data that currently supports a significant proportion of the income of a state entity, and for which the aim is to phase out such dependency where possible. Meta-data is openly published, and the data is openly available but not free.
4. The balance of grant-funded or state-generated outputs is freely and openly available and meta-data is published openly.

## Separation of Policy and License

An important aspect to consider in respect of maintainability:

1. Policy and the license in support of the policy must be separated. Quality public domain licenses (such as Creative Commons) exist, and these should be used by preference.
2. Creative Commons has released a new family of licenses that can accommodate 'local' conditions (note that these are not restrictions). These conditions map to the exclusions mentioned in our policy advice: conservation considerations, privacy and ethics, business confidentiality, etc.
3. The two most commonly used licenses are:
  - a. [Creative Commons Attribution-ShareAlike 3.0 Unported License](#)
  - b. [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](#)

For both these recommended licenses, access is unrestricted, and attribution is required. In the latter case, commercial uses are specifically excluded.

This document is made available under the [Creative Commons Attribution-ShareAlike 3.0 Unported License](#)

