Greenhouse Gas Observation & Climate-Smart Agriculture

NRF National Research Foundation | SAEON South African Environmental Observation Network

Data Policy Considerations



Wim Hugo

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730995

Project: 730995 - Supporting EU-African Cooperation on Research Infrastructures for Food Security and Greenhouse Gas Observations (SEACRIFOG)

Work package number: 5
Work package title: Interoperability of infrastructures, access and data sharing
Deliverable number: 5.2
Deliverable title: Guidance on Data Policy and Supporting Open Licenses
Data Policy Considerations
Lead beneficiary: NRF SAEON
Lead authors: Wim Hugo (wim@saeon.ac.za)
Contributors: Johannes Beck (SASSCAL)
Submitted by: Wim Hugo

Part 1: Review

Context

This document provides a basis for a future operational SEACRIFOG infrastructure to agree policies for itself, provide guidance to members on policy, and provide policy to end users of SEACFRIFOG services.

Policy Attributes

Why do we need policy, and what are the attributes of a good policy? The following assessment provides a synthesis from a number of sources [17], [18], [19], [20], [25]:

- **Serve a Need**: Firstly, a policy should address one or more objectives, goals, or aspirations of the organisation or initiative. In many cases, policy is needed to ensure compliance with legislation or regulations, but the objectives may also result from an organisation's strategy.
- **Concise Policy Statement:** Communication is improved by having a concise statement of policy that can be used as a placeholder for the comprehensive policy document.
- **Detailed Policy Statement:** This should cover the following aspects:
    - **Definitions** of terms and role-players.
    - **Objectives**, based on the need.
    - **Authorisation or Mandate:** who authorises the policy, and what authority do they have to issue the policy?
    - **Applicability and Scope**: defining the role-players or use cases covered by the policy, as well as important **exclusions**. Definition of the research area or focus of the policy, and the type of research output covered by the policy (Annexure C).
    - **Effective Date Range**: start date, and optionally an end date for the policy.
- **Responsibilities**: The roles and responsibilities of stakeholders associated with the policy and its implementation should be stated clearly. These stakeholders are often outside the organisation or entity that creates the policy, and hence consideration should be given to the efficacy of policy provisions in situations where direct control is not possible.
- **Instruments**: Policy implementation is done via instruments, of which there are several generic classes:
    - **Procedures**: The main instrument whereby policies are implemented. Each policy statement may in practice result in several (standard) operating procedures.
    - **Specifications and Standards**: These are the tests or norms that apply to elements of policy or procedure[1].

[1] Licenses are a good example in this context: policy might require that an appropriate license is provided for a published research output. The **specification** might be that such a license support open science and be machine-readable, and the **standard** or norm will be one of the Creative Commons licenses.

- **Compliance**: Measurement of compliance with the policy should be elaborated or considered and defined. Compliance is not binary: some provisions need to be complied with, while other may be suggestions of good practice.
    - **Mandatory**: Compliance with these provisions are mandatory, and in essence the policy needs to be enforceable in such circumstances. It includes aspects of legal and regulatory compliance, ethics provisions, and privacy.
    - **Optional/ Guidance**: Compliance cannot be enforced, but is requested or suggested - an appeal to implement good practices and act responsibly[2].
    - **Conditional**: Some provisions only apply in cases where one of the optional compliance elements are present[3].
- **Incentives and Consequences**: Non-compliance should result in consequences for those that do not execute their responsibilities, but incentives or positive outcomes of compliance are generally more effective.
- **Review**: conditions under which the policy will be reviewed, and the review frequency.
- **Infrastructure**: Policy is often published without due consideration of the infrastructure and funding implications of the policy. It is especially important to consider the impact on third parties - for example funder policy implications for institutions and grantholders.

By and large, policy formulation should also take into account that there may be established community practice, standards, and infrastructure to consider [25].

Open Access, Open Science, and Open Data

Open Science [1], which has as one of its most important goals to make science and research available to all interested parties, requires some important technical elements to allow its implementation. The most important and obvious requirement would be for research outputs to be made available with largely open licenses, based on policies that support variations of Open Access. There are two additional nuances to these concerns:

1. Combination of research outputs require creation of a new license that honours the provisions of the input licenses, even though these may differ in conditions placed on the end user. For this reason, the field of Legal Interoperability [2] has received attention, and it can be viewed as an extension of the other aspects of interoperability.
2. While Open Science promotes liberal licenses, there are some research outputs that need to be restricted in respect of application and dissemination[4]. In general, to process such licenses in an automated way requires well-established authentication and identity services.

[2] In license provisions, attribution is requested but is not practically enforceable. Hence it should be viewed as 'good practice' or 'responsible behaviour'..

[3] Using license provisions as an example again: attribution is conditional on subsequent publication of a derivative work, and does not apply if a derivative is used privately.

[4] Typically on conservation, ethics, or privacy grounds.

The achievement of Open Science has recently received significant impetus through the publication of and support for 'Plan S', which aims to create open access for all scientific output and publication in a short space of time [50]. This initiative is supported by a number of funders, and critically, by the EU [51].

International Trends

The following trends - roughly emerging about a decade ago - are important from a data policy perspective:

1. Data sharing and availability of publicly funded RDI[5] outputs at no cost, or at the cost of dissemination, is a widely adopted trend in the developed world [3], [4], and is the basis of engagement for a number of influential organisations [5], including GEO and ICS. 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[6] 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 cost 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. Table 1 summarises data on R&D expenditures worldwide, classified by broad global affiliation. As can be seen, the bulk of global R&D expenditure is done in the 36 OECD countries (these subscribe almost universally to Open Science/ Open Data policy). BRICS countries spend

[5] Research, Development and Innovation outputs
[6] Brussels Declaration: http://www.dlib.org/dlib/january11/smit/01smit.html

approximately 50% of the OECD total, with a small amount originating in the remaining 49 countries for which data is available.

Not only does the flow of open data point overwhelmingly in the direction of the developing world (95% vs 5% of R&D expenditure), but spending per capita in OECD and BRICS countries is 4 times higher than in the rest. In essence, developing countries can focus their limited resources on population development and rely on scientific data generated openly by the rest of the world.

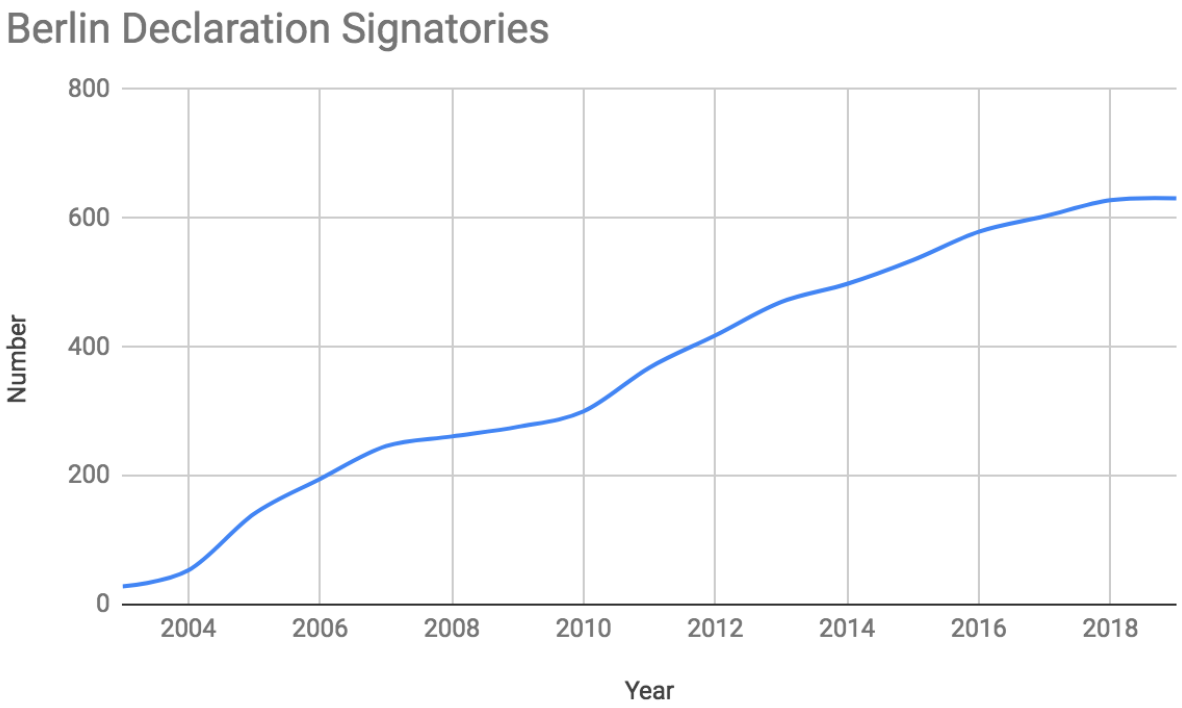
Table 1: R&D Expenditures by Global Grouping [6]

	Expenditure ($ billion)	% of GDP
OECD	1,253	2.43%
BRICS	604	1.64%
Others	104	0.62%

The Berlin Declaration [7], [16] provided an early impetus and rallying point for Open Access, and provided institutions (academic and research institutions, funders, initiatives) with a publicly accessible commitment to Open Access. To date, more than 600 signatories have committed to the principles of the declaration. The declaration makes the case for open access to data:

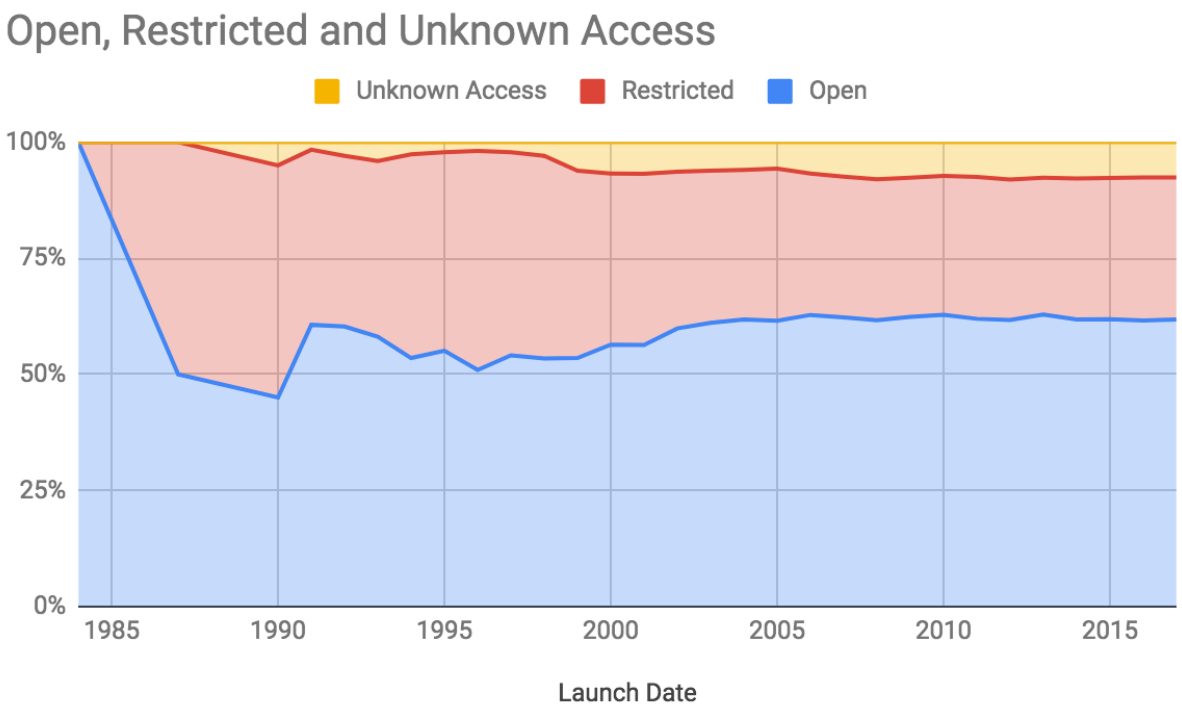
“"Raw research data should be made freely available to all researchers. Publishers encourage the public posting of the raw data outputs of research. Sets or sub-sets of data that are submitted with a paper to a journal should wherever possible be made freely accessible to other scholars.”

Figure 1: Signatories to the Berlin Declaration [7]



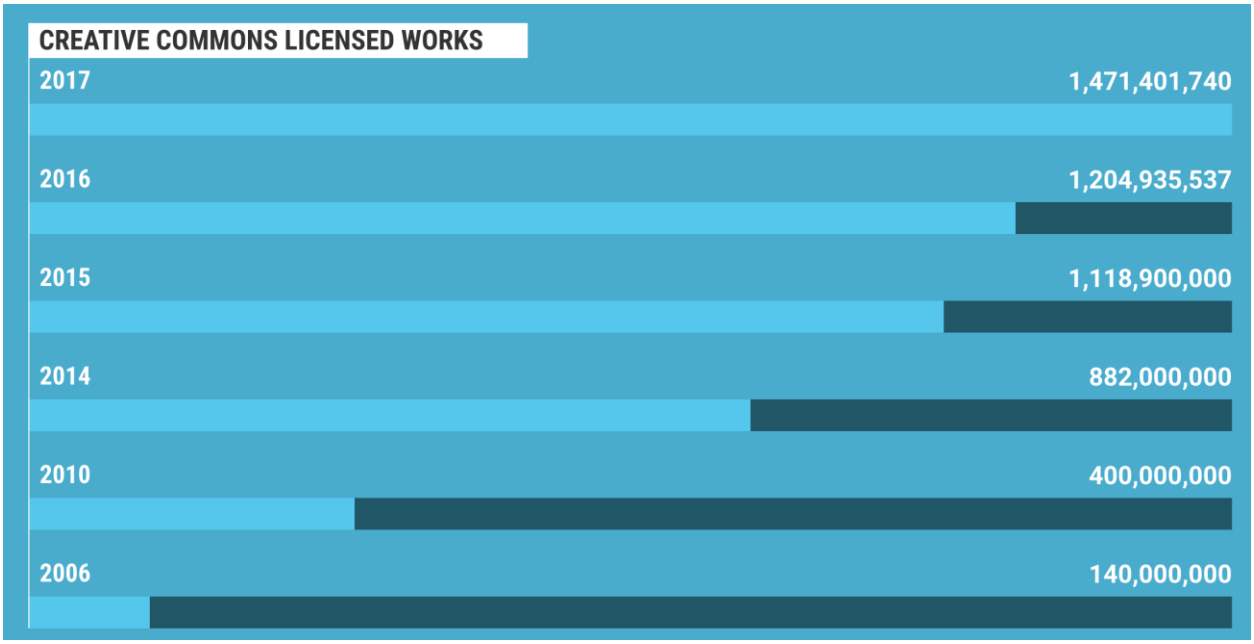
The earth observation community, through CEOS and GEO, have been pioneers in the promotion of open access since the 1990s, and while some outputs are likely to remain unavailable due to commercial or security concerns, the bulk of satellite imagery remains openly available and is modestly increasing [41].

Figure 2: Percentage of Open Satellite Missions [41].



Finally, there has been an extraordinary growth in the number of digital objects (data, text, code, media) that have now been published under so-called free culture licenses [8] (such as the Creative Commons licenses, or open source software licenses). Using Creative Commons as a case study, the following trends are of interest: works available using some form of Creative commons license have grown from 140m a decade ago to more than 1.4 billion.

Figure 3: Growth in Creative Commons Licenses [55]



Arguments For and Against Open Access

There are potentially many motivations for promotion of Free and Open Access to data, but most can be distilled into just three broad lines of reasoning [10]:

- **Investment Return** [6], [23]: 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 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.
- **Publicly Owned**: Secondly, the public has funded the outputs from state department data collection and grant-supported 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.
- **Credibility of Science**: Reproducibility of science is contingent on having the data used to obtain a result available for open scrutiny and peer review [11], [12].

Counter-arguments generally fall into the following categories [10], [21], [22]:

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 implication***s – especially in the case of natural resources [21], [22]. 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
    a. ***apply the data incorrectly***: this is the least defensible argument, in the sense that the onus is on the end user to ensure the responsible application of the data - provided reusability is maximised by way of proper metadata;
    b. will apply it to ***challenge government***: in democratic society it is reasonable and increasingly possible [54] for the public to assess and challenge government in respect of its conclusions and performance, but this aspect of accountability is reported by some as an important impediment to open access in developing countries [22];
    c. ***gain financially*** from it: the public investment in research output, if applied for financial gain outside of government, creates jobs, opportunities, and growth [22], [23].

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.

There is, however, progress possible in developing countries in respect of open data, largely due to a number of drivers [22]:

- Political will to improve government effectiveness;
- The interest of government to increase accountability [54],
- Internal pressure;
- External pressure and
- Reputation gain.

Defensible limitations on use of open data is discussed in more detail on the section below dealing with licenses.

Legal Aspects

A comprehensive review of the considerations applicable to interoperable, open data is provided by Uhlir et al. [2] in a consensus guideline developed under the auspices of RDA and CoDATA.

Licenses

License Considerations

Licenses need to be separated from policies, since licenses need to be machine-readable. In addition, a single policy can reference or require many licenses.

Creative Commons licenses are preferred because [14]

1. There are human readable, legal, and machine readable versions of each license;

2. The licenses have been verified for many jurisdictions.

Machine readability is important for two reasons:

1. This allows automated determination of conditions applicable to download and visualisation of digital objects (if any), and allows search facilities to distinguish results on the basis of license conditions;

2. Composite objects (mash-ups) and derived objects require automated computation of a resulting license.

Creative Commons licenses allow three types of qualifications [13], [14], [15]:

1. **Permissions**: rights of the owner that are accorded to the end user;
2. **Conditions**: performance expected from the end user in return for the permissions - for example the requirement to cite a work;
3. **Limitations**: rights that are expressly not granted to the end user (for example disallowing commercial applications).

There are legitimate considerations for making licenses restrictive. These are:

1. Release of the data can harm an individual, community, or natural environment. Such restrictions apply in the case of ethics and privacy concerns (personal data); indigenous community resources, and endangered species, for example.

2. Release of the data can impact the commercial and financial affairs or rights of an individual or organisation.

Creative Commons licenses cannot be made more restrictive, and hence cannot be applied for the conditions of use identified above.

License Candidates

Table 2 provides a summary of licenses that should be available within a typical policy. Some of the more restrictive licenses may not be required in an open data policy [10], [14].

License **URL**	**Short Description**
Public Domain: CC0	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)	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)	Free to share, even for commercial purposes, provided that you give appropriate credit.
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.
Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)	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)	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)	Free to share, provided that you give appropriate credit, apply the work for non-commercial purposes only, and do not distribute modifications or derivatives. **Note that this is not a free culture license.**
Restricted: Ethics and Privacy ("No Harm")	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.

Part 2: Analysis

Methodology

Uhlir et al. (2016) [2] provide a comprehensive set of guiding principles applicable to the legal interoperability of open research data. Many of these are applicable as a basis for policy formulation. The OECD (2017) [3] provides an equally comprehensive review of the considerations informing open data policy and the principles that should be applied. CoDATA has published guidance in respect of data policy [25], and so have several of the major funding agencies [28], [29], [30], [36]. The Belmont Forum has also been active in consolidating data policy guidance for its members and stakeholders [31]. It is also common for global infrastructures, global initiatives, and individual projects or research institutions to publish policies. All this makes for a truly divergent and complex landscape in need of standardisation:

1. Principles, scope, and topic coverage obtained from review studies, such as those mentioned above.
2. General policies published by international bodies and initiatives in our field of study, such as GEO [37], Future Earth [38], ICS [39], WDS [40] , CEOS [41], IOC [42], and similar.
3. Review against detailed policies from specific institutions and networks in our field of study - ICOS [43], SAEON [44], ILTER [45], TERN [46], IMOS [47], NEON [48], and DataOne [49] amongst others.

In addition, several international initiatives and organisations have published guidelines in respect of data management - these include the GEO Data Management Principles [32], FAIR principles [33], CoreTrustSeal certification criteria [34], and declarations such as the Bari Manifesto [35]. All of these, to some degree, have implications for data management and policy.

From these sources, we have compiled the following analyses:

- Principles and implications of those principles for policies, procedures, and systems.
- Elements covered by data policies.
- Roles and responsibilities identified in policies.
- Assessment of a large number of example policies in respect of principles and elements covered, and the roles and responsibilities assigned.
- A set of policy design considerations that can be used to synthesize generic policy candidates.

Principles and Implications

In this section, we consider the principles that can be considered for inclusion in a policy to cover the management and dissemination of research outputs, and what these principles imply for implementation, policy objectives, and infrastructure [5],[24],[25],[33] [38],[40],[47],[48],[52],[54] .

#	Principle	Discussion	Implications

1	Research Outputs and Government Data are Public Assets	They require public funding to create, they have a replacement value, and require funding for maintenance and preservation	Maximise ROI Preservation Infrastructure
2	Reproducibility is a Major Objective ("Open Science")	The process of science relies on peer review and incremental rectification of incorrect findings.	Accessibility Re-usability
3	Open by Default	"Free access to, and subsequent re-use of, open data are of significant value to society and the economy".	National Policy
4	Quality assured, full disclosure	Release data and supporting information and metadata after due diligence and quality assurance has been performed, and the data is described properly. Maximum granularity	Ethics Release Event
5	Free and Universal Access, Usable by All	Data should be usable by all, not discriminating on the basis of societal sector. In essence, data should be FAIR.	Machine Readable Human Readable Interoperable Limit Impediments Discoverable Accessible Re-usable
6	Transparency	Publication of methods, standards, and protocols, all relevant data, code, and support materials included	All research outputs Re-usability
7	Realisation of Benefits, Positive Outcomes	Re-use of data for improved research and innovation ROI, economic benefits, governance benefits, societal benefits in respect of planning, preparedness, disaster response.	Maximise ROI

Elements of Research Output Policy

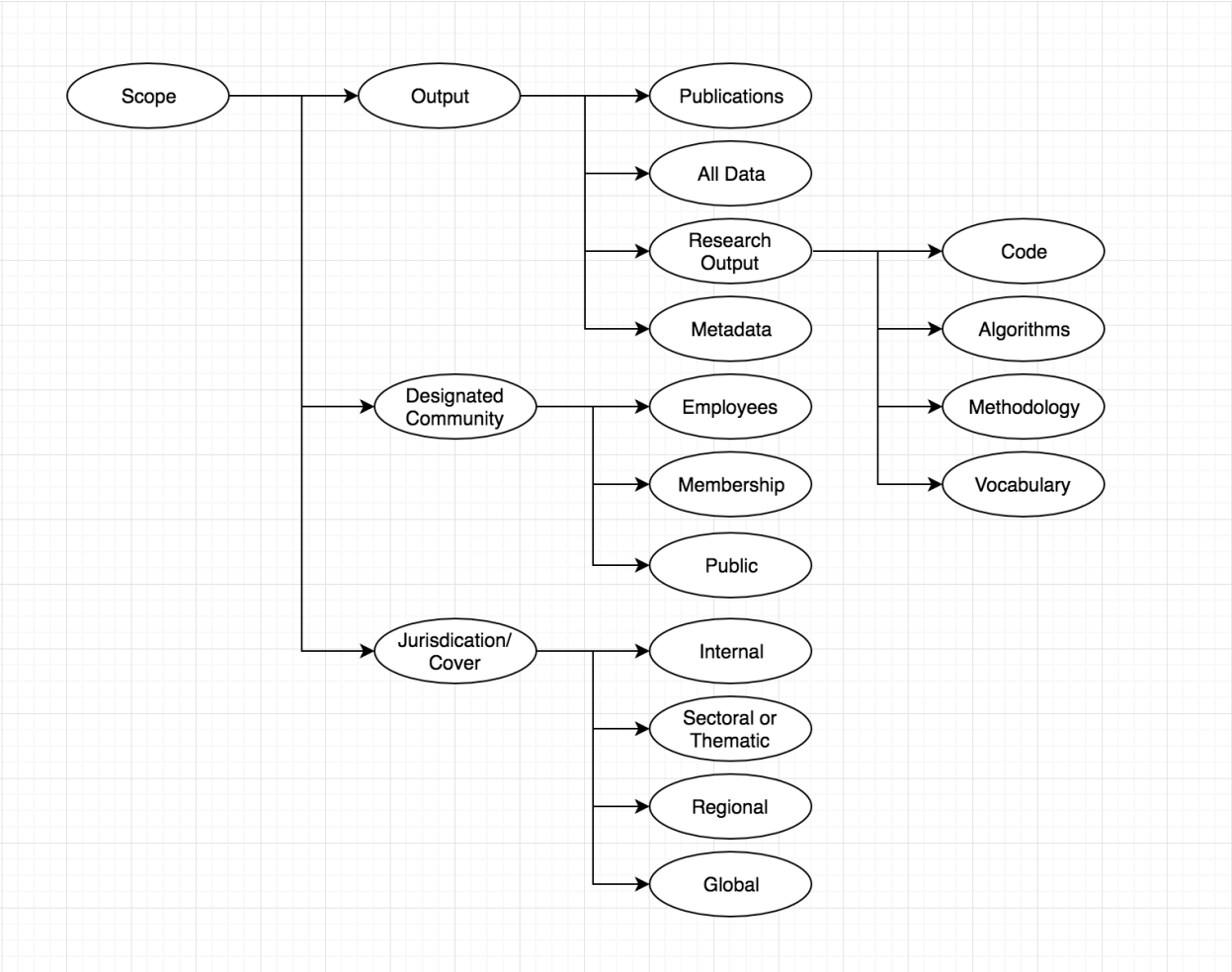
The following analysis is based on assessment and classification of 13 policy statements in the domain of earth and environmental observation [5],[24],[25],[38],[40],[42],[47],[48],[54]. The frequency reported indicates the relative adoption rate of each element, and if two values are reported, the first applies to the entire scope of the policy (Stakeholders and products or services). If a second frequency is reported, it applies to a designated set of stakeholders or products only.

The elements are based on a hierarchy, this is indicated by a numbering system in the leftmost column. The most detailed elements are framed in language that can be used as elements of a data policy.

Some elements, if adopted, imply that other elements must also be present. This aspect is addressed by cross-references following on the element definition.

Policies apply to publicly funded outputs in the research and government data landscape, and include research outputs, products, and services, dependent on the scope of the policy. See Annexure C for more detail on this aspect.

Scope

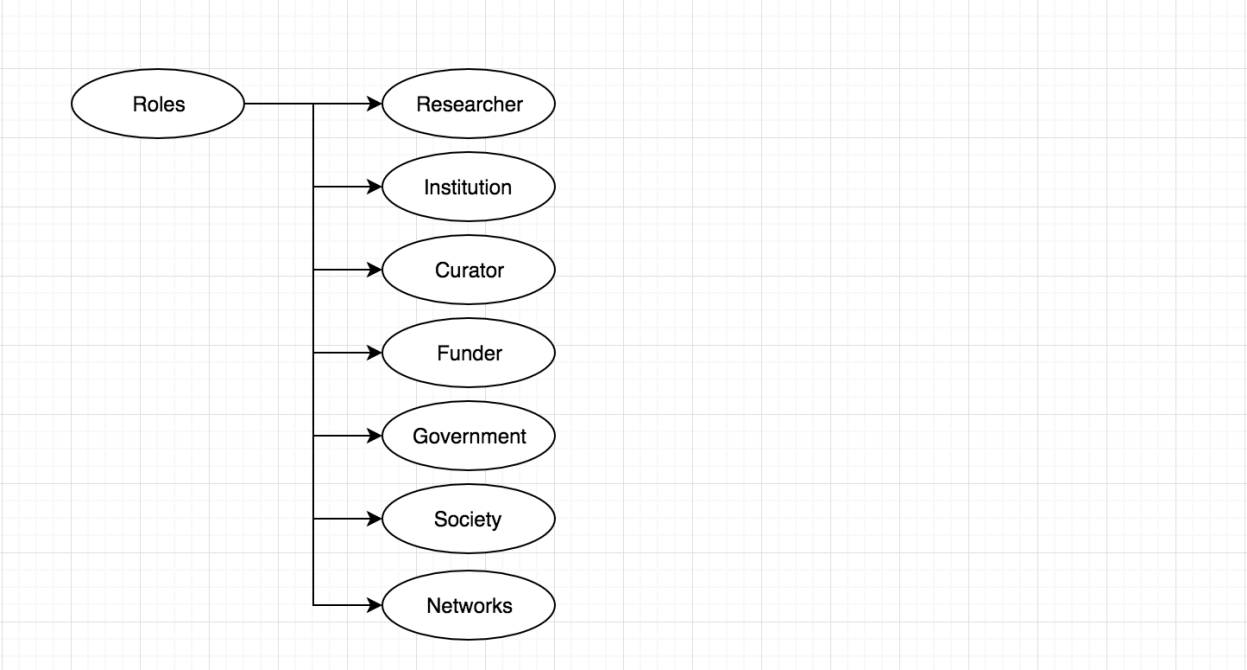


#	Element	Element Definition/ Discussion	Freq
1	Scope	The scope of a policy needs to be defined. Most policies deal with data only, but this is not always the case, and sometimes multiple types of research output or services are included. Furthermore, most policies define a designated community that needs to adopt or adhere to the policy. Finally, a policy has a jurisdiction or a coverage.	100%
1.1	Output	Responsibilities include the manner in which creators and custodians deal with research outputs (data, services, code, …)	100%
1.1.1	Publications	All scholarly publications and the data required to support or reproduce the findings in such a publication is covered by the policy.	46% 46%

1.1.2	Data	All data products, including raw data, are covered by the policy	100%
1.1.3	All Research Output	All research outputs, covering scholarly publications, and data, code, algorithms, services, vocabularies, and methods are covered by the policy	46% 46%
1.1.4	Metadata	Comprehensive, community accepted, and standards-compliant metadata will be maintained and made available for purpose of discovery	54% 15%
1.2	Designated Community	The designated community for the policy.	100%
1.2.1	Employees	Employees are often the designated community for an internal policy, but it is very common to extend the reach to collaborators, subcontractors, and service providers.	0% 31%
1.2.2	Membership	Membership of an organisation, grouping (such as government or academia), network, or initiative requires individuals or institutions to subscribe to a policy.	23% 31%
1.2.3	Public	Policy is applicable to any member of the public that participates in or makes use of the services or products covered by the policy.	54% 8%
1.3	Jurisdiction/ Cover	The applicability of a policy is often limited in respect of a region or jurisdiction. Sime limitations apply due to a thematic or domain consideration (e.g. space science, marine domain, etc.)	100%
1.3.1	Internal	The scope of a policy if often internal - no jurisdiction or effect outside the organisation.	0 31%
1.3.2	Sectoral or Thematic	The policy applies only to the formal or informal membership of a thematic or domain grouping (e.g. all taxonomists, all educational data repositories, all funding agencies, marine domain…)	0% 8%
1.3.3	Regional	The policy applies to a community defined by a region - administratively or virtually - such as a country or continent, a regional infrastructure, etc.	15% 15%
1.3.4	Global	The policy applies to all stakeholders globally	46% 15%

Role-Players

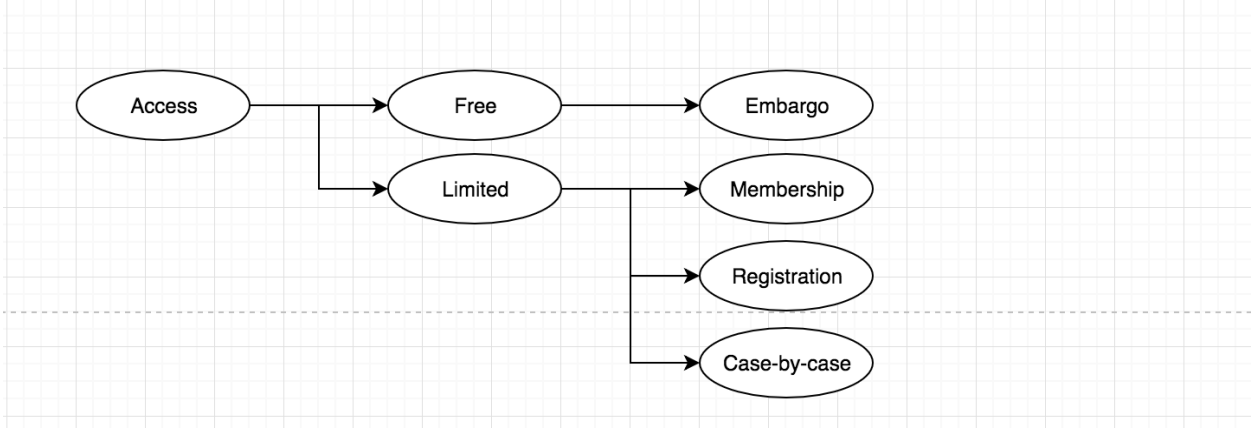
Roles and responsibilities are discussed in more detail in a following section.



#	Element	Element Definition/ Discussion	Freq
2	Role-Players	Role-players are not the same as stakeholders, and within a given stakeholder community one or more roles can be defined.	100%
2.1	Researchers (Creators) (Authors)	Researchers (or other designated individuals) are most often the creators or owners of the products and services that are covered by the policy. Usually, they need to conform to policy expressed in funding conditions during the lifetime of the project Some responsibilities may remain after the completion of the project.	46% 15%
2.2	Institutions	Ensure that policy requirements are executed by grant-funded lead researchers and that this effort is adequately supported; Develop institutional policies and and guidance, particularly for creation and execution of data management plans and proof of deposit; Depending on national data infrastructure availability or disciplinary repositories, the research performing organisation may also need to provide long-term stewardship for some data.	31% 0%
2.3	Funders	Develop and communicate Research Data Management (RDM) policy Provide advice directly or through services and intermediaries Review implementation and measure compliance.	-
2.4	Curator	Provide long-term stewardship for specific data in accordance with funder policies Provide guidance and support according to role designated by funder Meet preservation obligations on behalf of	-
2.5	Government	Government often has a role in respect of open access to government data, and this is increasingly the norm in the developed world. They also	23% 8%

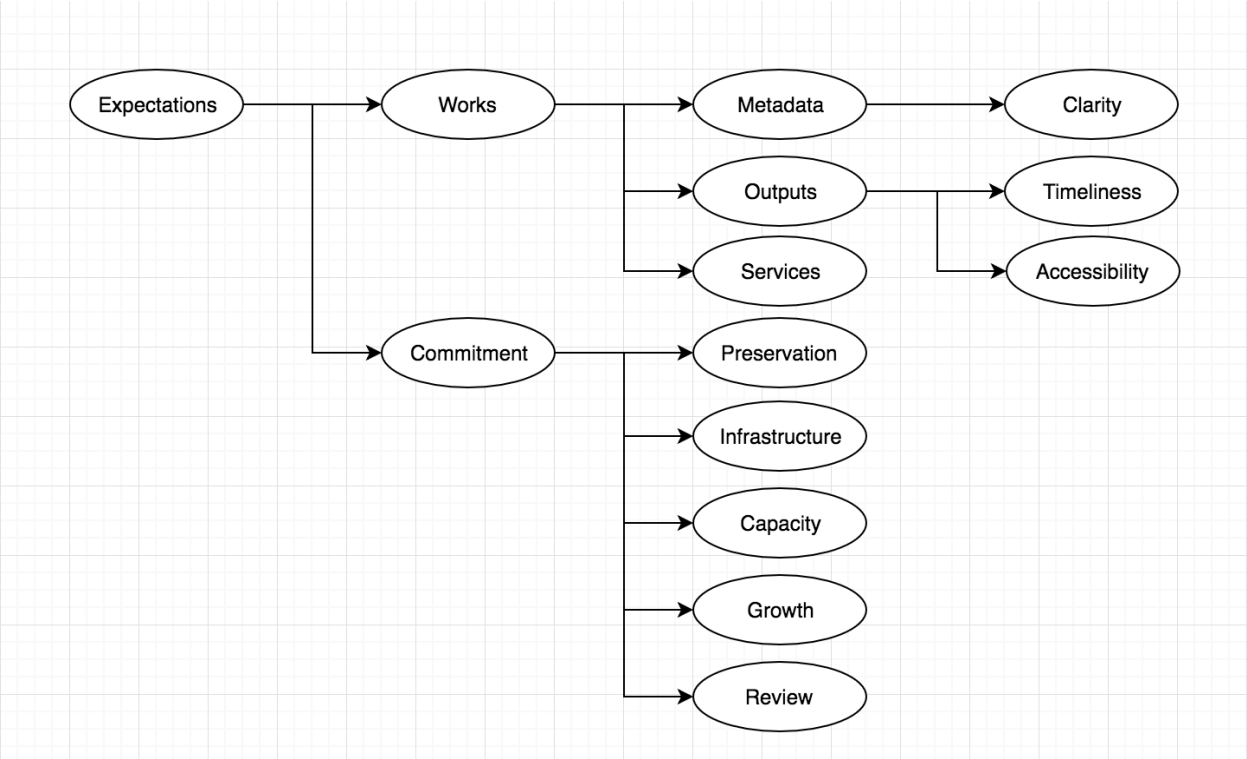
		ultimately provide funding for much of research and research infrastructure.	
2.6	Society	Validates and uses products and services covered by the policy, and can be a source of feedback	-

Access



#	Element	Element Definition/ Discussion	Freq
3	Access Options	Policy needs to define the principles guiding access, and the nature of such access. Note that the specifics of access and application is defined by an appropriate license.	100%
3.1	Free	Free and open access is a common principle in most policies. Note that 'free' is not used in a monetary sense, but refers to the fact that everyone has equal freedom of access to outputs or products.	69% 15%
3.1.1	Embargo	Products and services will be embargoed for a sensible period to allow the creators of the research outputs time to publish their findings. Access will not be withheld unreasonably beyond the initial embargo period unless a motivation for extension is provided,	23% 0%
3.2	Limitation	Limitations to access are usually in place to prevent breach of privacy and ethics, or to protect commercial interests.	0 31%
3.2.1	Membership	Access is restricted or limited to accredited or registered members only. Membership may or may not involve payment of fees.	-
3.2.2	Registration	Access is only permissible for registered end users. This is a common policy requirement and is generally needed to maintain detailed usage statistics for funders and data providers.	-
3.2.3	Merits	Applications to use data are evaluated on a case-by-case basis, and if policy guidelines do not specify the evaluation process, it could be subjective and arbitrary.	-

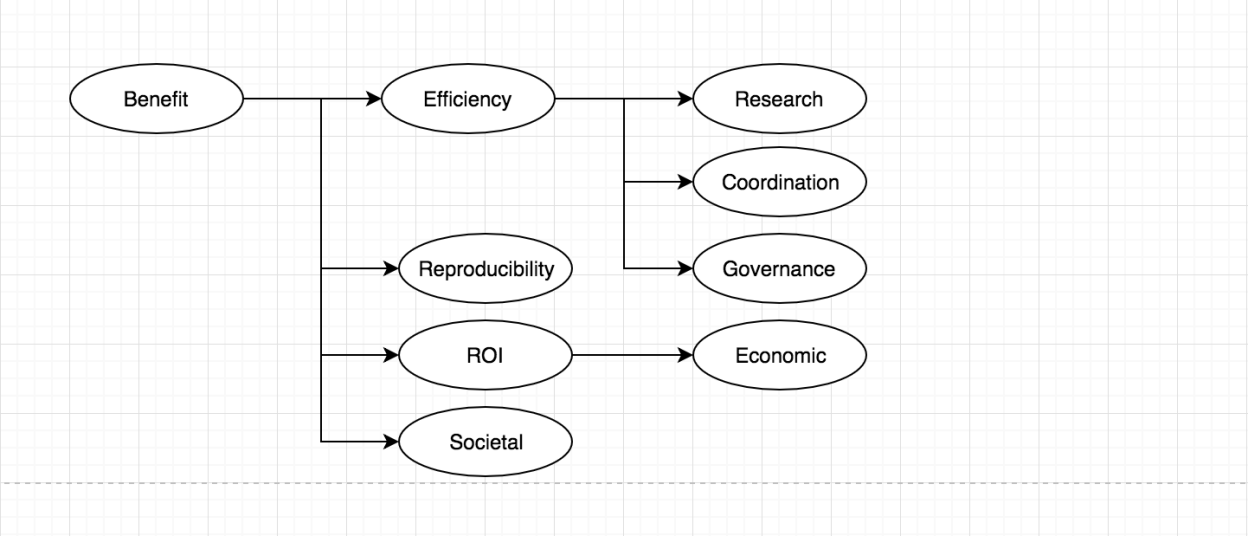
Expectations



#	Element	Element Definition/ Discussion	Freq
4	Responsibilities/ Expectations	Policy generally places responsibilities and expectations on the target audience for the policy, whether that be members of a network or initiative, employees in an organisation, or end users of a service.	100%
4.1	Works	The works provided (research outputs, services, products) will meet expectations expressed by the policy	100%
4.1.1	Metadata/ Discovery	Comprehensive, community accepted, and standards-compliant metadata will be maintained and made available for purpose of discovery	54% 15%
4.1.1.1	Clarity	Metadata will be adequate to clearly indicate the provenance of the data product, and how it could be applied	8%
4.1.2	Research Outputs	Research outputs, as specified in the scope of products and services, will be made available with an appropriate license	46% 46%
4.1.2.1	Timely Release	The creators of products will ensure its timely release, limited only be quality assurance, metadata completeness, and defensible embargo considerations	54% 8%

4.1.2.2	Accessibility	Products and services are accessible after discovery of such services via metadata	8%
4.1.3	Services	Data services and processing services will be made available with an appropriate license	38% 46%
4.2	Commitment	Many policies specify the commitments that signatories of the policy undertake. These may involve infrastructure, long-term sustainability and preservation, or commitment to community-directed activities, such as outreach, communication, and review.	77% 0%
4.2.1	Capacity	In addition to products and services, the policy supports capacity development and assistance in the designated community.	15% 8%
4.2.2	Preservation	Long-term preservation is in place, and the signatories to the policy undertake to preserve research outputs based on a preservation plan.	62%
4.2.3	Infrastructure	The signatory to the policy will ensure availability of adequate ICT infrastructure in respect of hosting, security, and availability	31%
4.2.4	Growth	The signatories of the policy commit to promotion of the aims of the policy and to assist with growth in policy adoption	8% 8%
4.2.5	Feedback and Review	End users will be afforded a mechanism for provision of feedback and review, and appropriate action will be taken.	8%

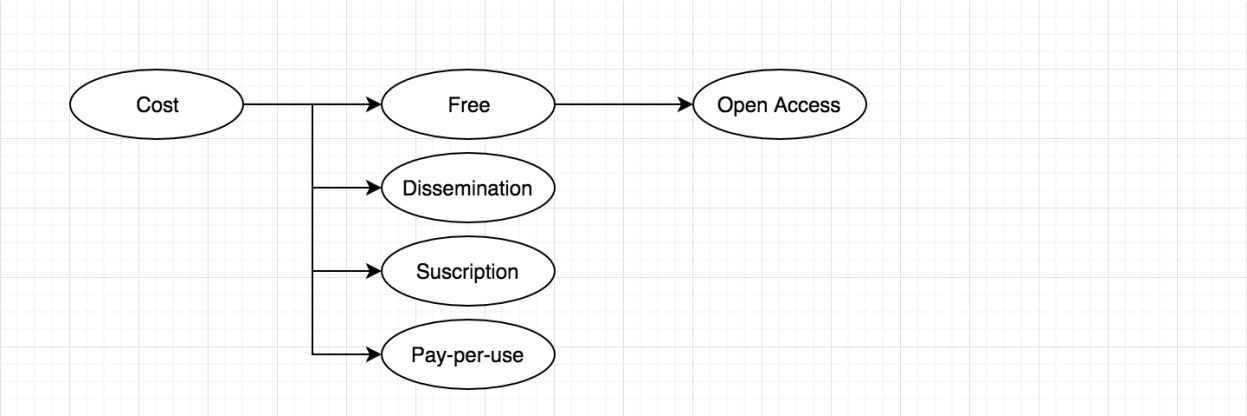
Benefits



#	Element	Element Definition/ Discussion	Freq
5	Benefits	Policies are universal in their aim to improve one or more outcomes for their stakeholders.	100%

5.1	Efficiency	Policy aimed at improving efficiency.	46% 0%
5.1.1	Research Efficiency	Policy is aimed at improving research efficiency - re-use of data, verification and peer review, reproducibility	31% 0%
5.1.2	Improved Coordination	Research coordination (avoidance of duplication of effort) is achieved by making comprehensive metadata available on existing and future products and services.	0% 15%
5.1.3	Improved Governance	Policy is aimed at improving the governance of a stakeholder group	8% 0%
5.2	Reproducibility	Open access to data in support of scholarly publication will enable published research to be verified and improve scientific rigour.	8% 0%
5.3	Return on Investment	Ensure that research outputs, including data, is accessible and re-usable to maximise return from publicly funded research.	23% 0%
5.3.1	Economic Benefit	Specifically stimulate the knowledge economy by making scientific research outputs easily and freely available.	8% 0%
5.4	Societal/ Community Benefit	Improve society by supporting knowledge dissemination: decision-making, policy, economic, and planning benefits. Data policy is sometimes aimed at supporting or serving a community. The elements of such benefits are stated in the policy.	62% 15%

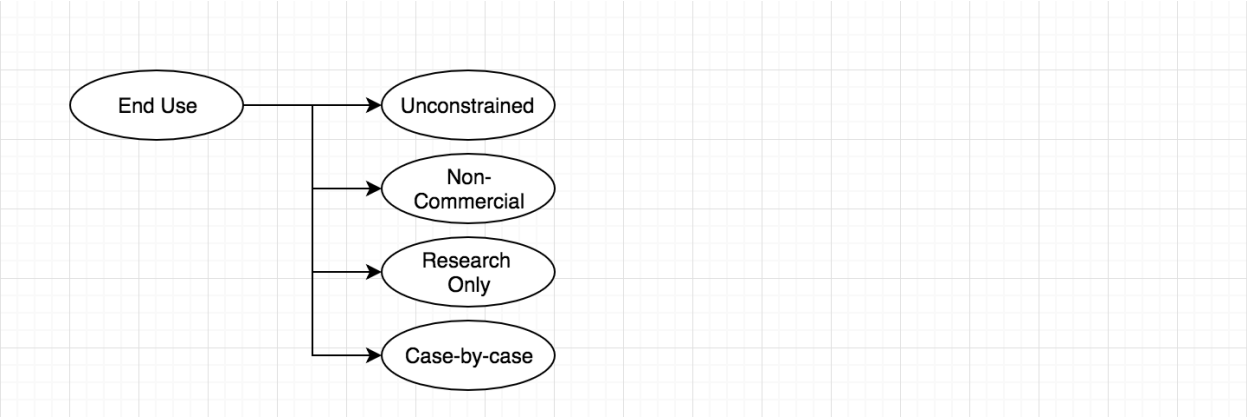
Costs



#	Element	Element Definition/ Discussion	Freq
6	Costs	Policy makes provision for a variety of cost recovery mechanisms.	100%
6.1	Free	Many policies explicitly state that research outputs should be available without any costs (the sustainability of this is dependent on funding for such access).	62% 15%

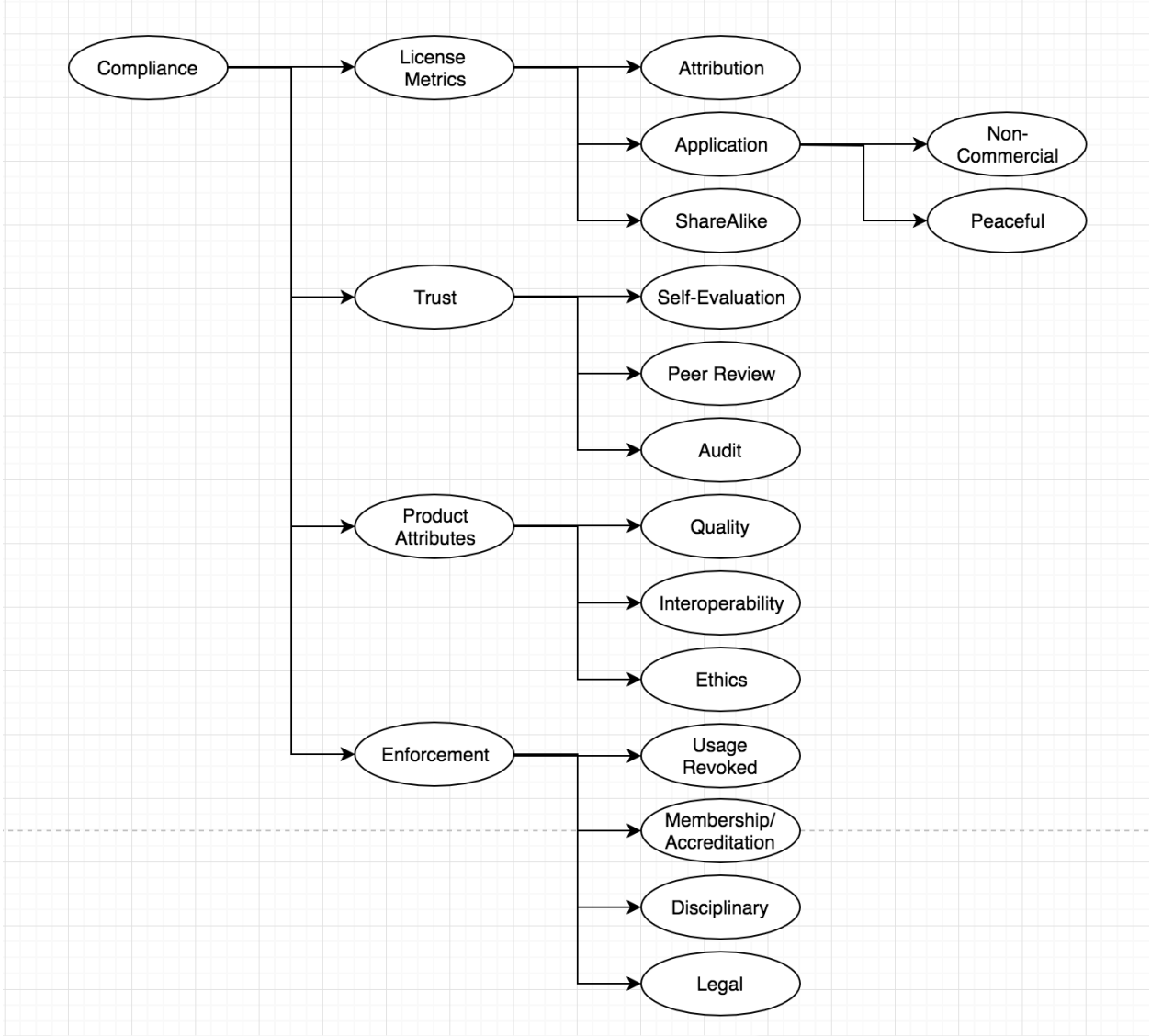
6.1.1	Open Access	'Open Access' accomplished the goal of having cost-free access for the end user, but costs of publication and dissemination is borne by the creator on publication.	69% 0%
6.2	Dissemination Costs	Some policies require the end user to fund dissemination costs, and these should be limited to the cost of download/ dissemination.	46% 15%
6.3	Subscription	Policy makes provision for access to research outputs, products, or services on the basis of a subscription	-
6.4	Pay-per-Use	Policy makes provision for access to research outputs, products, or services on a pay-per-use basis	-

End Use



#	Element	Element Definition/ Discussion	Freq
7	End Use	End uses are sometimes limited by data policy and applicable licenses	100%
7.1	Unconstrained	Policy allows application of products, services, and research outputs in any context, provided applicable license conditions are adhered to.	-
7.2	Non-commercial	Policies may require end users not to use the research outputs, services, or products for commercial gain	8% 15%
7.3	Research Only	Research outputs, products, and services may only be used for research purposes	-
7.4	Case-by-case	Proposed end use is evaluated on a case-by-case basis.	-

Compliance



#	Element	Element Definition/ Discussion	Freq
8	Compliance	All policies indicate the manner in which compliance will be measured, enforced, and how compliance will be promoted or non-compliance dealt with	100%
8.1	License Attributes	Compliance with license provisions is often required and metrics for evaluation of compliance can be defined	
8.1.1	Attribution	This a a very common license provision, and requires end users to cite the sources that are used for new works.	
8.1.2	Application	Policy on occasion controls or attempts to control the end use of the product or service.	

8.1.2.1	Non-Commercial	Policies may require end users not to use the research outputs, services, or products for commercial gain	8% 15%
8.1.2.2	Peaceful use	Policy defines acceptable use, including use only for peaceful purposes	0 8%
8.2	Trust	Policies are often aimed at improving trust between partners - in this instance largely between the providers and consumers of works (products, services, research outputs).	-
8.2.1	Self-Evaluation	The policy requires participants to comply with the policy through self-evaluation	8% 0%
8.2.2	Peer Review	Policy compliance is evaluated by way of peer review	-
8.2.3	Audit	Audit is performed to determine policy compliance. This could take the form of periodic review of compliance by auditors or spot checks/ sampling of compliance indicators.	-
8.4	Attributes	Policy dictates or provides guidance on the desirable attributes of works	100%
8.4.1	Quality	Quality assurance and review prior to publication of dissemination is a common aspect of policy. The creators of research outputs and services will ensure that the products are quality assured to community standards and expectations	54% 0%
8.4.2	Interoperability	Datasets, metadata, research outputs, and services, as applicable, will conform with interoperability standards endorsed by the community	54% 0%
8.4.3	Ethics	The publisher or curator of the product will ensure that relevant privacy and ethics considerations are adhered to	15% 0%
8.5	Enforcement/ Incentives	Policy sometimes provide measures for enforcement or compliance, or consequences of non-compliance. This is at times congruent with incentives for adhering to policy	100%
8.5.1	Usage Revoked	Policy can make provision to terminate the access or end use due to non-compliance	-
8.5.2	Membership/ Accreditation	Policy compliance can be viewed as an incentive if it determines continued membership of an initiative or network, and as a deterrent if membership can be revoked due to non-compliance.	0% 0%
8.5.3	Disciplinary Action	Disciplinary action is most often the consequence of non-compliance for internal policies, but less common for policies aimed at networks or initiatives	-
8.5.4	Legal Action	Legal action is highly unlikely in the case of free culture or open access licenses and the policies that govern them. It may be appropriate should the license be restrictive, especially of commercial, privacy, or ethics considerations are at stake.	0% 0%

1.2.3	Growth	The signatories of the policy commit to promotion of the aims of the policy and to assist with growth in policy adoption	8% 8%
1.3.6	Feedback and Review	End users will be afforded a mechanism for provision of feedback and review, and appropriate action will be taken.	8%

Roles and Responsibilities

#	Stakeholder	Typical Responsibilities	Typical Actions	References
1	Funders	Responsible for determining the policy requirements associated with funding and grants. In many jurisdictions, publicly funded research are now required to be piublished with open licenses.	Policy formulation Compliance Monitoring	[25]
2	Individual Researchers	Comply with grant requirements in respect of research outputs, and ensure that adequate metadata is produced for citation, re-use, and discoverability	Create/ produce research output(s)	[25]
3	Research Institutions	Encourage open publication of research outputs, provide infrastructure to allow researchers to publish and disseminate research outputs. Formulate institutional policies aligned with international best practice.	Policy formulation Publication Compliance Monitoring	[25]
4	Data Services/ Repositories/ Curators	Provide mechanisms for citation support, promote open licenses for content, and confirm that depositors have the rights to allocate a license to the work. Monitor that end users are aware of license requirements. Publish datasets and research outputs on behalf of creators.	Hosting services and product repositories Curation Compliance Monitoring	[25]
5	International, disciplinary, or regional networks and initiatives	Promote open publication of research outputs, and develop network or programme policies in support of open publication. Assist the community with development of interoperability specifications and standards. Promote license compliance.	Policy guidance Standards and Specifications Compliance Monitoring	
7	Society at Large	Provide appropriate citation credit and respect licence provisions	Adhere to license conditions Cite use of works	

Part 3: Synthesis

This section consolidates our analysis of the policy landscape with a view to synthesis of one or more generic policies.

Design Considerations

1. All data, information, and research outputs generated by publicly 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. Institutions should align policies with such national policies, but for multinational or global initiatives, such alignment could prove difficult. For this reason, such initiatives need to provide more than one policy candidate that can be adopted by participants.
3. Access control measures 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 intellectual property legislation;
    e. Infringement of current or future legal rights to exploit natural resources.
    f. In developing country context, provision ahould be made for a slate of licenses that range from free culture licenses to licenses restricting access on the basis of national or cultural intellectual property. This last provision should be the exception and not the rule.
4. The number of licenses applicable to a policy or required to implement a policy should be limited in number, based on international best practice, and be machine readable.

Access Control Process Model

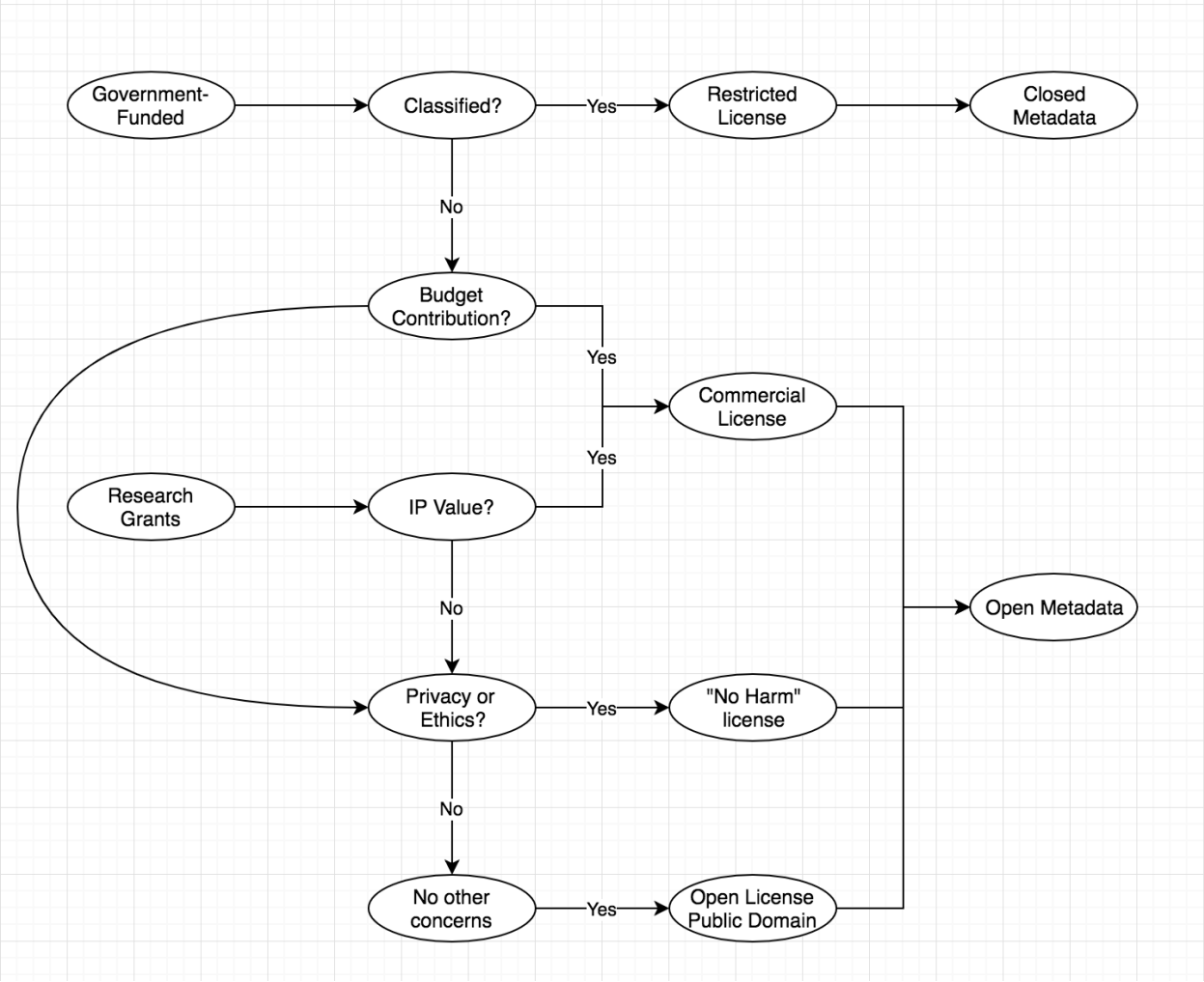
The process model makes provision for the following states of access:

1. State-generated data that is classified: not openly and freely accessible, and no meta-data is openly published.
2. Publicly funded outputs for which access is controlled, due to four situations:
    a. Grant-funded research that has a commercial exploitation potential that far outweighs its general public good, and is offered protection under Intellectual Property Rights legislation in various jurisdictions;
    b. Publicly funded outputs for which access is controlled because of conservation considerations, privacy issues, or legal rights of third parties.
    c. In all of these cases, it remains good practice to publish the metadata openly, unless the outputs are classified: this means that the outputs are discoverable but may not openly or freely accessible.
    d. 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.

3. The balance of grant-funded or state-generated outputs is freely and openly available and meta-data is published openly.

Figure 4: Access Control Process Model



Candidate Policies

From assessment of the analysis results, once can determine that there are at least three policy candidates, depending on the role-players:

1. A ***internal policy*** that can be adopted by any research institution;
2. A ***network*** or ***member policy*** for member institutions that participate in a multinational initiative such as SEACRIFOG;
3. An ***end user policy*** for the beneficiaries of the SEACRIFOG initiative.

The expectations and actions for each of these stakeholder groups will be different, and the consequences of non-compliance with the policy will also be different.

Candidate 1: Institutional Policy

#	Policy Element	Description
1	Why is the policy required?	Policy is required to ensure that research outputs produced by the institution are managed properly, providing maximum benefit from publicly funded research.
2	Policy Statement	All research outputs produced at the institution will be published at the earliest opportunity, with standards compliant metadata, and made available with the least restrictive license applicable to the situation.
3	Objectives	1. Obtaining maximum benefit from publicly funded research through re-use of outputs; 2. Allowing verification of published results; 3. Contributing to evidence-based societal benefit through decision, policy, and planning support; 4. Preserving research outputs for future use, especially in cases where observed data cannot be reproduced; 5. Maximising utility for end users through provision of standardised metadata and standardised data schema or services, appropriate to the scientific discipline and community; 6. Allowing appropriate time for researchers to publish, but otherwise limiting embargo periods to what is absolutely necessary[7].
4	Mandate	The policy is authorised by the Chief Executive Officer/ Managing Director/ President of the institution.
5	Applicability	The policy applies to all research outputs produced at the institution, and that are funded publicly. These outputs include, but are not limited to 1. Datasets and services; 2. Code and algorithms; 3. Methodologies, experiment designs, and protocols; 4. Vocabularies, conceptual models and ontologies; 5. Scholarly publications; 6. Research infrastructure and platform information.
6	Term	The policy is effective from date of publication, and applies to all research commencing after the date of publication. The policy will be reviewed annually.
7	Compliance	Researchers that fail to comply with policy will be subject to institutional disciplinary procedures.
8	Planning and Costs	Researchers should include costs of research output management, including open access publication costs, computing costs, data storage costs, and future presrvation costs into the calculation of funding requirements for the research project. Data Management Plans can assist with this process.
9	Detailed Responsibilities	1. The institution will provide researchers with

[7] Guideline embargo periods are typically 18 months

		a. Infrastructure for the hosting and publication of research outputs, or designate community or discipline-specific repositories that researchers can use; b. Data Management Planning advice and services, or indicate which publicly available services to use; c. Curation services to assure preservation, archiving, and assist with standardised publication of research outputs. 2. The institution will a. Monitor compliance of researchers with the policy provisions; b. Review the policy on an annual basis to determine continued applicability; c. Select and recommend open, machine readable licenses for researchers to apply, and provide guidance on license selection; d. Determine a mutually agreed embargo period that allows researchers adequate time to publish research before making research outputs publicly available; e. Maintain an inventory of research infrastructure and observation platforms as agreed with network organisations or compliant with community standards; f. Determine portfolios of applicable metadata and data standards for researchers to use, including directives in respect of minimum mandatory metadata. 3. Individual researchers will a. Select the least restrictive license applicable to the work; b. Provide minimum mandatory metadata as per institutional directives; c. Provide information on research infrastructure and platforms compliant with institutional norms; d. Provide curators with research outputs compliant with agreed standards and formats; e. Provide proof of deposit and compliant metadata to the institution on completion of a research project.

Candidate 2: Network Policy

#	Policy Element	Description
1	Why is the policy required?	Policy is required to ensure that member organisations provide contributions to the network in an agreed and timely manner.
2	Policy Statement	All research outputs agreed by network members will be made available with minimum delay, at no cost to the network, in a format and schema compliant with the metadata and data standards agreed by the network members.
3	Objectives	In addition to objectives of individual members, the network aims to add value and benefit to researchers, policy makers, and the general public by aggregating standardised observations across different regions, time scales, and domains.
4	Mandate	The policy is authorised by the members at their annual general meeting.
5	Applicability	The policy applies to all research outputs agreed by members in duly constituted meetings, and includes 1. Datasets and services; 2. Code and algorithms; 3. Methodologies, experiment designs, and protocols; 4. Vocabularies, conceptual models and ontologies;

		5. Scholarly publications; 6. Research infrastructure and platform information.
6	Term	The policy is effective from date of publication, and applies to all member contributions submitted after the date of publication. The policy will be reviewed annually.
7	Compliance	Network membership is contingent on compliance with the policies of the network.
8	Planning and Costs	It is assumed that costs of open dissemination of research outputs, as agreed by the members, will be borne by the members.
9	Detailed Responsibilities	1. The network will provide members with a. Infrastructure for the aggregation of research infrastructure and platform information; b. Redirection services whereby datasets applicable to specific standard variables and resarch infrastructures or platforms can be obtained; c. Data and metadata infrastructure to use should members not be able to provide such infrastructure themselves. 2. The network will a. Agree a set of standard variables, applicable metadata, data, and protocol schema/ definitions that will be contributed and used by members; b. Monitor compliance of mebers with the policy provisions; c. Review the policy on an annual basis to determine continued applicability; d. Select and recommend open, machine readable licenses for members to apply, and provide guidance on license selection; 3. Individual members will a. Select the least restrictive license applicable to the agreed shared output; b. Provide minimum mandatory metadata as per network directives; c. Include links and redirection to standards-compliant data services hosted by the member (or on behalf of the member) in the metadata; d. Provide research infrastructure/ platform data in the agreed format and schema to the network, and keep this up to date with a mutually agreed frequency.

Candidate 3: End User Policy

#	Policy Element	Description
1	Why is the policy required?	Policy is required to ensure that end users comply with license provisions applicable to the research outputs provided by members via the network.
2	Policy Statement	All research outputs obtained via the network, either directly from members or indirectly from the network, must be applied in compliance with license provisions.
3	Objectives	Support of open science principles via citation and license compliance.
4	Mandate	The policy is authorised by the members at their annual general meeting.
5	Applicability	The policy applies to all research outputs agreed by members in duly constituted meetings, and includes 1. Datasets and services; 2. Code and algorithms; 3. Methodologies, experiment designs, and protocols; 4. Vocabularies, conceptual models and ontologies;

		5. Scholarly publications; 6. Research infrastructure and platform information.
6	Term	The policy is effective from date of publication, and applies to all end user activity after the date of publication. The policy will be reviewed annually.
7	Compliance	End users will acknowledge the content of licenses prior to use of the research outputs. Instances of non-compliance will be recorded by the network.
8	Detailed Responsibilities	1. The network will provide end users with a. Access to the license applicable to the research output being requested; b. A means of acknowledging familiarity with license provisions. 2. The end user will a. Acknowledge familiarity with license provisions prior to access.

References

[1] Knoth, Petr; Pontika, Nancy (2015): Open Science Taxonomy. figshare. Figure. https://doi.org/10.6084/m9.figshare.1508606.v3 [Last accessed 2019-02-01]

[2] Uhlir, Paul F. (editor) (2016): LEGAL INTEROPERABILITY OF RESEARCH DATA: PRINCIPLES AND IMPLEMENTATION GUIDELINES RDA-CODATA Legal Interoperability Interest Group September 8, 2016, http://www.codata.org/uploads/Legal%20Interoperability%20Principles%20and%20Implementation%20Guidelines_Final2.pdf [Last accessed 2019-02-01]

[3] OECD (2017): H2020 Programme - Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2, 21 March 2017. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf [Last accessed 2019-02-01]

[4] Pilat, D. and Fukasaku, Y. (2007): OECD Principles and Guidelines for Access to Research Data from Public Funding, http://www.oecd.org/dataoecd/9/61/38500813.pdf [Last accessed 2019-02-01]

[5] CoDATA (2019): List of International Organisations with Open Data Policies, codata.info http://www.codata.info/data_access/policies.html. [Last accessed 2019-02-01]

[6] https://en.wikipedia.org/wiki/List_of_countries_by_research_and_development_spending [Last accessed 2019-02-01]

[7] Various (2003): Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003. https://openaccess.mpg.de/Berlin-Declaration [Last accessed 2019-02-01]

[8] Various (2019): Review of Free Culture Licenses https://freedomdefined.org/Licenses#Copyleft [Last accessed 2019-02-01]

[9] Creative Commons (2019): State of Creative Commons. https://stateof.creativecommons.org/ [Last accessed 2019-02-01]

[10] Hugo, W and Pauw, J (2013): Notes on Data Sharing Policy in South Africa - prepared for DST. https://docs.google.com/document/d/1KnMAqwRbvfnkn0PeLlXTSjANB_DHgT4qbaEjm-HsY6k/edit?usp=sharing [Last accessed 2019-02-01]

[11] Fidler, F. & Gordon, A. (2013): Science is in a reproducibility crisis: How do we resolve it? Phys.Org, September 20, 2013. https://phys.org/news/2013-09-science-crisis.html [Last accessed 2019-02-01]

[12] Leyser, O, Kingsley, D., and Grange, J. (2017): The science 'reproducibility crisis' – and what can be done about it, The Conversation, March 15, 2017 11.49am SAST

https://theconversation.com/the-science-reproducibility-crisis-and-what-can-be-done-about-it-74198 [Last accessed 2019-02-01]

[13] Various (2019): Choose a License, https://choosealicense.com/ [Last accessed 2019-02-01]

[14] Creative Commons (2019): License Features, https://creativecommons.org/choose/ [Last accessed 2019-02-01]

[15] Creative Commons (2019): Considerations for licensors and licensees, https://wiki.creativecommons.org/wiki/Considerations_for_licensors_and_licensees [Last accessed 2019-02-01]

[16] Smit, E (2011): Abelard and Héloise: Why Data and Publications Belong Together, D-Lib Magazine
January/February 2011, Volume 17, Number 1/2. doi:10.1045/january2011-smit
http://www.dlib.org/dlib/january11/smit/01smit.html [Last accessed 2019-02-01]

[17] UCSC (1994): Guide to Writing Policy and Procedure Documents, Internal Working Draft, Policy and Procedures Team, UCSC, 2-12-1994
https://policy.ucsc.edu/resources/images/GuidetoWritingPolicy.pdf [Last accessed 2019-02-02]

[18] Murray, S. (2013): How to Write Good Policies - Policy Elements to Consider for Your Organization August 30, 2013 CyberTears.Org Dr. Shawn P. Murray, 2013 – Murray, Shawn P. https://www.slideshare.net/DrShawnPMurray/how-to-write-good-policies [Last accessed 2019-02-02]

[19] University of Colorado (2019): User Guide to Writing Policies, Office of Policy and Efficiency ope@cu.edu, https://www.cu.edu/sites/default/files/APSwritingguide.pdf [Last accessed 2019-02-02]

[20] Purdue University (2019): WRITING AND FORMATTING POLICIES AND SUPPORTING DOCUMENTS, University Policy Office. https://www.purdue.edu/policies/writing-guidelines.html [Last accessed 2019-02-02]

[21] Ndzabandzaba, C. (2015): Brief for GSDR 2015 - Data sharing for sustainable development in less developed and developing countries, Institute for Water Research, Rhodes University, South Africa. https://sustainabledevelopment.un.org/content/documents/615860-Ndzabandzaba-Data%20sharing%20for%20sd%20in%20less%20developed%20and%20developing%20countries.pdf [Last accessed 2019-02-02]

[22] Schwegmann, C. (2013): Open Data in Developing Countries. European Public Sector Information Platform Topic Report No. 2013 /02 Published: February 2012.
https://www.europeandataportal.eu/sites/default/files/2013_open_data_in_developing_countries.pdf [Last accessed 2019-02-02]

[23] De Sherbinin, A. and Hugo, W (2018): Trends in Scientific Data Management, ICSU-WDS Scientific Data Management Workshop, 17 April 2018, Brazilian Academy of Sciences, Rio De Janeiro, Brazil https://docs.google.com/presentation/d/1Te6olqJQzNKlbwePDXFx189b3-8Z5D0hMQo-vGcogdY/edit#slide=id.p32 [Last accessed 2019-02-02]

[24] WMO (1995): Resolution 40, 12th Congress of the World Meteorological Organisation, Geneva, 1995. http://www.wmo.int/pages/prog/hwrp/documents/wmo_827_enCG-XII-Res40.pdf [Last accessed 2019-02-02]

[25] Hodson, S. and Molloy, L. (2014): Current Best Practice for Research Data Management Policies, CoDATA on behalf of the Danish e--Infrastructure Cooperation and the Danish Digital Library, 2014. https://zenodo.org/record/27872#.XFVcUM8zbYJ [Last accessed 2019-02-02]

[26] Dillo, I, Gärtner-Roer, I, Hugo, W. (2018): "Thoughts on the Future of Trust", Draft DIscussion Paper prepared for the ICS World Data System Scientific Committee and presented at the biannual Members' Forum, Gaborone, Botswana, November 2018. https://docs.google.com/document/d/1YXbYQ7-DBn0Xq-CKQzHocdc7KEUNY8aeGIt1qC9E8wk/edit?usp=sharing [Last accessed 2019-02-02]

[27] Open Government Partnership (2019). https://www.opengovpartnership.org [Last accessed 2019-02-02]

[28] NSF (2017): Chapter XI - Other Post Award Requirements and Considerations https://www.nsf.gov/pubs/policydocs/pappg17_1/pappg_11.jsp#XID4 [Last accessed 2019-02-02]

[29] NSF (2018): Data and Sample Policy, Division of Earth Sciences, National Science Foundation, Division of Earth Sciences https://www.nsf.gov/geo/geo-data-policies/ear/ear-data-policy-apr2018.pdf [Last accessed 2019-02-02]

[30] EU (2017): H2020 Programme - Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 Version 3.2 21 March 2017 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf [Last accessed 2019-02-02]

[31] Belmont Forum (2019): Data Policy and Principles http://www.bfe-inf.org/about/data-policy-and-principles [Last accessed 2019-02-02]

[32] GEO (2015): Data Management Principles (DMP) Implementation Guidelines: Life-Cycle Data Management Principles. 2015. Group on Earth Observations. https://www.earthobservations.org/documents/geo_xii/GEO-XII_10_Data%20Management%20Principles%20Implementation%20Guidelines.pdf [Last accessed 2019-02-02]

[33] Wilkinson, M. et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 2016/03/15/online, vol 3, 160018 https://doi.org/10.1038/sdata.2016.18 [Last accessed 2019-02-02]

[34] CTS (2016): Core Trustworthy Data Repositories Requirements, https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf [Last accessed 2019-02-02]

[35] Hardisty, A. et al. (2019): The Bari Manifesto: An interoperability framework for essential biodiversity variables, Ecological Informatics Volume 49, January 2019, Pages 22-31 https://doi.org/10.1016/j.ecoinf.2018.11.003

[36] Productivity Commission (2017): Data Availability and Use: Overview & Recommendations, Report No. 82, Canberra https://www.pc.gov.au/inquiries/completed/data-access/report/data-access-overview.pdf [Last accessed 2019-02-02]

[37] GEO (2015): GEO Strategic Plan 2016-2025: Implementing GEOSS, http://earthobservations.org/documents/GEO_Strategic_Plan_2016_2025_Implementing_GEOSS.pdf [Last accessed 2019-02-02]

[38] Future Earth (2013) Future Earth Initial Design: Report of the Transition Team. Paris: International Council for Science (ICSU). http://futureearth.org/sites/default/files/Future-Earth-Design-Report_web.pdf [Last accessed 2019-02-04]

[39] ICS (2018): Advancing science as a global public good - High level strategy https://council.science/cms/2018/06/HLS_EN.pdf [Last accessed 2019-02-04]

[40] WDS (2015): Data Sharing Principles, http://dx.doi.org/10.5281/zenodo.34354 [Last accessed 2019-02-04]

[41] CEOS (2019): Data Policy Portal, http://www.ceos-datapolicy.org/en/data_policy/ [Last accessed 2019-02-04]

[42] IOC (2003): IOC Oceanographic Data Exchange Policy https://www.iode.org/index.php?option=com_content&view=article&id=51&Itemid=95 [Last accessed 2019-02-04]

[43] ICOS (2015): ICOS Data Policy https://www.icos-ri.eu/sites/default/files/cmis/ICOS%20RI%20Data%20Policy.pdf [Last accessed 2019-02-04]

[44] SAEON (2017): SAEON Data Policy, https://docs.google.com/document/d/1Ts6pv7oxpYG96UVQKzwSks0a7vTcCvQSNubfJhGS3Ck/edit?usp=sharing [Last accessed 2019-02-04]

[45] Vanderbilt, K. L., C.-C. Lin, S.-S. Lu, A. R. Kassim, H. He, X. Guo, I. San Gil, D. Blankman, and J. H. Porter. 2015. Fostering ecological data sharing: collaborations in the International Long Term Ecological Research Network. Ecosphere 6(10):204. http://dx.doi.org/10.1890/ES14-00281.1 [Last accessed 2019-02-04]

[46] TERN (2015): TERN Data Licensing Policy V3.0 https://www.tern.org.au/rs/7/sites/998/user_uploads/File/Data%20Licensing%20Documents/TERN%20Data%20Licensing%20Policy%20v3_0.pdf [Last accessed 2019-02-04]

[47] IMOS (2016): IMOS Data Policy, v2. http://imos.org.au/fileadmin/user_upload/shared/IMOS%20General/Framework_Policy/2016_May_update/4.2_IMOS_Data_Policy_May16_Final_14062016.pdf [Last accessed 2019-02-04]

[48] NEON (2017): NEON Data Usage and Citation Policy http://data.neonscience.org/data-policy [Last accessed 2019-02-04]

[49] O'Brien, M. et al. (2019) Sharing data: legal and policy considerations, DataOne, https://www.dataone.org/best-practices/sharing-data-legal-and-policy-considerations [Last accessed 2019-02-04]

[50] Coalition-S (2018): Open Access is Foundational to the Scientific Enterprise https://www.coalition-s.org/why-plan-s/ [Last accessed 2019-02-04]

[51] European Commission (2018): 'Plan S' and 'cOAlition S' – Accelerating the transition to full and immediate Open Access to scientific publications. 4 Sep 2018. https://ec.europa.eu/commission/commissioners/2014-2019/moedas/announcements/plan-s-and-coalition-s-accelerating-transition-full-and-immediate-open-access-scientific_en [Last accessed 2019-02-04]

[52] LTER (2017): Long Term Ecological Research Network Data Access Policy https://lternet.edu/data-access-policy/ [Last accessed 2019-02-04]

[52] Michener, W. (2015): Ecological data sharing, Ecological Informatics, Volume 29, Part 1, 2015, Pages 33-44, ISSN 1574-9541, https://doi.org/10.1016/j.ecoinf.2015.06.010. (http://www.sciencedirect.com/science/article/pii/S1574954115001004) [Last accessed 2019-02-04]

[53] Normile, D. (2018): China asserts firm grip on research data, https://www.sciencemag.org/news/2018/04/china-asserts-firm-grip-research-data?r3f_986=https://www.google.com/ [Last accessed 2019-02-04]

[54] UK Government, (2013): G8 Open Data Charter and Technical Annex, https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex [Last accessed 2019-02-05]

[55] Creative Commons (2017): State of the Commons, https://stateof.creativecommons.org/ [Last accessed 2019-02-05]

[56] Create Commons (2002): Creative Commons Unveils Machine-Readable Copyright Licenses, https://creativecommons.org/2002/12/16/creativecommonsunveilsmachinereadablecopyrightlicenses/ [Last accessed 27-02-2019]

Annexure A: Legal Aspects in Selected Jurisdictions

A.1 South Africa

Adapted from [9].

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. Promotion of Access to Information Act, 2000 (Act No. 2 of 2000):
4. Protection of Personal Information Act: policy or regulation required to declassify and exempt the data.
5. 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 legislation but the potential to abuse legislation and unduly or unfairly restrict access.

Annexure B: Licenses

Based on inputs from 'Freedom Defined' [8], Creative Commons [12]

B.1 Research Outputs

The following research outputs should be considered for licensing:

1. Scholarly Publications
2. Reports and Studies
3. Datasets and Services
4. Vocabularies and Name Services
5. Code
6. Algorithms
7. Protocols and Methodologies

B.2 Options for Licensing and Applicability

B.3 Criteria for Selection of Licenses

#	Criterion	Discussion	References
1	Community Alignment	It is generally good practice (especially in the case of software) to align with the licensing tradition of the community.	[13]
2	Simplicity	The simplest license that covers all requirements is more than often the best one.	[13]
3	Sharing	Sharing improvements is important - this is the basis of the virtuous circle of community development and free culture. There may be cases where this is not desirable, but should be the default.	[13], [14]
4	Modification		[13].[14]
5	Irrevocability	Generally, most licenses cannot be revoked.	[14], [15]
6	Applicability	Not all licenses apply to all research outputs types	[14], [15]
7	License Scope	Important in cases where a work has multiple components	[15]

8	Adequacy	Licenses apply in generally two situations - public domain, which cannot be copyrighted, and copyrighted material or equivalent (*sui generis* works, for example).	[15]
9			

B.4 License Elements

Based on [13], [14], [15]

Licenses cover a range of options that are bound by two extremes:

1. 'All rights reserved' - which generally prohibits use by anyone that is not the owner of the work [15], and
2. 'Public Domain' - where all of the rights have been ceded.

The table lists these rights (elements) - which can generally be seen as either **permissions** (rights that are waived or not exercised), **conditions** (rights that are not waived or are implied), and **limitations** (rights that cannot be claimed by the user).

Aspect	Element	Description	Applicability
Permissions	Commercial use	Software/ work may be used in commercial applications	All Works
	Distribution	Software/ work may be distributed onwards	All Works
	Modification	Software/ work may be modified	All Works
	Patent use	This license provides an express grant of patent rights from contributors.	Software
	Private use	This software/ work may be used and modified in private.	All Works
Conditions	Disclose source	Source code must be made available when the software is distributed.	Software
	Attribution	Acknowledge the creators of the work and cite properly whenever used.You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that	All Non-Software Works

		suggests the licensor endorses you or your use.	
	Non-Commercial	NonCommercial — You may not use the material for commercial purposes.	All Non-Software Works
	License and copyright notice	A copy of the license and copyright notice must be included with the software.	All Works
	Network use is distribution	Users who interact with the software via network are given the right to receive a copy of the source code.	Software
	Same license/ ShareAlike	Modifications must be released under the same license when distributing the software or work. In some cases a similar or related license may be used. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.	All Works
	State changes	Changes made to the code must be documented.	Software
Limitations	Liability	This license includes a limitation of liability.	Software
	Patent use	This license explicitly states that it does NOT grant any rights in the patents of contributors.	Software, Data
	Trademark Use	This license explicitly states that it does NOT grant trademark rights, even though licenses without such a statement probably do not grant any implicit trademark rights.	Software
	Warranty	The license explicitly states that it does NOT provide any warranty.	Software

B.5 Community Preferences

#	Community	License Preference/ **Requirement**	References
1	Apache Projects	Must use the Apache License	[13]
2	WordPress Plug-Ins	Use GNU GPLv2	[13]

Annexure C: Scope of Publicly Funded Outputs

C.1 Publicly Funded Outputs

Generally speaking, a narrow definition of data policy extends to data produced via research grants, and required to prove the results obtained by others.

There are, however, many other publicly funded outputs in the broad realm of research and development that can be considered: government gathers data using public funding, and in many countries the Open Government movement [27] is gathering pace. In addition, global institutions (UN, World Bank, etc.) is also largely funded by public money and a case can be made that their outputs should be publicly available (preferably in the public domain).

Research increasingly produces much more than publications and data - and hence the scope of policy could extend to these other increasingly important research outputs.

Table C.1.1: Scope of Policy and Benefit Achieved

#	Scope	Benefit [9]		
		Reproducibility	Investment	Ownership
1	All tax-funded outputs, including research and government data	Yes	Yes	Yes
2	All grant-funded research outputs	Yes	Yes	Partly
3	Scholarly publications and data required for validation	Yes	Partly	Partly

C.2 Scope of Research Outputs

The table below provides a sense of what is regarded as research outputs by the community, how established the infrastructure for trusted sharing is, and how important the outputs themselves are thought to be for proper functioning of science[8] [26].

[8] Based on a survey of trusted repositories in the membership of the ICS World Data System, to which 33 repositories responded - approximately a third of the membership.

Table C.2.1 ICS World Data System Member Survey: Importance of Research Outputs

#	**Research Output**	**Current Implementation** [26]
1	Data and Data Services	Good
2	Scholarly Publications	Good
3	Code and web-based processing services	Emerging
4	Algorithms (pseudocode, logic)	Poor
5	Protocols and Methodologies	Poor
6	Registries of Research Entities and Outputs	Emerging
7	Vocabularies and Name Services	Emerging
8	Experiment Design	Poor
9	Brokering and Mediation	Poor
9	Workflows	Poor
10	Specimens and Samples	Poor
11	Assay/ Analysis Standardisation	Poor
12	Notebooks and Field Records	Not Evaluated