



Australian Government  
Bureau of Meteorology

# Bureau of Meteorology

---

Submission to the Productivity  
Commission Inquiry into Data  
Availability and Use

August 2016

For **further information** on this submission, please contact:

Dr. Lesley Seebeck  
Division Head  
Information Systems and Services

Ms. Vicki Middleton  
Acting Director of Meteorology and CEO

Version No: 1.1

Date of Issue: 22 August 2016

### **Confidentiality**

This is a public submission; it does NOT contain 'in confidence' material and can be placed on the Commission's website.

© Commonwealth of Australia 2016

This work is copyright. Apart from any use as permitted under the Copyright Act 1968, no part may be reproduced without prior written permission from the Bureau of Meteorology. Requests and inquiries concerning reproduction and rights should be addressed to the Production Manager, Communication Section, Bureau of Meteorology, GPO Box 1289, Melbourne 3001. Information regarding requests for reproduction of material from the Bureau website can be found at

<http://www.bom.gov.au/other/copyright.shtml>.

## Table of Contents

<b>Executive summary</b> .....	1
<b>Background</b> .....	2
Our Service.....	2
International alliances and obligations .....	5
Economic benefits .....	6
Data in the Bureau .....	7
Data Context.....	7
Data Sharing.....	8
Data Volume.....	8
Data Complexity .....	9
Data Quality, Security and Integrity.....	9
User Expectations.....	10
Regulation .....	10
Costs and Funding.....	11
Conclusion .....	12
<b>TERMS OF REFERENCE RESPONSES</b> .....	13
TOR 1: Public Sector Data Availability – Benefits and Costs .....	14
Improving availability of data through cost recovery.....	14
Processed data as open data .....	14
Attributing value and price to data.....	15
Licensing and charging for data .....	16
Open access licensing .....	17
Situation in similar jurisdictions .....	17
Data as an asset.....	19
Restrictions of using data provided under international treaties .....	19
Limited bandwidth infrastructure .....	20
Licence restrictions .....	20
Restricted data sharing formats .....	20
Further considerations .....	21
TOR 2: Private Sector Data Availability - Benefits and Costs .....	22
TOR 3: Improved Access to Stored Data .....	24
TOR 4: Benefits and Costs of Standardisation.....	25
TOR 5: Individual and Business Trust.....	26

## Executive summary

The Bureau welcomes this opportunity to provide a submission to the Productivity Commission's inquiry into Data Availability and Use. The Bureau creates, and receives from third parties, huge quantities of data to support its operations. The use of this data is critical to providing the meteorological services on which much of the Australian population, as well as critical sectors of the Australian Government and industry (such as aviation, shipping, primary industries and Defence) rely.

As weather affects all social and economic activity, the Bureau is committed to ensuring that its provision of data benefits the Australian economy and community, including through making data openly available wherever possible. This inquiry into increasing the availability and use of data to boost innovation and competition in Australia is therefore of considerable interest to the Bureau and is a matter in which the Bureau has much practical experience.

Both the *Meteorology Act 1955* and the *Water Act 2007* emphasise the role of the Bureau in collecting data from the natural environment and its provision of information in the form of value-added data, forecasts and warnings to the public. Through membership of the World Meteorological Organization (WMO), the Bureau provides its observations and model data, as well as scientific data to member countries, and in return receives global data from its peer agencies.

The Bureau is not merely a collector and conduit for data; it also provides assurance, storage and management. It generates data in the form of model outputs, as well as information products, knowledge, and advice based on the value-add and insight provided by the systems and expertise within the Bureau. The Bureau deals in huge amounts of data; we currently hold around 23 petabytes of data, forecast to increase to around 250 petabytes in 2021.

Ever changing technology and increasing demands from users mean that our current methods for managing data are becoming unresponsive and irrelevant. The Bureau is looking carefully at models for explicitly managing the costs of maintaining and sharing data, in addition to the traditional funding models for maintaining ICT systems.

In light of the discussion presented in this submission, the Bureau makes the following key points:

- Australia should develop a world-leading approach to public data provision that should be reflected in clear and detailed policy advice to agencies. This requires consideration of the position and experience in similar countries (including the United Kingdom, Canada and New Zealand) as well as taking a nuanced and practical approach to data provision from a licensing and cost perspective while also avoiding one-size-fits-all solutions;
- Where barriers unnecessarily obstruct the provision of data on an open basis, the Bureau supports resolving such situations. It is also important that public data policy recognise the full range of legitimate situations in which data need not be made available on an open access basis;
- Recognition should be accorded to the real costs for public sector agencies in collecting, storing and making available data sets through-life. The exponential growth in storage required, as well as the need to obtain the right skillsets and continually procure the necessary technology is becoming increasingly difficult in a tight fiscal environment. These

true data costs should be expressly recognised when considering how best to make public sector data available;

- Where there is a robust commercial market for public sector data (such as some weather and environment data in the Bureau's case), the sale of that data by agencies helps fund future high quality public sector data collection and development by those agencies and reduces the burden on the taxpayer;
- A flexible approach to licensing public sector data is critical to ensure that public sector data most effectively boosts innovation. Such an approach acknowledges the value of open access licensing systems (such as Creative Commons and the Public Access Licence) and allows public sector agencies the flexibility to apply the most appropriate licence. This ensures that each dataset is treated in a way that maximises its value to the public, whether through:
  - making it freely available under a range of open access licensing frameworks to maximise its appeal to users; or
  - through enabling agencies to invest in developing high quality data which can then be sold to a robust private market in order to offset the cost of investment (including collecting, developing, storing and providing access to that data).

It should be noted here that the use of a one-size-fits-all approach to open access licensing (such as sole reliance on Creative Commons to the exclusion of other open access licensing frameworks) can actually reduce the likelihood that some users will want to use that data;

- Thought should be given to the formal adoption of a government-developed open access licence (such as the Public Access Licence) as has occurred in the United Kingdom and Canada. This could be made available for use by Australian regional governments as well (as occurs in the United Kingdom and Canada) to provide a consistent approach and experience for the Australian public; and
- A single format to share data should not be mandated—a one-size-fits-all approach would be counter-productive. Instead, a suite of accepted data formats should be promoted.

## Background

### Our Service

The Bureau of Meteorology is Australia's national provider of weather, climate, water, ocean and space weather information. This provision of information is one of the most fundamental and widely used services of government. The Bureau is and has been the definitive source of environmental intelligence for over a century. Our services assist Australians in dealing with the harsh realities of their natural environment, including drought, floods, fires, storms, tsunami and tropical cyclones. The Bureau is the largest provider of meteorological information in the Southern Hemisphere, and plays a crucial role in the sharing of meteorological and related data within the global community of national meteorological agencies. Australia's meteorological responsibilities cover continental Australia, all

offshore territories, marine environments and the Australian Antarctic Territory, approximately 26 million square kilometres.

The Bureau’s functions include the supply, publication and promotion of meteorological and water data, as well as the provision of extreme weather warnings under the *Meteorology Act 1955* and *Water Act 2007*. The Bureau performs its functions in the public interest and for the purposes of special interests, including Defence, shipping, civil aviation and primary industries.

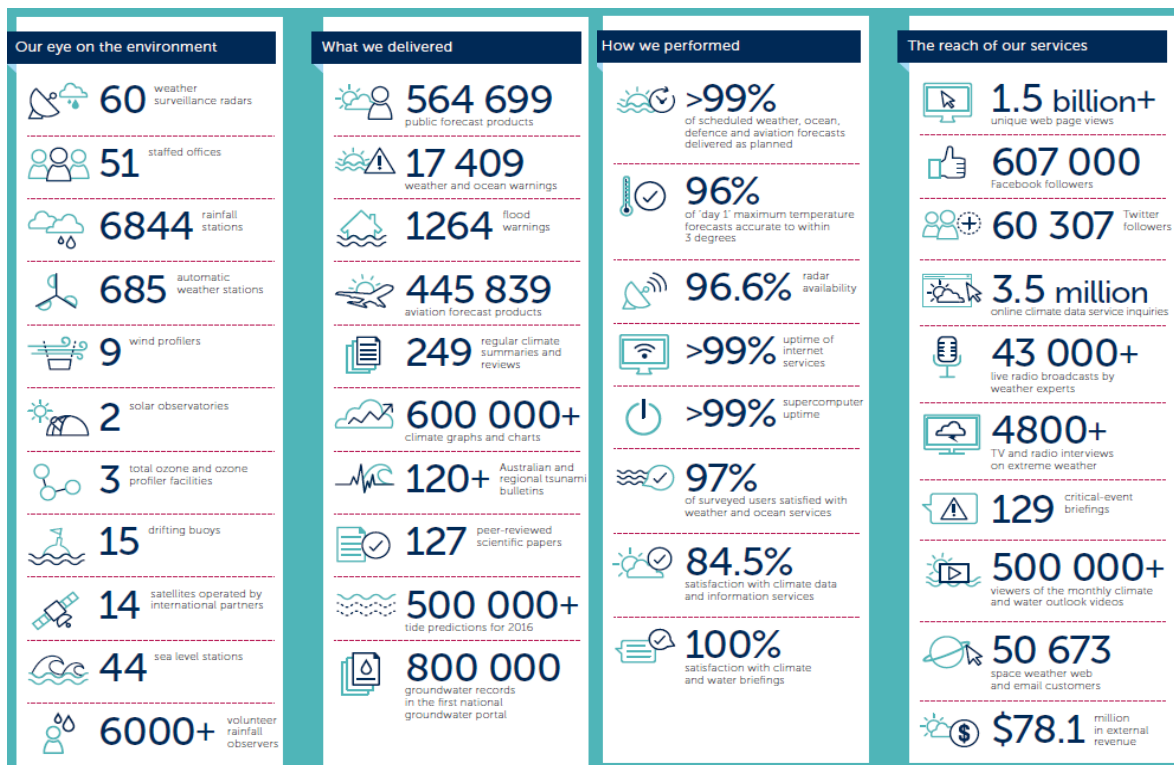


Figure 1: Bureau Data Fast Facts

Data is central to the Bureau’s role and function, and focuses on all aspects of the environment (see Figure 2). Observations are gathered from the environment in real-time or near real-time from a variety of instruments, and drawn from and provided to our international partners through the WMO framework. Some products are delivered in as close to real-time as possible (tsunami warnings) while others are generated through regular (and increasingly frequent) numerical modelling on the Bureau’s supercomputer. Some products require additional correction, verification, and value-adding (particularly longer-range assessments, seasonal outlooks, and scientific understanding) or rely on the integration of data from partners such as Geoscience Australia.

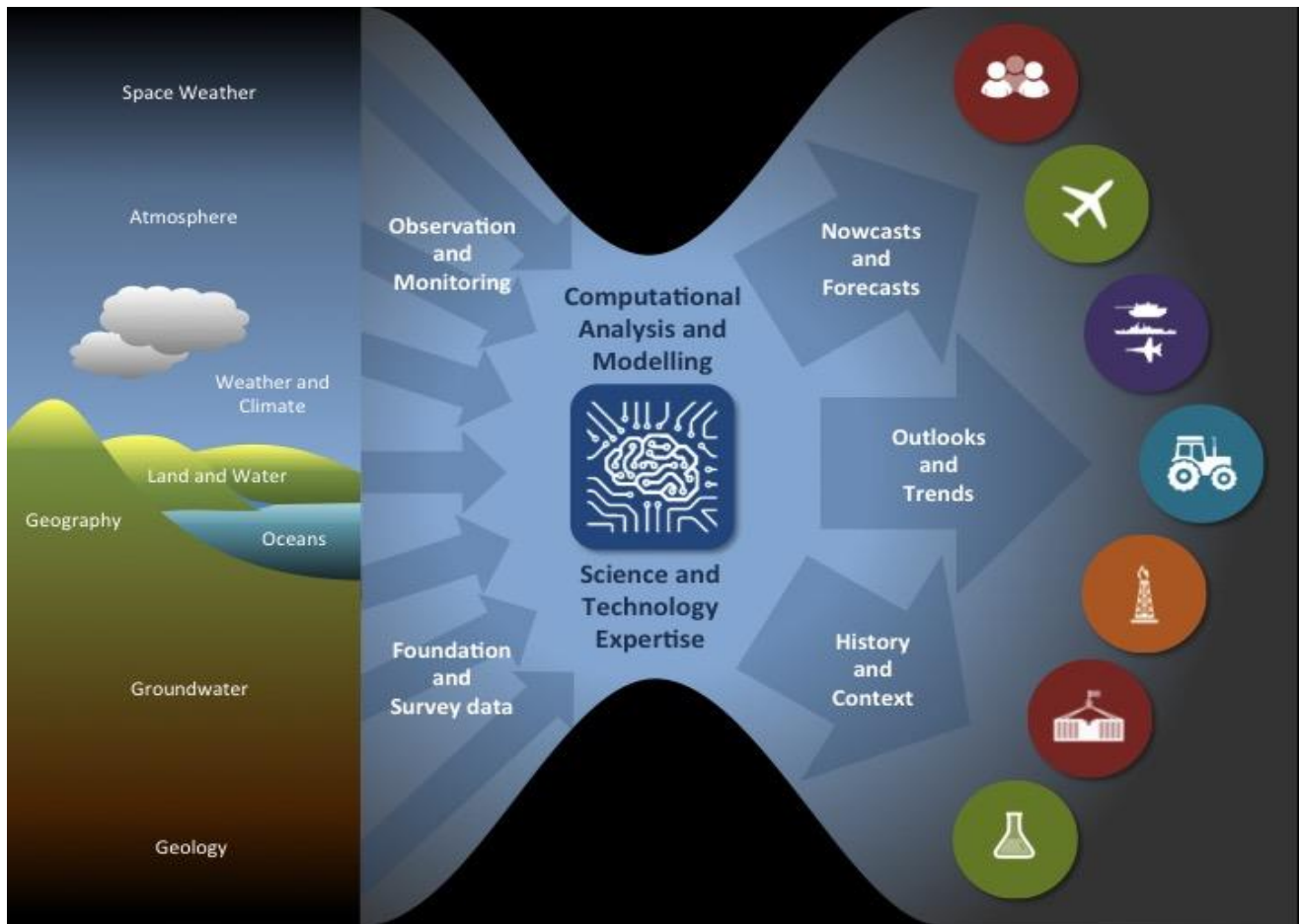


Figure 2: High-level representation of Bureau functions

Bureau data and expertise is used by the Bureau for:

- timely information to support public safety (for example, from heat wave notification to forecasts of the paths of strengths of cyclones to wind shifts during fires);
- improving knowledge of environmental information to support national security and border protection (for example, ocean modelling, support to air capability, advice on sea state, direct support to operations, training);
- supporting safe and efficient transport (for example, through its aviation, flood and storm forecasting services);
- timely and accurate information to support critical industries (for example, lightning warnings, wave modelling and improved information on cyclone behaviour affecting supply chains and operations, space weather alerts);
- improving investment in food security (through, for example, seasonal forecast, soil moisture, and streamflow, and long-term climate predictions);
- monitoring and protection of natural resources (for example, through monitoring ground water and water quality and turbulence around the Great Barrier Reef);

- meeting Australia's international obligations and interests (for example, through its contribution to Australia's support of the international rules-based order, and assisting regional developing countries);
- meeting the government's digital, innovation and data agenda (the Bureau operates the most popular website in the Australian government, and provides over 80 per cent of all its products and services online, and a number of companies now draw directly on Bureau data to support their own digital applications (for example, Weatherzone, Pocket Weather)); and
- continuing to underpin to improved productivity (for example, through improving accuracy and responsiveness of environmental conditions affecting mining supply chains and deployment of resources in primary industries).

The Bureau contributes to the building of national resilience through the analysis of past climate variability and change, and by providing predictions of future changes. It provides an understanding of the risks of natural hazards and facilitates the design and implementation of mitigation measures.

### **International alliances and obligations**

As part of its responsibilities under its Acts, the Bureau fulfils Australia's obligations under the Convention of the World Meteorological Organisation (WMO) and related international treaties and agreements. Many of those obligations involve the collection, management (standardisation and curation) and sharing of data, especially observations data. The Bureau is also responsible for the coordination of observations for WMO Region 5 (shown in Figure 3) and is the only major meteorological centre in the southern hemisphere. Data on global environmental conditions are critical to the accuracy and reliability of the numerical weather modelling that is core to the forecasting and analysis undertaken by all meteorological agencies.



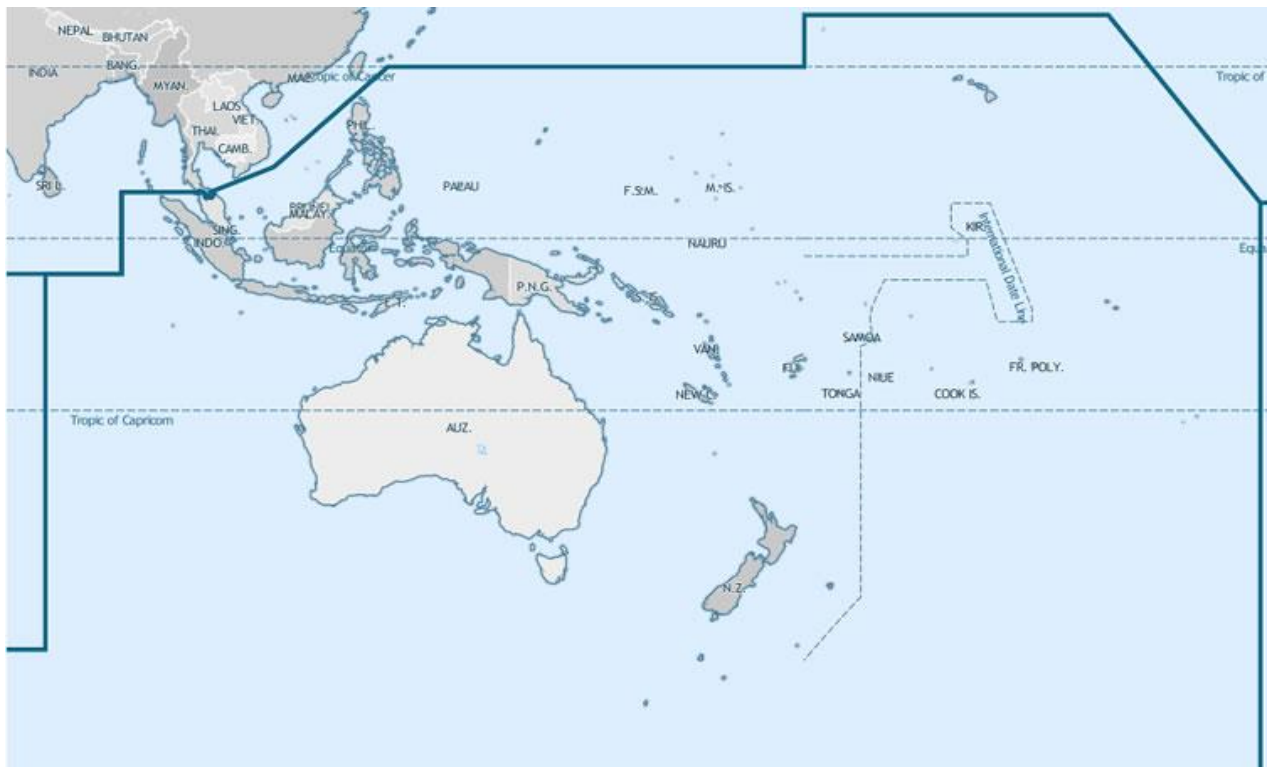


Figure 3: Bureau area of responsibility

The Bureau sends its WMO partners around 700,000 messages each day (approximately one terabyte), for use in their models, and receives observations in return. The Bureau also shares its model data—and receives model data from key partners with modelling capability—for comparison and use in internal analysis and forecasting processes. Overall, Australia is both a major beneficiary of and contributor to the WMO arrangements, while also supporting bilateral cooperation under the WMO aegis. The estimated cost of collection of the observational data obtained from international partners through WMO and bilateral arrangements is several billion dollars annually.

### Economic benefits

The Bureau contributes to increasing national economic growth capability and resilience in meeting all the needs of end-users and partner agencies, and providing an essential, often critical element to the decision-making of individuals, companies, and governments. Our products and services address a range of time-scales from the immediate to distant event horizons. Scenarios, real-time events, and mitigation measures can be tested using Bureau data, and against Bureau forecasts and insights.

Many sectors of the Australian economy and society rely on Bureau data, information, and advice. Estimations of economic value to national economies from peer agencies internationally range from the UK Met Office's conservative £29.5 billion to the US National Oceanic and Atmospheric Administration's USD\$485 billion. A recent study undertaken on behalf of the UK Met Office (UKMO), found a return of around 14 to one; for every pound invested in the UKMO, the return to the UK

economy was fourteen fold<sup>1</sup>. While the UKMO has slightly different functions to the Bureau, there are similarities to the Bureau's contributions.

## Data in the Bureau

The Bureau occupies a unique position in the Australian Government due to the volume, veracity, velocity and variety of data it has to manage, process and disseminate. Data content complexity occurs at both ingest and dissemination points due to changing or maturing scientific content standards, content encoding requirements and increasingly unstructured data.

The Bureau's mission poses considerable and growing challenges. Those challenges are categorised and discussed below and cover data sharing, volumes, complexity, content and quality (including security and integrity), user expectations, technology, regulation, licensing and cost.

## Data Context

Being a public science organisation, the Bureau values the free flow of data. As an agency that is operational 24 hours a day, 365 days a year—the Bureau has specific needs and constraints regarding the data it ingests, generates and disseminates.

Constraints such as license restrictions and intellectual property rights of third parties restrict the provision of some data. To meet the Bureau's mandate, and ensure the trust accorded its forecasts, warnings and advice, the provenance of data and the context of its use are critically important. Provision of certain data may be restricted by context, particularly in cases where specialist tools are required by end users to make sense of the content.

Value-added and assured products and services are of greatest use to end-users. Some of the processes for value-adding to data include consistency adjustments—for example, to account and adjust for the effect of the shade of a tree on an automated weather station. Understanding this context is crucial to making decisions when relying on the observational data.

Security, as a means of ensuring the integrity of data, is increasingly important and costly as it must be implemented from collection right through to delivery. Experience is showing that the management and provision of data in a digital world can be expensive in a way not accounted for in traditional architectures and budget models. For example, the shift to cloud and as-a-service provision of services impinges operating expenditure, not capital funds, which have been the traditional, planned for source of funding for ICT capability.

The Bureau provides selected products and services free of charge to meet its basic service responsibilities. But it also provides some data on cost recovery terms and in certain cases, full cost recovery rates for tailored services and products specifically requested by users. The ability to have flexibility in charging models enables the Bureau to invest in the quality and value of data provided,

---

<sup>1</sup> London Economics, *Met Office—General Review*, April 2015, available at [http://www.metoffice.gov.uk/media/pdf/1/n/Met\\_Office\\_General\\_Review\\_Economic\\_Summary.pdf](http://www.metoffice.gov.uk/media/pdf/1/n/Met_Office_General_Review_Economic_Summary.pdf).

maintain capability critical to support data provision, and in the development of new products and capabilities that benefit the economy and public interest in the long-term.

### Data Sharing

The Bureau participates in and directly benefits from regional, national and international data sharing initiatives and obligations. We would not be able to perform many of our functions without the data-sharing arrangements of the members of the WMO, which include private, public and research sectors.

Some shared data is only available under certain restrictions from the owner or provider, which may include a limited ability to use that data in situations other than those prescribed, or to charge for products derived from that data.

### Data Volume

Whole-of-government 2014-15 data showed that the Bureau managed 8.6 per cent of total government data holdings but was responsible for only three per cent of the total government expenditure attributed to storage<sup>2</sup>. The Bureau currently holds around 23 petabytes (PB) of data. That is expected to increase ten-fold over the next five years, and we expect that the current rate of increase will itself accelerate.

We hold the information on the initial conditions and the models applied, which allows us to replicate the information if required. At the moment the Bureau keeps this and other high volume data indefinitely and on high cost media. We are exploring policies that will pragmatically and responsibly limit the time that certain data is held, thus limiting the rate of growth and ensuring best value is achieved.

The main drivers of data growth are:

- the ingest of satellite image data (Himawari-8 generates around 330 terabytes (TB) per year), which can increase by an order of magnitude with each new generation of satellite. Korea and China will each launch satellites with similar capabilities and data volumes within the next two years, which will also be available to the Bureau;
- each generation of numerical weather prediction (NWP) models, requiring support and storage for both ingest and model output (our new supercomputer, Australis, requires six petabytes per year in storage for the new generation models)—data support for NWP is anticipated to increase 12 times over five years;
- climate, seasonal and earth system models, though run much less frequently than NWP models, are some of the most computationally and data intensive models around (research,

---

<sup>2</sup> Data provided by the Department of Finance, drawing on material supporting its *Australian Government ICT Trends Report 2014-15*.

climate and earth system models are run at the National Computational Infrastructure, using Bureau data); and,

- continued growth in the observation record (both over time and through increased instrumentation of the environment), product generation and interaction with customers, translating into storage growth of around 10 per cent per year.

Growing data volumes affect storage requirements, transmission capabilities, accessibility and research.

The main drivers of costs associated with data are:

- initial outlay for additions to existing systems, or development of new systems to cater for increased customer demand;
- ongoing costs of compliance with national/international content and technical delivery standards;
- initial and ongoing costs of implementing appropriate security and access protocols, particularly to legacy systems; and
- escalating costs associated with software licensing and supporting increasing delivery infrastructure.

### **Data Complexity**

Aside from the huge volumes of data that the Bureau handles, there is an associated high level of complexity of that data, which also contributes to the cost of managing that data. Some of the parameters of the complexity include:

- increasing variety in the sources and types of data that we do not control and that provide disparate, unstructured data on a scale we have not dealt with before;
- the various time scales and events that the data represents. Data must be available from minimal processing through to high value-add in near 'real-time' to seasonal, decadal and centennial periods in both an operational and support capacity;
- complexity and richness of parameters in available data represented by the increased level of detail that is required, in multiple contexts;
- the variety in user needs and delivery channels; and
- probabilistic modelling in addition to the existing deterministic modelling approaches.

### **Data Quality, Security and Integrity**

Data for the climate record needs to be accurate, precise to several decimal places and needs to be historically consistent. Scientific content exchange standards apply. Maintaining this can be a challenge when that data is open and shared and subject to security threats.

In 2012, the Australian Signals Directorate assessed the Bureau as being in the top five government organisations essential for national security due to the critical importance of its services for key sectors such as Defence and aviation. To ensure continuity of service provision and data integrity, sufficient redundancy must be built into the system to insure against failure of infrastructure and

provide the ability to recover rapidly when required. Again, this adds to the overall cost of dealing with and providing data.

## User Expectations

User expectations have increased dramatically with the use of smart devices (phones, tablets, wearables) that require user responsive outputs. There is an increased expectation that information will be available on demand and in user-centric formats.

Users expect hyper-local forecasts and information on the potential implications of localised events, with a focus on short and medium-term future events. This sort of context would require links to data and sources that are not directly controlled by the Bureau. It requires a translation from the high accuracy and precision that the Bureau uses to generate its products to a user-centred understanding. Further, the focus is on shorter cycles of updating information, in user responsive formats linking with contextual information from other sources. The information architecture associated with such responsive systems is vastly different from what the Bureau currently has in place.

Such demands extend across the full range of users. For example, the *Review of the Bureau of Meteorology's capacity to respond to future extreme weather and natural disaster events and to provide seasonal forecasting services* (The Munro Review) found that an increased frequency of natural disaster events led to a heightened level of risk awareness particularly by State and Local Government agencies. Consequently, the Bureau is under ongoing pressure to increase its own weather observation and monitoring networks. The Bureau has also found that commercial customers preference a tailored 'look and feel' of information over accuracy alone, and is adjusting its approach accordingly.

The Bureau anticipates the demand for its services in disaster planning and mitigation will continue to grow in the future. Our challenge lies in ensuring that we are able to deliver both on the needs of the immediate—minute-by-minute, hourly, daily updates—in a user centric format, as well as maintaining the capability and capacity to continue to deliver information that will inform long-term strategic decision making. This requires continuous investment to meet the ever-growing demand of users.

There is currently no recognition for the funding needed to implement offsite Infrastructure, Platform or Software-as-a-Service capability that can meet user's expectations. In reality, the internal systems developed for point-in-time solutions have limited flexibility to adapt to changing user expectations and are often required to be retrofitted with additional features that include data capture, management and delivery capability. As these features were not planned, funding bids do not include their development or maintenance costs. This results in a spiralling technical debt and introduces system fragility.

## Regulation

The list of legislation, international treaties and agreements, national strategies and policy statements that applies to the Bureau's activities is extensive. Policy and regulation development is struggling to keep pace with government administrative changes, let alone the consequences of the development of digital technologies, increased user expectations, and increasing volumes of data. Navigating between differing interpretations and applications can lead to suboptimal outcomes and impose

considerable overhead and often inertia. Moreover, many are imposed without consideration of the resourcing needed for compliance. As the Bureau has high-levels of scientific, technological and highly specific requirements, yet must deliver continuously and reliably in a data-intensive, high-response, operational environment, it often finds itself as an outlier on assumptions made and policies devised for the 'average' agency. The Bureau considers that its users would benefit from a lighter touch and more consistency across applicable policies and regulation relevant to data and to ICT as the medium of data generation, transmission, storage, analysis and provision.

## Costs and Funding

The Bureau receives a combination of government funding and external revenue from Bureau cost recovery activities. Operations are funded primarily from government budget appropriations. Some revenue is derived from the charge for services to the aviation industry, the private sector, and other government agencies, including Defence. In 2015-16 the Bureau's operating revenue was \$288.6 million. Of this, the government contributed \$213.9 million. The remaining \$78.4 million was derived from 'own-source income'.

There is a tension between the potential to recover costs for certain activities and the government's open data agenda. As an example of this, the National Commission of Audit recommended that agencies consider assessing the value of their data and making better use of the data they currently manage<sup>3</sup>. Clear advice on the linkages between these imperatives would be helpful for agencies. Also, it is important that policies recognise the diversity of different agencies. While one type of agency that does not deal with much data may want to adopt an approach of making all of its publications available on an open access basis, other agencies that deal with large amounts of data may not be able to do this as easily without substantially reducing the services provided to the community.

The Bureau currently provides data, information and services across the four levels of charging mechanisms available to it:

- free, or public good products, delivery via web and mass media;
- free public good products, with a cost to access for some services;
- incremental cost recovery where the Bureau has added value through analysis and processing, to provide tailored information for specific industries; and
- commercial products, which are bespoke, where we provide considerable value-add, for which the Bureau charges.

---

<sup>3</sup> National Commission of Audit, *Towards Responsible Government: The Report of the National Commission of Audit – Phase One*, February 2014, available at [http://www.ncoa.gov.au/report/docs/phase\\_one\\_report.pdf](http://www.ncoa.gov.au/report/docs/phase_one_report.pdf). See Part B -10.5, page 223.

## Conclusion

The Bureau maintains a vast store of data, and provides both data and information to users in multiple forms across various timeframes and levels of urgency. The demands on us to provide more user-centric, high-response services are increasing. The Bureau's ability to meet the challenge—and ensure the economic, security and community benefits—associated with a data-centric and data-driven world is dependent on a realisation of the costs associated with data, and ability to meet user-needs agilely.

Open data is a challenge that we are working through. It is an area that requires more definition to facilitate consistent treatment across government and users of data. It hasn't been explored adequately with regards to funding and this is fast becoming a growing concern.

## **TERMS OF REFERENCE RESPONSES**



## TOR 1: Public Sector Data Availability - Benefits and Costs

*Examine the benefits and costs of options for increasing availability of public sector data to other public sector agencies (including between the different levels of government), the private sector, research sector, academics and the community. Where there are clear benefits, recommend ways to increase and improve data linking and availability.*

### Improving availability of data through cost recovery

As discussed in this submission, the Bureau's mission is to collect, analyse and process data and distribute information to entities across the public and private sector. Data availability can be improved through having the right mix of cost recovery, to ensure that new forms of data are developed and provided in a way that users find useful.

**The Bureau recommends that consideration be given to understanding the circumstances in which improved data availability can be enhanced through cost recovery models for some data. The costs of developing, managing and storing data through its lifecycle should be recognised.**

### Processed data as open data

According to the Productivity Commission's Issues Paper (2016) data is the raw content that is devoid of context. Almost all of the data that the Bureau releases has been processed (including quality assured) to some extent. As is noted elsewhere, observational data—often considered the most 'raw' of data in meteorology—typically undergoes some processing as part of ingest. Observational data is spatially indexed to enable topological calculations for use in further processing, numerical model input and to ensure consistency in the historical record. In this sense, 'raw' observational data may be misleading or of considerably lesser value than that processed and spatially indexed.

The Bureau would like to draw attention to the regime adopted by the UK Met Office, given the similarity in complexity, scale and responsibilities associated with data and its management<sup>4</sup>. The advantage of the Met Office policy is that not only does it reflect the broader government impetus towards open data, it recognises issues associated with costs, sustainability, context and third party data. It also explicitly differentiates between the different categories of data—open, managed and internal—and the fact that the Met Office, like the Bureau, does not simply collect but generates substantial amounts of data.

Last, the Met Office relies on funding by external parties, including many government agencies. It manages achieve a balance between cost, charging and open data policies. As noted above, the Bureau uses opportunities to charge for products and services, even on a cost recovery basis, to

---

<sup>4</sup> UK Met Office, *Open Data Policy*, 25 November 2014, available at [http://www.metoffice.gov.uk/media/pdf/3/5/OpenDataPolicy\\_MetOffice\\_v1.0.pdf](http://www.metoffice.gov.uk/media/pdf/3/5/OpenDataPolicy_MetOffice_v1.0.pdf). See Table 1, pages 5-6.

fund new and additional services that it would not be able to provide within its appropriation envelope.

**The Bureau recommends clarification of the concepts of open data, cost of access, and value-adding to data in order to improve the ability to explore revenue opportunities for government.**

*a) identify the characteristics and provide examples of public sector datasets that would provide high-value to the public sector, research sector, academics and the community to assist public sector agencies to identify their most valuable data*

### **Attributing value and price to data**

The value of data changes across users and across time. Its value is associated with the context of the user and of the data itself.

The context includes temporal attributes, scale, resolution, quality and availability to form a complex pattern to which an end user will assign 'value.' No user has a need for data or information in the absence of context.

For some use cases, in academia and scientific realms for example, the complete set of data, including errors, is considered the quality data set for algorithm improvement or machine learning training. In contrast, an individual member of the public may prefer to know today's weather forecast temperature (in whole numbers) and what the day 'feels like'.

The data that the Bureau considers of the highest value is raw data because this is what is used to produce other products that consumers value. But this is generally not what we release externally. The majority of data and information the Bureau releases externally is high-value to one or more of our clients and has undergone some level of processing. We provide data that is context specific, and can be consistently reproduced.

At the moment the only way the Bureau can determine what the appropriate quality and timeliness that is required by consumers is by analysing lag indicators such as hit rates, number of downloads and direct client enquiries. These provide very limited information on what users might want. To anticipate and build services and products that provide further value and be user responsive we would need to have a much higher understanding of users' contexts.

Active monitoring and needs analysis is currently an unfunded activity—but such information could go towards being able to deliver in ways that could generate revenue for government.

We are also conscious that provision of data bears its own set of responsibilities, inasmuch as

- it may be time-limited, and so we would want to avoid confusion with data that has expired; and
- that once a data set has been provided, there is generally an inherent commitment to the ongoing provision of that data and its upkeep.

Consequently, we would need to ensure we are able to support the sustainable provision of data—and that, given fiscal constraints and current system capabilities, is challenging.

**The Bureau suggests that conventions on how data can be valued be developed. This could then inform the development of funding models for data management.**

### Licensing and charging for data

The Bureau licences data in a number of ways:

- On an open access basis to enable full and free re-use. This occurs under an open access licence (either Creative Commons or the Public Access Licence);
- On a standard agreement licensing basis, either free or following payment of a fee. This occurs via the Bureau's two short-form standard licence agreements. The four-page Access Agreement is used for ongoing paid data subscriptions which is principally used where substantial data feeds are provided over a period of time to organisations, and the two page Short Licensing Agreement is used for one-off provisions of data whether for a fee or otherwise. These agreements are used where it is not suitable to make data available on an open access basis, either because of the licence restrictions involved or because of the cost of provision. The cost involved in providing and maintaining datasets cannot be fully funded by government alone, particularly in an environment of pressure on public sector spending. As a result, it is necessary for some public sector data to be made available on an incremental cost recovery basis;
- On a bespoke basis. Where a request for data cannot be provided on an open access basis or via the Bureau's standard agreements, then bespoke terms are created to enable access to the data that are appropriate to the circumstances. In these cases, the Bureau is creating or obtaining data in response to specific requests from users, and in order to be able to undertake such activities it is necessary to charge the recipients for the services provided; and
- On a default basis: providing limited rights where no licensing approach has been applied.

The Bureau takes a sophisticated approach to determining how data should be made available. First, the Bureau looks at the practical considerations including the nature of the data, the public interest in the data and the level of value-adding and quality assurance that is required by the Bureau. Second, the Bureau applies a due diligence licensing tool developed by the Bureau. That licensing tool requires consideration of:

- the extent to which the data to be provided contains material owned by third parties which are subject to licensing restrictions or copyright protection;
- whether there are any legislative restrictions to making the data available;
- whether there are any contractual restrictions to making the data available;
- whether there are any confidentiality or security restrictions on making the data available;
- whether provision of the data would undermine the commercial interests of the Bureau;
- how the data might be used by the public;
- whether there are quality issues with the data that might affect its usefulness or need to be rescinded;
- whether the Bureau wishes to warrant data against risk of copyright infringement to actively encourage private sector use (to provide confidence to users that they can safely invest in

incorporating that data in their products without risk of intellectual property claims against them);

- whether the data should not be allowed to be changed by users from a public interest or public safety point of view (for example, extreme weather warnings); and
- whether any special conditions need to be attached to the data.

The Bureau notes that this list of due diligence activities is quite similar to the guidance given by the New Zealand government to its agencies in considering making data openly available.

### **Open access licensing**

The Bureau uses two open access licence suites for making data freely available on open access terms. The first is Creative Commons and the second is the Public Access Licence (created by the Australian Government Solicitor in conjunction with the Commonwealth Attorney-General's Department). The reason that the Bureau uses two sets of licences is that the features of either set of licences are not always appropriate in every situation where the Bureau wishes to make data available to the public on open access terms.

For example, some private sector organisations will not use material that is licensed under Creative Commons because it does not suit the way in which they licence their products to end users. Similarly, where licence restrictions are imposed by third parties on material that an agency wishes to make available on an open access basis, and those licence restrictions do not fit with a Creative Commons licence, Creative Commons cannot be used. However, in both cases, because the Public Access Licence is more flexible, it can be used instead to ensure that the material can still be provided on an open access basis. These and other issues with Creative Commons licences are one of the reasons that jurisdictions such as the United Kingdom and Canada use their own government open access licences.

In the Bureau's view, the more widespread adoption of the Public Access Licence by public sector agencies would improve the availability of public sector data on an open access basis. In some cases, data could be made available under both a Public Access Licence and a Creative Commons Licence to enable users to choose the licence which best suits their purposes. In other cases, either a Creative Commons licence or a Public Access Licence may be more beneficial, depending on the specific nature of the data being provided and the circumstances surrounding its provision to the public.

### **Situation in similar jurisdictions**

#### ***Licensing***

- The United Kingdom and Canada use their own government open access licence.
- The United Kingdom releases open access data through the Open Government Licence. Its other standard approaches to licencing are the Non-Commercial Government Licence and the Charged Licence.
- Canada uses its own Open Government Licence to make data openly available. This licence has been adopted intermittently at a provincial, territorial and municipal level.

- New Zealand uses Creative Commons 4.0 International licences as a default but not in every case. The circumstances where Creative Commons is not used<sup>5</sup> include where release of information would:
  - conflict with legislation, policy or an order by a court;
  - amount to a breach of contract, confidence, privacy or otherwise result in an actionable wrong;
  - result in disclosure of a trade secret;
  - conflict with the legitimate commercial interests or business model of an agency;
  - publish a patentable invention for which the agency may want to apply for the patent;
  - be against the public interest 'in having a single, authoritative and non-adapted version of a specific data source';
  - result in the publication of incomplete work or material which would be 'materially misleading' or which would 'cause or contribute to material error on the part of recipients or licensees' where disclaimers or other statements could not appropriately mitigate these risks; or
  - influence the control, integrity, economy or other potential of the traditional knowledge or culturally sensitive material of Indigenous persons.

### **Fees**

- The United Kingdom, Canada and New Zealand permit charging for government data although all acknowledge and prefer the approach of providing the data free where appropriate.
- The United Kingdom has developed a template Charged Licence which can be used by agencies to make data available for a fee.
- Canada has committed to reviewing and eliminating fees being charged for access to government information in accordance with the Action Plan on the G8 Open Data Charter.
- New Zealand discourages charging a fee for the use and re-use of data and only considers it appropriate to do so where it would not act as a barrier to such use or re-use.
- However, in the United Kingdom, Canada and New Zealand, considerable meteorological data is made available for a fee on subscription basis, see:
  - United Kingdom: <http://www.metoffice.gov.uk/services/wholesale>
  - Canada: <http://www.ec.gc.ca/meteo-weather/default.asp?lang=En&n=E6B571EB-1>
  - New Zealand: <http://about.metservice.com/weather-for-business/metconnect/>

---

• <sup>5</sup> See: <https://www.ict.govt.nz/guidance-and-resources/open-government/new-zealand-government-open-access-and-licensing-nzgoal-framework/nzgoal2/>

- The Bureau's peer agency in the United States, the National Oceanic and Atmospheric Administration (NOAA) has found that unanticipated demand, due to rapid uptake of apps, for example, can impose unexpected costs on the agency, and is exploring different models of data provision<sup>6</sup>.
- Experience in other countries have shown that successful models of open data provision require infrastructures with different regimes of legislation, cost recovery models and vastly superior network connectivity than are currently available in Australia. The French meteorological agency (Météo France) addressed this with a more constrained model and limited their costs by only providing 20 open public data sets through a redesigned portal.

The Bureau's view is that the availability of public sector data can be improved by providing public sector agencies with options to make data available for free or at a reasonable cost, as may be appropriate; and under a range of licensing terms that provide flexibility to account for the nature and restrictions on data and the way in which users wish to use that data.

**The Bureau recommends that the Productivity Commission consider the costs involved in data creation and provision, the limitations for public sector funding of such activities and the need for flexible approaches to licensing and charging in order to provide users with access to a range of data in sustainable ways that suit their demand for it.**

### Data as an asset

The Commission's Issues Paper correctly identified that data as an asset is a unique, accumulating resource that increases in value the more it is used without being consumed. Collections such as Australian Climate Observations Reference Network—Surface Air Temperature and the Climate Record accumulate value with the passage of time and ongoing updating. The resulting long-term and accumulative Record provides an anchor for observation validation, current forecast model tuning and future long-term policy setting.

*b) examine legislation or other impediments that may unnecessarily restrict the availability and linking of data, including where the costs are substantial, and consider options to reduce or remove those impediments*

### Restrictions of using data provided under international treaties

The Bureau operates within internationally agreed treaties under the WMO that restrict the release of data. The treaties include the *International Convention of Safety of Life at Sea* and *Convention of International Civil Aviation*, which operates a shared model for internal Bureau use only. The restrictions are in place to provide direct benefit to the Bureau for generating forecasts, contributing to international research and providing dedicated services to emergency services and aviation. There is no ability for the Bureau to resupply this data in a form other than the original or even re-purpose it

---

<sup>6</sup> See NOAA's Big Data Project: <https://data-alliance.noaa.gov>

for anything other than the agreed use. Such data and products derived from it also cannot be provided on a user pays basis.

This restriction may put the data covered by international agreements out of scope for this review.

### **Limited bandwidth infrastructure**

Data management within the Bureau has become an increasing cost due to its high volume. Considerable traffic (in 2014-15, 854.1 TB upload, 253.2 TB download) passes through the Bureau's internet gateway in support of operations, as well as servicing public and user demand, and it serves as the back-up to critical observational data links. In 2015-16, the annual internet gateway traffic cost was in the order of \$1 million. The demand on our data due to the variety, veracity and velocity of information required has placed considerable pressure on our internet gateway infrastructure, with demand necessitating an increase in bandwidth (from 300 Mb/s to 400 Mb/s), at additional cost. The pressure on the gateway is not decreasing, with saturation of the new bandwidth occurring regularly. That will require new management approaches and further cost.

### **Licence restrictions**

There are often restrictions placed on how the Bureau can licence data to the public. This will occur where third parties provide all or part of a data set which the Bureau then wishes to provide in some form to the public. Australian Government policy (as set out in the Australian Government Public Data Policy Statement released in December 2015) requires Australian Government entities to publish government data by default under a Creative Commons BY Attribution licence unless a clear case is made to the Department of the Prime Minister and Cabinet for another open licence. Creative Commons does not allow its licences to be branded as 'Creative Commons' licences where any of the terms of the licence are changed. In circumstances where part of a data set is subject to a restriction from a third party which is required to be passed through to users of the data set, the Bureau is unable to use a Creative Commons licence to make it available on an open access basis because the restriction would alter the terms of the Creative Commons licence. In such circumstances, unless an alternative open access licence is used, it is not possible for the Bureau to make the material available on an open access basis.

The only alternative open access licence the Bureau is aware of in the Australian context is the Public Access Licence which, allows more flexibility. It is, we understand, also a licence developed by the Australian Government specifically for Australian Government use. In this context, it would be helpful for the Australian Government Public Data Policy Statement to recognise and support the use of this licence. Doing so will assist agencies to increase the availability of public sector data.

### **Restricted data sharing formats**

In terms of availability models the format of data needs to be considered by the Commission with regard to ensuring that a one-size-fits-all proposal is not the outcome. The Bureau is acutely aware of the current format for ubiquitous data exchange being XML due to its interoperability and ability to be processed with no or little prior specific knowledge. The XML exchange format also provides the ability to link data sets easily. To achieve this interoperability the XML format stores data in a less efficient manner than other, more specialised formats. As XML data is stored less efficiently this increases the cost of managing the data through increasing the storage and transfer footprint of the

data. Of particular relevance to the Bureau, XML's storage problem increases dramatically with the addition of complex scientific content.

**The Bureau recommends that a single format to share data should not be mandated: a same size fits all approach would be counter-productive. Instead, a suite of accepted data formats should be espoused.**

A related issue is that complex and large amounts of data currently necessitates special analysis and querying tools which are not widely available or easy to use.

The scientific complexity of data held by the Bureau includes point-in-time observations that require additional context, such as location and statistical processing for the information to be useful outside the Bureau. A fundamental tenet of the UK Met Office Open Data Criteria is that it does not require specialist tools or knowledge to interpret the data in a meaningful way.

### **Further considerations**

An oft-quoted position, particularly by the open data access industry, is that as public sector data is funded by the taxpayer, it should all be made freely available to the taxpayer. This raises an important point, namely the consideration that should be given to who benefits from release of a particular data set. Is it the Australian taxpayer? Or is it overseas corporations? A more nuanced approach to free public data provision would be to consider what value is delivered to Australian taxpayers for not seeking a cost recovery return on the provision of data.



## TOR 2: Private Sector Data Availability - Benefits and Costs

*Examine the benefits and costs of options for increasing availability of private sector data for other private sector firms, the public sector, the research sector, academics and the community. Where there are clear benefits, consider ways to increase and improve availability. The Commission should:*

- a) identify the characteristics and provide examples of private sector datasets that would provide high value to the private sector, public sector, the research sector, academics and the community in developing or providing products and services and undertaking research and developing policy*

State government and private sector entities already supplement the Bureau's observations data in some locations. These are extremely valuable for providing more relevant forecasts, warnings, and seasonal predictions. We receive observation input from marine-based oil rigs and provide specialised forecast services. We also rely on private sector entities and the public for lightning strike data. There can be financial sensitivities associated with this data and we ensure that is protected while making the information available to emergency services during fire season activity.

The sensitivity in using such data lies in ensuring that we do not compromise any competitive advantage that the private sector entity may have related to that information.

We believe there is scope to collect more such data, applying concepts such as the Internet of Things. There would be substantial benefits in turning such data into public goods and services. We are currently shackled by our capacity, capability and systems that require modernisation and uplift to deal with the larger amounts of data this would entail.

- b) identify the concerns of private sector data owners and provide recommendations on principles or protocols to manage these concerns*

The Bureau engages with private providers as third parties in commercial agreements with exchanges of data subject to embargoes, or release restrictions for a defined period of time. This protects entities' commercial and competitive advantages but provides the Bureau and government with valuable information in hazards and warning forecasts without compromising the source of the information.

- c) examine legislation or other impediments that unnecessarily restrict the availability of data, including where the costs are substantial, and consider options to reduce or remove those impediments*

The greatest concern here, as discussed above, is the financial sensitivity and competitive advantage that private companies have invested in any data they collect and maintain. These may be overcome through the use of embargo periods, content or volume throttling and data cleansing.

From the Bureau's perspective the development of minimum acceptable standards of instrumentation, calibration and metadata would ensure that the data collected would be useful as scientific content. There would also be additional and unfunded costs in providing automated data streams for both provider and receiver.

*d) provide an update on existing data sharing initiatives in Australia, including the uptake of the credit reporting framework. Consider recommendations for improving participation in such initiatives*

The Bureau already shares data with various entities, domestically and internationally, as mentioned throughout this report.

### **TOR 3: Improved Access to Stored Data**

*Identify options to improve individuals' access to public and private sector data about themselves and examine the benefits and costs of those options. The Commission should:*

- a) *examine how individuals can currently access their data, including data about them held by multiple government agencies, and develop recommendations to streamline access*

The Bureau holds limited data that allows identification of individuals or entities, usually where the observation area is particularly remote and unique. The relative size of this data is small in comparison to the volume of data currently shared. In those circumstances we ensure we meet privacy conditions and cleanse or aggregate data to maintain anonymity.

- b) *identify datasets, including datasets of aggregated data on consumer outcomes at the product or provider level, that would provide high value to consumers in making informed decisions and any impediments to their use. Develop guidance to assist in identification of other high value datasets*

Currently, this does not fit into the Bureau's operating model.

- c) *examine the possible role of third party intermediaries to assist consumers in making use of their data.*

Currently, this does not fit into the Bureau's operating model.

## TOR 4: Benefits and Costs of Standardisation

*Examine the options for, and benefits and costs of, standardising the collection, sharing and release of public and private sector data.*

Standardising the collection, sharing and release of public and private sector data would prove problematic for the Bureau as we are obligated to conform to a wide range of scientific domain-based international standards. Further, the Bureau must constantly contend with the four main attributes of big data—data velocity, volume, variety and veracity. Typically, other government agencies may find one or two of these attributes challenging, while the others will remain stable or are a non-issue. Additionally, the Bureau shares data with peer agencies across the globe for which established standards are already in use. As a further complicating factor, the Bureau must collect and share data over a significant geographic area.

The Bureau already approaches the data it owns with the intent of standardising and sharing to the maximum extent possible. The exceptions to this are where conditions apply to data shared under international treaties referred to in earlier parts of this submission.

The Bureau also plays an important national role in standards setting for other domains including national water and flood information (for example, the Water Data Transfer Format), supporting aggregation and interoperability of meteorological data (for example, through the Weather Observations Website) and developing national approaches for interoperability of environmental monitoring site data across environmental domains (National Environmental Monitoring Sites Register). The Bureau's experience and success reflects the importance of an authorising authority, establishing effective governance arrangements and adequately resourcing the function. Such a model can yield significant benefits in achieving national data interoperability where a series of entities holistically manage the data lifecycle. Existing Bureau models could be readily leveraged and applied to achieve similar national outcomes for related environmental domains including air quality information, water quality, soils data and third-party observations of climate parameters (including rainfall, temperature and solar observations).

Data collection and management should remain as flexible as possible to allow for the incorporation of disparate data, and collections in downstream systems. Any form of standardisation is a time-bound phenomenon that requires stability for a period of time to gain benefit against incurred development costs.

## TOR 5: Individual and Business Trust

*Examine ways to enhance and maintain individuals' and businesses' confidence and trust in the way data are used. Having regard to current legislation and practice, advise on the need for further protocols to facilitate disclosure and use of data about individuals and businesses while protecting privacy and commercial interests and, if recommended, advise on what these should be. The Commission should:*

- a) balance the benefits of greater disclosure and use of data with protecting the privacy of the individual and providing sufficient control to individuals as to who has their information and how it can be used*

The Bureau supports open access to data consistent with the comments made in this submission.

- b) benchmark Australia's data protection laws, privacy principles and protocols against leading jurisdictions*

The Bureau supports open access to data consistent with the comments made in this submission.

- c) examine whether there is adequate understanding across government about what data can be made openly available given existing legislation*

The Bureau supports open access to data consistent with the comments made in this submission.

- d) consider the effectiveness and impacts of existing approaches to confidentialisation and data security in facilitating data sharing and linking while protecting privacy*

The Bureau supports open access to data consistent with the comments made in this submission.

- e) consider the merits of codifying the treatment and classification of business data*

The Bureau supports open access to data consistent with the comments made in this submission.