



# DIRISA Guidelines

---

*Policies and Licensing*

Preliminary Guideline  
Wim Hugo, Johan Pauw  
18 December 2015  
Document Reference: G805.4.1.5.1  
Version: Draft



## Document Information

### Schedule of Documents

The document is based on several preceding documents:

Document	Date Issued	Authors	Notes
G342.1.15.4 Notes on Data Sharing Policy in South Africa	January 2013	Wim Hugo Johan Pauw	Served as an input to NRF policy

### Version Control

Version	Date Issued	Authors	Description of Major Changes
G805.4.1.5.1	18 Dec 2015	Wim Hugo	Updates to and extension of G342.1.15.4

### Sign-Off

Designation	Date	Organisation	Name and Signature
Client Principal		DST	
Client – Contractual		Meraka Institute, CSIR	
Supplier	21-05-2015	SAEON	Wim Hugo

### License

Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)	Free to share, adapt, and apply the work, even for commercial purposes, provided that you give appropriate credit and distribute the adaptations under the same license.
<a href="http://creativecommons.org/licenses/by-sa/4.0/">http://creativecommons.org/licenses/by-sa/4.0/</a>	

## Table of Contents

<b>DOCUMENT INFORMATION .....</b>	<b>2</b>
<b>SCHEDULE OF DOCUMENTS.....</b>	<b>2</b>
<b>VERSION CONTROL.....</b>	<b>2</b>
<b>SIGN-OFF .....</b>	<b>2</b>
<b>LICENSE.....</b>	<b>2</b>
<b>TABLE OF CONTENTS.....</b>	<b>3</b>
<b>INTERNATIONAL TRENDS.....</b>	<b>4</b>
<b>MOTIVATIONS FOR FREE AND OPEN DATA ACCESS.....</b>	<b>4</b>
<b>COUNTER-ARGUMENTS.....</b>	<b>5</b>
<b>LEGAL ASPECTS .....</b>	<b>6</b>
<b>POLICY RECOMMENDATIONS .....</b>	<b>6</b>
<b>POLICY PROCESS MODEL.....</b>	<b>7</b>
<b>SEPARATION OF POLICY AND LICENSE.....</b>	<b>9</b>
<b>ANNEXURE: LICENSES.....</b>	<b>10</b>

## 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.

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

---

<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>

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. Protection of Personal Information Act, 2013 – places necessary constraints on the gathering and application of personal information.
7. 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>88</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;

---

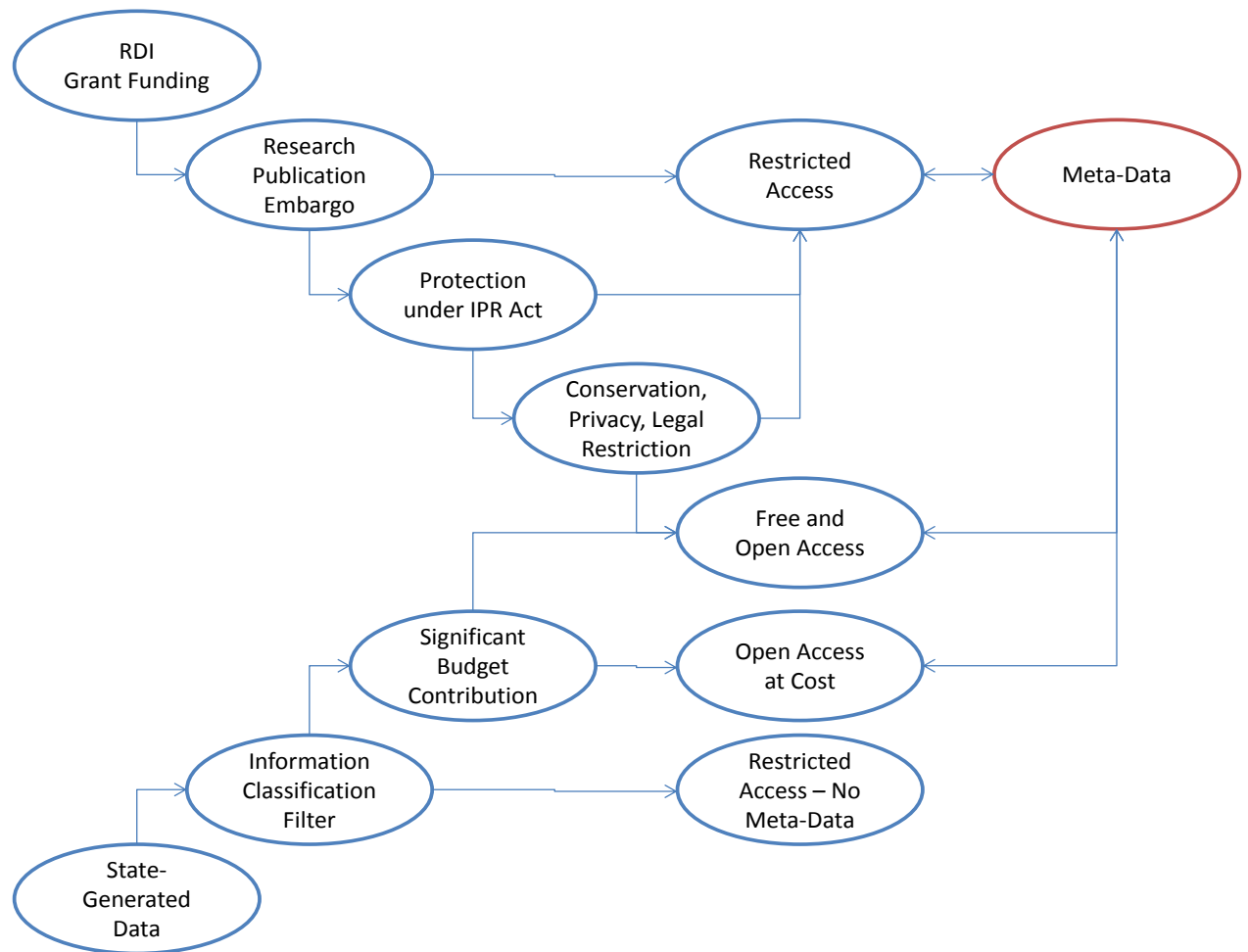
<sup>88</sup> Article 19, Authority to Declassify

- 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.
  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.

## **Policy Process Model**

---

<sup>9</sup> This programme has commenced, utilising the DST-funded infrastructure provided through DIRISA, CHPC and SANReN



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;
  - 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 licenses cannot be made more restrictive, and hence cannot be applied for the conditions of use identified earlier.
3. Licenses should be machine-readable, so that the following can be achieved:
  - a. Automation of access;
  - b. Automation of licensing of derivative works

### The Licenses



**Attribution  
CC BY**

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

[View License Deed](#) | [View Legal Code](#)



**Attribution-ShareAlike  
CC BY-SA**

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

[View License Deed](#) | [View Legal Code](#)



**Attribution-NonCommercial  
CC BY-NC**

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

[View License Deed](#) | [View Legal Code](#)



**Attribution-NonCommercial  
CC BY-NC**

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

[View License Deed](#) | [View Legal Code](#)



**Attribution-NonCommercial-ShareAlike  
CC BY-NC-SA**

This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.

[View License Deed](#) | [View Legal Code](#)



**Attribution-NonCommercial-NoDerivs  
CC BY-NC-ND**

This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

[View License Deed](#) | [View Legal Code](#)

DIRISA has identified additional licenses to support the conditions and exceptions identified in this document, these are presented in the Annexure together with the applicable Creative Commons licenses.

This document is made available under the [Attribution-ShareAlike 4.0 International \(CC BY-SA 4.0\)](#) license. 😊

## Annexure: Licenses

License URL	Short Description
Public Domain: CCO <a href="https://creativecommons.org/about/cc0#">https://creativecommons.org/about/cc0#</a>	Use this universal tool if you are a holder of copyright or database rights, and you wish to waive all your interests in your work worldwide.
Attribution 4.0 International (CC BY 4.0) <a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a>	Free to share, adapt, and apply the work, even for commercial purposes, provided that you give appropriate credit.
Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0) <a href="http://creativecommons.org/licenses/by-nd/4.0/">http://creativecommons.org/licenses/by-nd/4.0/</a>	Free to share, even for commercial purposes, provided that you give appropriate credit.
Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) <a href="http://creativecommons.org/licenses/by-sa/4.0/">http://creativecommons.org/licenses/by-sa/4.0/</a>	Free to share, adapt, and apply the work, even for commercial purposes, provided that you give appropriate credit and distribute the adaptations under the same license.
Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) <a href="http://creativecommons.org/licenses/by-nc/4.0/">http://creativecommons.org/licenses/by-nc/4.0/</a>	Free to share, adapt, and apply the work, for non-commercial purposes, provided that you give appropriate credit.
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) <a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a>	Free to share, adapt, and apply the work, for non-commercial purposes, provided that you give appropriate credit and distribute the adaptations under the same license.
Attribution-NonCommercial-NoDerivs : CC BY-NC-ND <a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a>	Free to share, provided that you give appropriate credit, apply the work for non-commercial purposes only, and do not distribute modifications or derivatives. <b>Note that this is not a free culture license.</b>
Restricted: Ethics and Privacy <a href="http://noframe.data.dirisa.org/platform/policy-and-licenses/restricted-ethics-and-privacy">http://noframe.data.dirisa.org/platform/policy-and-licenses/restricted-ethics-and-privacy</a>	The work is restricted in its use in one of the following ways: the end user must be known to the provider of the data, and/ or can only access the data under controlled conditions. Publication of derivatives to be agreed with the data provider.
Restricted: Commercial Value Non-standard - URL provided by owner	The work is available only on payment of an access fee or license fee. The owner of the work is free to determine any further restrictions on use by way of a license for the work.
Restricted: Classified Government Data Non-standard - URL provided by owner	Access to the work is restricted due to government classification. Access to the work is determined on individual merit and the owner of the work has full discretion in this regard.