Submission to the Productivity Commission's Data Availability and Use public inquiry

Centre for Policy Development
July 2016

Author: Geoff Shuetrim

MELBOURNE
15-31 PELHAM STREET
CARLTON
VIC 3053
(03) 9929 9915

SYDNEY
LEVEL 6, 115 PITT STREET
SYDNEY
NSW 2000
(02) 8199 9407

www.cpd.org.au
ACN: 124 425 896
ABN: 82 124 425 896

About

The Centre for Policy Development (CPD)

CPD is an independent, non-partisan public policy institute. Our mission is to foster an Australia that embraces the ‘long-term now’. In doing so we seek a future for Australia based on shared prosperity and sustainable wellbeing. One of our three key research programs - Effective Government - is dedicated to understanding the role for an active, capable and effective government in the 21st century.

Geoff Shuetrim is a member of CPD’s Research Committee.

Table of Contents

Overview 4

Individual access to personal data 5

Utility bill verification 5

Utility vendor selection 5

Personal data for delivery of migrant settlement services 6

Credit assessment case study 6

Security considerations 7

Encouraging adoption 7

Maximising data value 7

Imposing a low usage burden on users 8

User confidence in the system 9

Imposing a high cost to attacking the system 9

Minimising the value of data to attackers 10

Data-access and security-breach notifications 10

Delegating access to personal data 11

Recommendations relating to personal data accessibility 11

Open access to publicly-funded research 13

Recommendations relating to publicly-funded research 13

Publicly-owned natural history multimedia 15

The current mosaic of multimedia resources 15

Applying existing open data policies consistently 16

Funding collection publication 17

Extending the range of multimedia resources that are available 18

Recommendations for open access to multimedia resources 18

Open and structured financial reporting 20

Recommendations for open structured financial reporting 21

Australian Bureau of Statistics web services 22

Publish fully structured rather than semi-structured data 22

Format choices for structuring statistical data 22

Providing access to more detailed data via web services 23

Recommendations in relation to structured ABS data 24

Overview

> "Data! Data! Data!" he cried impatiently. "I can't make bricks without clay."
> – Sherlock Holmes in "The Adventure of the Speckled Band"

This submission has been prepared in response to the "Data Availability and Use" issues paper, released by the Productivity Commission in April 2016. It addresses data availability and use across a disparate range of fields. In all fields, the focus is on:

1. increasing data availability;
2. increasing data usability; and
3. ensuring that transfers of sensitive data use infrastructure with appropriately high levels of user control, privacy and trust.

In the context of this submission, data is deemed to be available if it can be accessed via the Internet. It is deemed to be usable if it is structured in a way that facilitates automated provision of data through web services in a format that enables automated consumption by recipient systems.

In some cases, where data is not sensitive, this submission argues for open and unrestricted network accessibility. In other cases, much greater control is required around how data is made available.

Where greater control is required, this submission sets out recommendations for how data should be provided. Those standards should be mandated at a national level in Australia. There is a role for national level IT infrastructure to support those standards. This submission describes security features that should be required of that IT infrastructure and the services it would provide.

This submission focuses specifically on the need for greater control over data accessibility in the context of data about individuals. Made available with appropriate user control, privacy, and trust, that data has great potential to generate social, economic and environmental benefits. If the individuals are not empowered to take full ownership of their control over who has access to their data and when, trust will be undermined. Without trust, many of the economic and social benefits of improved data access will not be realised because people will not opt in.

Individual access to personal data

Both the 2015 Competition Policy Review Inquiry and the 2015 Financial Services Inquiry recognise that there is a lot of value to be generated by increasing access to usable data about individuals.

Sensitive but valuable data about individuals is held by both public and private sector organisations. It includes:

- High frequency electricity, gas, and water consumption data, increasingly sourced from smart meters;
- Telecommunications and network service usage from telecoms providers;
- Financial account and transaction data from banks and other financial services providers;
- Insurance policy and claim details from insurers;
- Medical information from GPs, hospitals, the Medicare system and other health service providers;
- Public and private system education participation and achievement data;
- Employment services usage data;
- Public and private transport usage information.

In the hands of the right people, this data can improve both consumption and investment decisions. It can also create new opportunities for experts in various fields to generate value by acting as their agents, leveraging the additional data to provide them with suitably tailored and appropriate advice.

The following case studies provide examples of these potential benefits.

Utility bill verification

In the same way that restaurant customers are given an itemised invoice at the end of a meal, customers should also be able to independently verify the amounts in the invoices for utility services. This has become increasingly difficult as utilities gather richer usage data through smart meters and the like, and reflect that richer data in their pricing models.

It should not be acceptable that utility customers cannot independently reconstruct and so verify that they are being correctly invoiced for the services they receive. As a first step toward empowering customers, service providers should provide access to the same usage data that drives their invoicing. That data needs to be usable. It needs to be digital. It needs to be structured to support automated extraction and analysis. To meet these requirements the data should be exposed to the customer, upon request, through web services.

Ideally utilities would also expose their pricing algorithms as web services so that customers or their chosen agents can feed actual or scenario usage data back into the pricing algorithm to provide an additional check of their invoices and, more importantly, to assist them with decisions about how to modify their behaviour and their investment choices to better manage their utility costs.

Utility vendor selection

Customers should also be able to compare pricing systems across vendors and, again, they require access to their own usage data to do so. If they, or their agents, could feed actual or scenario data into the pricing algorithms exposed by utilities, value comparisons become much more informative and empowering.

Note that responsibility for tailoring data formats, so that usage data from one utility could be used with the pricing web services of another utility, would be a value-adding service that could be offered by suitable agents acting on behalf of customers.

Personal data for delivery of migrant settlement services

Each year, AMES Australia provides settlement, education, training and employment services for approximately 35,000 refugees and newly arrived migrants. They do this by working closely with community, business and the Government. They work with a wide variety of specialist settlement and mainstream agencies including education providers, health providers, real estate agents and community organisations to provide these services. They have strong relationships with employers to match job seekers with job opportunities, helping new refugees and migrants to gain work experience and paid
employment as a key to their independence in Australia.

AMES Australia begin their support with arrival into Australia. New arrivals, who come as part of the humanitarian intake, are met at the airport by AMES representatives. AMES works to ensure that these new arrivals have access to appropriate accommodation, health and education services and provides orientation to life in Australia as part of the Humanitarian Settlement Services contract.

These services would be easier to deliver if AMES Australia representatives had access to a more complete set of information about arrivals prior to the first meeting. The standard information that is currently provided with respect to humanitarian entrants is the number of people in the group, their names, ages, and ethnicity. For the cohort currently arriving from Syria, additional health screening information is being provided, indicating that there is capacity to improve the quality of pre arrival data.

AMES Australia understands that coordination of information and collection of consistent and comprehensive data from refugees pre departure is difficult and complex and hindered by a number of competing priorities. However, within these acknowledged limitations, access to increased data would be highly beneficial in providing services. AMES representatives would be in a position to prepare more effectively for new arrivals if they also had access to more comprehensive information including information about about health, education and previous work experience, English proficiency and any history of trauma related to their refugee experience. There may be new ways in which refugees
themselves could be involved in improving this information flow.

If humanitarian entrants could opt-in to sharing this information with AMES Australia, prior to their arrival, and if the Government had the capacity to share this information at the individual level with AMES (and other agencies contracted to deliver HSS), settlement and associated services could be delivered more effectively and efficiently from the outset. While the need for information is particularly critical for the humanitarian cohort, expanding access to data for other groups of migrants could also result in more targeted services that could support successful settlement.

Credit assessment case study

If people, and indeed other entities, were able to access their full financial history and share that history as they see fit, they would be able to provide a much clearer picture of their credit-worthiness. By giving potential lenders access to their full history, truly creditworthy individuals will be able to mount a stronger case for funds. Customers who are less creditworthy may not volunteer their information and that decision would also be taken into account by financial institutions that they approach for funds.

This ability to share detailed financial histories with financial institutions will improve lending standards. It will improve the ability of financial institutions to identify creditworthy customers who would otherwise be unable to provide sufficient evidence of their reliability. Most importantly, it would lower the barriers to entry for new financial institutions that do not have access to their own customer bases with lengthy financial histories. This, in turn will increase competition and drive down the prices of banking services.

Security considerations

Ours is now a digital economy. As Nicholas Negroponte put it in the last decade of the 20th century, transactions are increasingly manifesting as a movement of bits rather than atoms. Twenty years later this transformation still continues, as growth in the "Internet of Things" expands the range of devices that can send and receive data.

In any economy, economic performance improves as transaction costs fall. This is the case for transactions involving the exchange of objects. It is also true for digital transactions involving the exchange of information. By reducing those transaction costs, the performance of the digital economy can be enhanced. The promise of these performance enhancements have underpinned the business case for Australia's investment in the physical infrastructure that is the National Broadband Network.

There are however, other ways in which Australia can improve the efficiency of digital transactions. Much of the data being exchanged is personal in nature; information that people do not wish to be available in ways that they do not control. If people cannot trust that the security framework underpinning their data exchanges gives them that control and privacy, then transactions have a potentially high cost. For Australia's economy to thrive as the digital economy matures, it is just as important for Australia to invest in getting the security framework right as it is to invest in the underlying network infrastructure.

While detailed personal data can be used in very positive ways, enabling consumers to improve their vendor selection decisions, their investment, consumption and savings decisions, and their usage of public services, the data is extremely sensitive.

On its own, or worse, in combination with other data, personal information can be used in ways that are highly invasive and detrimental to the individual. With these high risks, no one should be required to make this data about themselves accessible via the Internet. Making such data accessible to themselves and/or others needs to be strictly opt-in.

Opt-in adoption can be a slow process. To encourage adoption, it is crucial that:

1. the personal benefits of opting in are high, as manifested in the value of the transactions that people can perform if they opt in;
2. the effort required of users to understand and control access to their personal information is low;
3. people have high confidence in the measures taken to ensure confidentiality of data and the authenticity of data-transfer participants are sufficient;
4. the difficulties involved in gaining unauthorised access to data about large numbers of individuals are sufficiently high that there is are low incentives to make attempts to compromise data systems;
5. the value of data that has been accessed in an unauthorised manner is kept low, so ensuring that the costs of unauthorised access to data also remain low for the affected individuals; and
6. people have timely notifications of data accesses and, crucially, notifications about potential security breaches.

Each of these requirements are addressed in detail below.

Encouraging adoption

Maximising data value

Members of the Australian public generally do not have the skills required to work with structured data obtained via web services. Most people can work with simple summaries of data, made available as web pages.

However, data in those unstructured or semi-structured formats are not as easily combined with data and algorithms from other sources. To maximise the value being generated, personal data needs to be published through web services and it needs to be structured.

Data on individuals should be available from many organisations. These organisations have very different IT infrastructures and will have different preferences in relation to how they structure data that they make available through web services.

Moreover, with many different types of personal data being published through web services, even if Australia achieves standardisation in some fields, there will still be considerable differentiation across fields (e.g. medical data compared to data about electricity consumption).

While data format standardisation formats might be attractive, it should not be mandated by government because the delays involved in brokering agreement on standards will cause unnecessary delays.

Instead, the data sharing system needs to enable provision of third-party data consumption and interpretation services on behalf of individuals. The provision of those third-party services will have value. The third-party data intermediaries will be able to fund their operations by charging for making this value accessible to their users. Delegation of responsibility for accessing and interpreting personal data is addressed at greater length later in this submission.

Though the Australian Government should not mandate standards for data formats, it should mandate standards for over-the-wire data security. Specifically, the Australian Government should require Transport Layer Security for all exchanges of personal data to ensure that those exchanges are based upon private and reliable connections between parties whose identities have been authenticated. The connections need to be private in the sense that they cannot be intercepted and understood by a third party. They need to be reliable in the sense that accidental or intended corruptions of the transmitted data are identifiable. They need to be between authenticated parties in the sense that the sending party needs to know that they are sending the information to a recipient known to have the rights to access the data and the receiving party needs to know that the data is being provided by the right organisation rather than some imposter.

The low-level encryption algorithms also need to be easily adapted to be resistant to quantum attack as the threat of quantum attack becomes possible. While quantum computing is a nascent technology, it is undergoing well-funded research, focused specifically on decryption applications and the low-level encryption algorithms in widespread use today will become vulnerable to quantum attack.

Imposing a low usage burden on users

A system of widespread sharing of personal data cannot lead to a proliferation of user authentication systems. It is not acceptable to impose an ever-growing set of user identities and associated passwords, fobs, tokens and other authentication mechanisms on users. Such a proliferation leads to inconvenience and it leads to behaviours that undermine the effectiveness of the authentication systems. Provisioning of an electronic identity for a user should be a once-only event, involving a high level of integrity around the identity verification process, similar to the processes around provision of a passport.

It is also important that people are able to easily authenticate their ownership of their electronic identity. There are many authentication systems that can be applied. They have widely varying levels of complexity and reliability. Ideally, the authentication system would be as simple as logging in on a vouched-for device, with a single user name and a single PIN. This simplicity is achievable while also ensuring a sufficiently low level of risk that a user's authentication name and PIN can be stolen and used to fraudulently access personal data.

Once they have authenticated ownership of their single electronic identity, people must be able to use that electronic identity to manage access to the data about themselves, regardless of which organisation is providing that data. The Open-ID Connect system, layered on top of the OAuth 2.0 protocol, should be used to enable that re-use of the user authentication.

User confidence in the system

It is important that the Australian Government owns and runs the Open-ID Connect provider service. While identities established with other private organisations can be re-used through the Open-ID Connect or similar approaches, they have a number of drawbacks that make them unsuitable. Specifically:

- The Australian Government has no control over the integrity of the identity verification process that provisions a user with an electronic identity.
- The Australian Government has no control over the quality of the authentication processes used by those private organisations.
- There are already significant concerns about the amount of information that such many such organisations have about their customers. It would be unacceptable to many if those organisations became integral to user authentication aspects of a system responsible for transmission of people's sensitive personal data.

Imposing a high cost to attacking the system

The technical foundations providing data security for users is never going to be comprehensible to those users. Instead, confidence in that system must develop over time as it demonstrates itself to be resistant to security violations. On the flip-side, confidence will quickly erode if the system shows itself open to being compromised in ways that lead to real costs for users.

To ensure that system credibility develops over time, the Australian Government needs to implement the user identity and authentication system using a foundation with the following characteristics:

- The authentication secret, known only to the individual, must not be stored on any device.
- Authentication secrets must always be obfuscated within the encrypted channel providing multiple layers of defense. The security framework must ensure that all components required for authentication are never in transmission together and never in plain text.
- Any authentication information stored on user devices must not be usable, in isolation, to determine authentication secrets.
- Any authentication information on the servers responsible for verifying authentication attempts must not be usable, in isolation, to determine authentication secrets.
- Data exchanged over the wire during authentication attempts must not be usable, in isolation, to determine authentication secrets, even in unencrypted form.
- Authentication-related information stored on servers and on local devices should have a very short life-span, ensuring that even if such data is obtained, whatever authentication value it has is quickly eliminated.
- Authentications attempts must only be accepted if they combine the user ID, the authentication secret, and a device that known to be under the control of the user.
- Users need to be able to authenticate with the system and then confirm their permanent or temporary control over a specific device that they wish to use in an authenticated manner.

A system with the features listed above provides a very small target for attackers. To maliciously authenticate, the attacker needs to discover both the user ID and the authentication secret. The attacker also needs to gain control of a device that has been vouched for by the user whose identity is being stolen.

Importantly, breaches of server security will not provide sufficient information to steal electronic identities. Decentralisation of authentication information removes the "honeypot scenarios" that motivate malicious attempts to breach server security. Thus, the system reduces the extent to which authentication servers are high-value targets because servers are completely neutralised as an identity attack vector.

Transmission interceptions will also not be sufficient, even if combined with access to the data held on authentication servers. Device theft will not facilitate identity theft unless also associated with a social engineering attack that simultaneously reveals the authentication secret.

Minimising the value of data to attackers

The system needs to ensure that personal data, exposed through web services, does not also convey the identity of the individual to which that data relates. The identity of the individual described by the data should only be possible to determine for the user or any organisation that the user has approved to access that data. By separating the data from the identity of the individual described by the data, the value to attackers is significantly reduced.

Data-access and security-breach notifications

To enhance transparency and to build an awareness of and confidence in the security framework, people need to be provided with real-time notifications about data accesses, including details of the data provider, the entity accessing the data, and the nature of the data exchanged.

The notification system would serve two purposes. Primarily, it would ensure that users are quickly made aware of unexpected and unintended accesses to their personal data. This will enable them to alter their data access controls in a timely manner to better suit their privacy requirements. The notification system will also to increase user trust in the security framework. By making the framework more visible, informing the user about the level of monitoring that is occurring in the background, the level of trust in the system will improve more rapidly.

The Government also has a role to play in establishing a legal requirement that security breaches involving personal data trigger rapid notification to all individuals who may have been affected. While such breaches may not be identified at the time they occur, it will still be vital that they are disclosed to affected individuals within hours of being recognised. This gives affected individuals the best opportunity to mitigate the impact of the breach event. Assuming the security framework is adequate, it will also gradually improve the level of trust in the system, as breach events occur infrequently and are restricted to impacting on a very limited (ideally one) number of individuals.

Delegating access to personal data

To encourage widespread use of personal data, it is crucial that people are willing and able to delegate responsibility for sourcing data from web services and for manipulating and interpreting that data in a manner that is valuable to them.

For people to be willing to delegate data access to third parties, they need to:

- trust that the third party is a reputable organisation;
- be confident that they are delegating access to the organisation that they intend to delegate access to and not some imposter; have
- have simple access to a centralised control system where they can alter their data access delegations in real time.

To address the issue of counterparty trust, the Government's identity management system needs to provide an identity verification process for original data providers and third-party handlers of data that enables system users to trust the identities of the parties that they are exchanging information with.

To address reputability, the identity verification system should be coupled with a system of regulation, perhaps managed through ASIC, that provides a level of assurance around the reputability of organisations that are authorised to be data access delegates.

Finally, the Government needs to provide a data access delegation control service to individuals with electronic identities so that they can manage their data access delegations in a manner that does not depend on facilitation by providers of third party data services. The control service should be accessible through a web front-end.

For the delegation control system to be effective, the authentication framework needs to be designed in a manner that forces all accesses to data services to be confirmed by the centralised data access delegation control system.

Recommendations relating to personal data accessibility

1. Personal records and transactions data about individuals should only be accessible via web services if individuals opt-in at the level of the information provider.
2. Require that personal and transaction data about individuals be published through web services in a structured format that facilitates automated consumption by other computer systems.
3. Do not mandate specific formats for structuring the personal data that is made available by different organisations.
4. Do mandate that all data exchanges are encrypted, using the Transport Layer Security Protocol 1.2.
5. Ensure that individuals are able to delegate responsibility for data access, processing and interpretation to third parties.
6. Ensure that individuals can manage their data-access delegations, turning them on and off, as desired, through a centralised web-accessible facility.
7. Government should provide a single user identity verification service, similar to that used for issuing passports. Users that have verified their identity should be provisioned with an electronic identity that they can use to manage access to data about themselves.
8. Authentication, the process of a user proving that they own their electronic entity, must be as simple as providing a user name and a PIN on a device that is known to be under the control of the user with the given electronic identity.
9. The Government should become an Open-ID Connect provider so that participants in the data sharing system can rely on a user's authentication with the Government electronic identity service instead of implementing their own varied authentication systems.

10. The identity authentication system, provided by the Government should be a low-value target for attackers, ensuring that confidence in the system does not get undermined by large-scale data thefts from breached servers.
11. The Government should provide a third party data handler identity verification service, ensuring that organisations wishing to provide data services are registered and regulated.
12. The Government should provide a data access delegation control service to individuals with electronic identities so that they can manage their data access delegations in a manner that does not depend on facilitation by providers of third party data services.
13. The Government should design the authentication system so that all accesses to data services need to be checked through the centralised data access delegation control system.
14. The security framework should ensure that users can receive real-time notifications about data accesses.
15. The legislative framework should require that all security breaches be reported to individuals that may have been affected, within hours of the security breaches being identified.

Open access to publicly-funded research

In 2013 the Australian Research Council (ARC) has adopted an open access policy. This initiative requires that:

> "Any publications arising from an ARC supported research Project must be deposited into an open access institutional repository within a twelve (12) month period from the date of publication."

This requirement for open access to publicly funded research findings is an important and welcome development. However, it comes with the caveat:

> "In cases where researchers may not be able to meet the requirements because of current legal or contractual obligations, Final Reports must provide reasons why publications derived from a Project, Award, or Fellowship have not been deposited in an open access institutional repository within the twelve month period."

The Australian Government should require greater transparency in relation to how extensively this caveat is applied. Specifically, the ARC should publish a list of finished research projects, along with links to any publications that have been deposited into an open access institutional repository. It should also publish summary statistics, indicating the extent to which their openness objectives have been met.

Along with greater transparency, the ARC should consider extending their open-access requirements to also cover the data sets used in the research and the code implementing the analytics applied to that data.

Increasingly, researchers are recognising the value that comes with replication of others' work as a first step toward building upon or repudiating their findings. Efficient replication depends in large part upon access to the datasets used in the work being replicated. Replication is also made more efficient if researchers also release the code used in their analytical work. Replication of a researcher's own work is also a valuable validation of the conclusions reached in that research.

For these reasons, the ARC open access policy should be extended to include:

- the data used in the research (including multimedia resources produced as part of the research programme and deemed to be relevant to the research findings); and
- the code used to produce results from the underlying data.

ARC grants are not the only form of Australian public funding for research. Many organisations, funded by Australian state or federal governments are also undertaking research and their research should be subject to the same open-access requirements as apply to research funded by the ARC.

Recommendations relating to publicly-funded research

1. The ARC should publish a list of finished research projects, along with links to any publications that have been deposited into an open access institutional repository.
2. The ARC annual report should contain a summary of the extent to which ARC funded research has been made available with open access.
3. The ARC open access policy should be extended to include provision of open access to underlying data sets used in the research that they fund.
4. The ARC open access policy should be extended to include provision of the code implementing the analytics applied to the underlying data sets used in the research that they fund.
5. The ARC open access policy should be extended to include multimedia resources (drawings, photos, videos, and sound recordings) that were gathered by researchers as part of their research programme and that were found to be relevant to the research results.

6. The Australian Government should consider extending the coverage of the open access policy to include research that has been publicly funded through other channels. This would include CSIRO research, research programmes within museums, and any research projects at university that have attracted public funding through other channels.

Publicly-owned natural history multimedia

In the face of climate change, competition and predation from introduced species, and land-clearing, there is a greater-than-ever need for Australians to support the measures necessary to preserve our unique and productive natural environment.

Because of their cost, people will only support these measures if they appreciate the importance of Australia's natural history. A diverse set of public institutions are the custodians of collections of multimedia resources recording Australian natural history. With Australia's great biodiversity, extensive land area, and relatively small population, these collections are often the only available documentation of the appearance and behaviour of Australian wildlife.

Many of the publicly-held multimedia resources are not available under open-access licensing terms and are correspondingly under-utilised. Steps should be taken to ensure that all such collections are available under licensing terms that facilitate maximum exposure to the Australian public.

Allowing open access to and reuse of publicly held multimedia resources will reduce pressure on custodian institutions to invest in their own costly systems to give more controlled access to the multimedia collections. It will also spur innovation by other public and private organisations that are able to incorporate publicly held multimedia resources into their own offerings to the Australian public.

The current mosaic of multimedia resources

Examples of offerings that could leverage appropriately licensed multimedia include:

- Wikipedia;
- iNaturalist, a species description and record capture system;
- Gaia Guide mobile field guides;
- Australian Online Reptiles Database;
- Fishes of Australia, operated by OzFishNet; and
- Fishbase.

Constructive steps are being taken to make multimedia resources more available. For example the CSIRO now provides open access images in the Australian National Fish Collection. This has substantively improved the completeness of the image library underpinning OzFishNet's "Fishes of Australia". However, these steps toward open access are incremental, inconsistent across organisations and collections, and poorly documented.

Some multimedia collections are not published digitally at all. Many have licensing terms that restrict access and reuse in ways that prevent multimedia resources from having maximum visibility.

An incomplete list of publicly held multimedia collections is provided below.

- The digitised collections at various Australian museums (Australian Museum, SA Museum, Museum of Western Australia, Museum of Victoria, Queensland Museum collections) including the Queensland Museum's image library);
- The CSIRO's Australian National Fish Collection
- The Australian National Herbarium National Plant Image Index co-managed by the CSIRO and the Australian National Botanic Gardens;
- The CSIRO's Australian National Insect Collection;
- The CSIRO's Australian National Wildlife Collection; and
- The NSW Office of Environment and Heritage's threatened species database.

A quick review of the resources listed above highlights the inconsistencies in the way that the resources are documented, made searchable, and made available for reuse by others. At one end of the spectrum, the Museum of Victoria provides an excellent facility, where each multimedia resource is documented in a web page giving access to the resource (e.g. the resources associated with Verreaux's Frog), along with the relevant information about the content of the resource itself, its owner, the source institution providing the resource, and the licensing terms for usage of the resource. At the other end of the spectrum, resources are available upon inquiry only and are, in some cases either not available for reuse or are available for a fee.

Applying existing open data policies consistently

The Australian Government already has a policy governing how these resources should be made available. It is set out in the Australian Government Public Data Policy Statement, released in 2015. That policy states:

> "The Australian Government commits to optimise the use and reuse of public data; to release non sensitive data as open by default; and to collaborate with the private and research sectors to extend the value of public data for the benefit of the Australian public."

The Australian Government should ensure that this policy is applied consistently to all public multimedia data resources as well as other more traditional types of data. This would bring the Australian approach into line with that of the US where organisations like National Oceanic and Atmospheric Administration place their multimedia resources into the Public Domain.

To maximise transparency around the terms of use for publicly-held multimedia, the application of this policy should be set out in more detail. This is particularly important because some of the multimedia resources held by Australia's public institutions are not owned by those institutions. To address these ownership complications, the institutions that are custodians of the data should ensure that:

1. Each multimedia resource is associated with its own unique and stable Uniform Resource Identifier (URI).
2. Each URI should resolve to a web page that provides:
   a. direct access to the resource;
   b. information about the owner of the resource;
   c. the licensing terms associated with the resource;
   d. a summary of the content of the resource providing the details required to understand what the resource documents; and
   e. a link to the web service query that would allow automated access to the multimedia resource, along with the metadata describing it.

This approach to documenting multimedia resources is consistent with the approach adopted by Internet-focused custodians of open-access multimedia resources such as Wikipedia (e.g a Wikipedia Platypus image). Among other benefits, this publication approach allows responsibility for making the collection searchable to be delegated to Internet search providers. It also ensures that users of multimedia resources are able to reference the source of the resources that they have used in a consistent manner.

This approach to providing open access to publicly-held multimedia collections would deliver substantial benefits to undertakings like the CSIRO's Atlas of Living Australia, which attempts to provide web services that list the multimedia resources describing each species in their underlying database. Because of the limitations in the documentation of the resources that the CSIRO is aggregating with this service, the value of the web service is undermined by missing, unsubstantiated, or incorrect licensing details. These limitations undermine the value of the multimedia features of the web services exposed through the Atlas of Living Australia.

Funding collection publication

The process of digitising and publishing collections can be slow and costly. The importance of these processes is well recognised though and this is reflected in the fact that these processes are currently underway in many of Australia's public institutions. For example the Australian Museum has adopted an innovative approach to digitising its collection, by drawing on assistance from volunteering members of the public to review collections images extracting and documenting relevant meta-data based on the labels included in those images.

As these digital resources accrue, there is a strong temptation for public institutions to package them up in value-added resources that promote the benefits of their ongoing collections management work and potentially to raise revenue through fees for value-adding services.

A good example of this kind of initiative is the set of "field guides" to the fauna in each state that have been developed as mobile apps. through a collaboration between various Australian museums (e.g. the NSW fauna field guide released by the Australian Museum). They package up information and multimedia resources in mobile apps that are made available for free to the general public. The collaborative approach whereby the same code base was used for all apps has been an innovative way to mitigate development costs. The apps also provide exposure for the museums.

However, the field guides have still involved significant cost and involve a number of fundamental drawbacks. These include:

- A lack of ongoing funding to ensure that the code base remains usable on new releases of mobile device operating systems;
- A very limited number of species in each type (birds/reptiles/frogs/fish/invertebrates etc.), preventing the app from reliably performing its core role of assisting users to identify species they encounter; and
- No ability to capture and upload records of users' encounters with the fauna documented in the apps.

With the limited species coverage and limited functionality, the app-store reviews of these "field guides" is very mixed, with a significant number of users being negative about their experiences. This public feedback undermines the very objective of creating and releasing the apps in the first place.

Given these limitations, the apps are little more than electronic coffee table books, providing a restrictive window into the Museum collections. They are destined for obsolescence. This can be seen today because the code bases for the iOS and Android versions have been made available as open source by the Museum of Victoria. Neither of these public code bases have been updated since their original release three years ago.[1]

Unfortunately, this process of innovation, funding exhaustion and eventual obsolescence is repeated across the public sector. Examples include:

- the nematode key system developed for Windows 95 by the Department of the Environment and still available on their website;
- the various hard-copy and electronic field guide books that are still being published by the CSIRO;
- the early foray into mobile applications by the Australian Museum in the form of its Field Guide to Australian Frogs; and
- the sea-slug forum, originally managed by the Australian Museum.

The sea-slug forum is a spectacular example of a failed initiative. Originally, it was so successful that it became the pre-eminent world identification resource for Nudibranchs and other exotic sea-slugs. After some minor

[1] There has been a bug-fix release in 2016 but that code update has not been made public.

problems with security breaches, the relatively inexpensive forum was entirely defunded by the Australian Museum. Its interactive capabilities were eliminated and the site is now an electronic mausoleum, with no further contributions since its defunding in 2010. The sad story is preserved in the last few comments posted to the discussion forum.

Making matters worse, the original focus on providing functionality rather than management of the underlying textual and multimedia data has meant that ownership of sea-slug forum data and the licensing terms under which it can be reused are unclear enough to prevent salvaging content from the site. This experience should guide the funding priorities for public-sector organisations with a responsibility for managing Australia's data.

The infrastructure funding necessary to provide open access to publicly-held multimedia resources should be prioritised ahead of funding for specific publication and promotional projects, like those listed above.

The cost of rolling out consistent and usable open-access to multimedia resources can also be significantly reduced by implementing the publication platform once and allowing that platform to be used by all Australian public institutions. The CSIRO's Atlas of Living Australia could be extended to perform this function.

Extending the range of multimedia resources that are available

Australia's research bodies, museums and universities together employ a large majority of the people with the expertise necessary to identify specimens. The valuable identification services provided by these experts are used by members of the public on a fairly ad-hoc basis. Often, members of the public send in images to collections managers at museums, requesting assistance with identification.

In cases where the multimedia resources provided by the public are of unusual species, or illustrate interesting behaviour, the public sector organization often publishes the images, with copyright and attribution to the member of the public. This is a valuable and productive service. It is excellent that the multimedia resources are often used to benefit the general public. However, it should be the default position, agreed with the contributing members of the public, that their relevant multimedia resources should be made openly available, through the public sector organization, with appropriate Creative Commons or similar licensing. That default position should not be necessary to get expert assistance. However, by introducing it as a default, the Australian government will be able to nudge members of the public into making their own contributions more widely usable.

Recommendations for open access to multimedia resources

1. The Australian Government's existing public data policy statement should be extended to explicitly include multimedia data.
2. Each multimedia resource needs to be associated with its own unique and stable Uniform Resource Identifier (URI).
3. Each URI should resolve to a web page that provides:
    a. direct access to the resource;
    b. information about the owner of the resource;
    c. the licensing terms associated with the resource;
    d. a summary of the content of the resource providing the details required to understand what the resource documents; and
    e. a link to a web service query that would allow automated access to the multimedia resource, along with the metadata describing it.
4. The infrastructure funding necessary to provide open access to publicly-held multimedia resources should be prioritised ahead of funding for specific publication and promotional projects for subsets of the multimedia resources and the organisations that manage them.

5. An extension to the CSIRO’s Atlas of Living Australia should be funded to enable it to be used as the multimedia resource publication platform that can be used by all public organisations in Australia.
6. Where public sector organisations in Australia provide expertise in relation to multimedia resources captured by a member of the public, the default position should be that those multimedia resources will be attributed to the appropriate member of the public and that they will be made available to others under a suitable Creative Commons or similar licence.

Open and structured financial reporting

Since December 1998, a global initiative has been underway to structure company financial disclosures in a way that facilitates automated consumption. This initiative began in the US at an American Institute of Chartered Practicing Accountants (AICPA) meeting. It quickly attracted participants from around the world. This process led to the creation of a global non-profit consortium: XBRL International.

Through XBRL International, a number of XML-based data-format specifications have been recommended for use in publishing financial reports with sufficient structure to enable automated data consumption. The core specifications were recommended in 2003 and have remained stable since.

In the financial reporting space, the XBRL initiative has been driven by the view that more accessible data, more precise data definitions, automated data discovery, improved data validation, and easier data extraction will all improve the efficiency of stock markets by lowering the costs of information discovery.

In the years since recommendation of the underlying specifications, a wide range of taxonomies have developed, setting out the definitions of data commonly found in financial reports. The US taxonomies include definitions of data appropriate for reporting under US GAAP accounting standards. Taxonomies for much of the rest of the world, including Australia, have been based on IFRS accounting standards. These taxonomies are updated annually, as accounting standards evolve. Countries, like Australia, develop their own taxonomy extensions to customise the core taxonomies to fit the specific nuances of their own accounting standards.

Companies wishing to provide open access to their structured financial data can publish electronic reports containing fully structured and self describing data, based on these taxonomies. Unfortunately, this has never occurred voluntarily except as part of experimentation with or demonstration of the XBRL standards by promoters of the standards. In all cases where publication of XBRL reports has been widespread, it has required government mandate. The US Securities and Exchange Commission (SEC) is the leading example of a mandated approach to XBRL reporting.

The absence of voluntary adoption reflects a number of factors, including:

- the complexity of the XBRL standard;
- companies' reservations about the increased clarity that XBRL brings around exactly how they are choosing to interpret and apply accounting standards;
- a lack of understanding, by business professionals, of the advantages of structured information, a lack of understanding of the nuances of financial reporting by information technology professionals, a lack of knowledge engineering expertise necessary to create appropriate software, and a communications gap between these three groups of professionals that is very challenging to cross;
- the significant cost of software designed to facilitate the creation of XBRL reports;
- The uneven quality of XBRL-capable software; and
- the relative diffuse influence of the main beneficiaries of XBRL reporting: the investor community.

Increasingly, mandates to publish in XBRL are occurring in other countries. XBRL International provides a useful summary of these mandates. As data becomes more available, intermediaries are emerging, making it more straightforward for data users to work with the increased volumes of structured data. An example provider of these data intermediation services is 28 msec, an organisation that translates raw XBRL data into data stores that are easily understood and efficiently queried. Among other services, 28 msec provide simple access to all data published through the SEC in XBRL format. Users can query individual reports or they can extract and compare data across all manner of different metrics, companies, and time frames.

Australia has already committed significantly to the XBRL standard, using it as a foundation format for the Standard Business Reporting Programme (SBR programme). For the most part, the SBR programme focuses on various types of tax and other reports to federal and state government entities. For those reports, the SBR programme provides tightly constrained guidance on the information required to be in each document.

The situation is different for company financial statements that need to be provided on an annual basis to the Australian Securities and Investments. The types of information contained in a financial report depends on the structure and activities undertaken by the reporting entity. Accounting standards reflect this dependence and need for flexibility. XBRL also reflects this need for flexibility. The ability to report structured financial reporting data to ASIC through the SBR system is similarly unstructured.

Currently the SBR programme provides guidance on submitting form 388 to ASIC using XBRL. Public companies, large proprietary companies, small proprietary foreign-controlled companies, registered schemes and trusts must lodge a copy of financial statements and reports annually accompanied by form 388. The form is a cover document, required to accompany the actual financial statements provided to ASIC. It can be submitted to ASIC in paper form or through the SBR system. It is unclear how many forms are being submitted through the SBR system.

More importantly, SBR and ASIC release the IFRS AU Taxonomy that is suitable for XBRL financial reporting to ASIC. Most recently, the 2015 taxonomy was published for financial reporting for years ending on or after 30 June 2015. According to the SBR website, it is:

> "based on the IFRS Taxonomy 2015 (as issued by the IFRS Foundation in March 2015) extended to include Australian specific disclosure requirements from the Corporations Act 2011, AASB accounting standards and ASX Listing Rules."

These XBRL taxonomies facilitate the creation and collection by ASIC of structured electronic financial statements for Australian companies. However, it is unclear whether they have ever been used. ASIC certainly do not make any such documents available to the public. This contrasts with the US where all XBRL reports are published through the SEC's Office of Public Disclosure.

Given the substantive steps towards adoption being taken elsewhere in the world, and given the upside of open and structured financial reporting, more should be done to ensure that Australia does not become less transparent, from a financial reporting perspective, than other developed and developing countries. Specifically, Australia needs to identify and set in place a workable and cost-effective pathway toward mandating the publication of open and structured financial reports. Given the lack of alternatives, XBRL should be the format required for these publications.

Recommendations for open structured financial reporting

1. ASIC needs to map out a pathway and timeline toward mandating the publication of all financial statements in structured electronic form.
2. The format used for structuring electronic financial statements should be XBRL, given that no other suitable or competing standards have emerged over the course of the last 18 years.
3. The pathway to full mandate needs to force companies to begin working with XBRL on a limited scale within the near term so that companies and their software vendors develop XBRL publication systems are a good fit to their specific requirements.
4. ASIC should begin the pathway with a very accommodating stance on tagging and numerical accuracy in XBRL reports. This accommodating stance should be tightened over time as companies traverse the necessary learning curve.

5. ASIC needs to highlight the connection between data quality and supporting business rules by developing and publishing machine-readable business rules that can be applied to XBRL financial statements to test the quality of information within the financial report.
6. ASIC should provide direct feedback to companies on detected errors and other data quality issues.
7. The pathway towards full mandate needs to incrementally extend the components of the full financial statements that are structured with XBRL over time until full coverage is achieved.
8. ASIC should publish all XBRL financial reports that it receives so that they are accessible to the general public at no cost.

Australian Bureau of Statistics web services

The Australian Bureau of Statistics (ABS) has made significant and valuable steps toward increasing the availability and accessibility of their statistical publications. In the pre-Internet days, statistics were published in hard copy or MS Excel spreadsheets on floppy disks and made available for a fee.

The ABS now makes its statistics releases available electronically via the Internet. These releases are published using a semi-structured Microsoft Excel® format. They are available at no charge to users with network access to the ABS website. This is a substantial improvement on the previous situation and the ABS is to be commended for making its data open in this way, well before such openness was being driven by government policy.

However, there are two related ways in which the ABS data availability could be further enhanced. These are addressed in the sections below.

Publish fully structured rather than semi-structured data

While the ABS should continue to publish its statistical publications in Microsoft Excel® format to support one-off requirements of individuals, the ABS should also publish its statistical releases as fully-structured data using web services.

The Microsoft Excel® format is suitable for casual access and usage by individuals. It can be read by most spreadsheet software and it presents the data in a form that is easily interpreted by people. However, it is less suitable for direct and automated transmission of data from one the ABS data storage systems to the data storage systems of regular statistics users. This is because the data is only semi-structured. The interpretation of the values in the spreadsheets is a function of the way that the data is organised by row and column and the way that it has been formatted. This approach to specifying the meaning of the data is brittle, substantially reliant on human interpretation of the data organisation. A minor change in the organisation of the spreadsheet can easily lead to breakdowns and errors in automated exchange of data between systems. This is why spreadsheets are not used for web-services in any other context.

Instead, the ABS should move to augment its spreadsheets with a publication format that is suitable for exposure through web services. This format needs to be self-documenting, ensuring that data transmission from ABS systems to user systems is not dependent upon human interpretation of the data and is not vulnerable to minor changes in spreadsheet structure.

Format choices for structuring statistical data

The next question is then, what specific format should be used to fully structure the statistical data being published by the ABS through web services? The natural fit is the Statistical Data and Metadata eXchange

(SDMX) format. SDMX is an XML format that has been specifically designed for the purpose of exchanging statistical data.

SDMX is an XML format that is used for the exchange of statistics data. It is supported, to some extent, by a wide variety of statistical agencies, especially within the European Union. An explanation of SDMX and its intended benefits provides valuable background on how SDMX should be able to improve statistical data exchange.

The SDMX format is stable, it has fairly widespread adoption among national statistical agencies, and it is recognised as an ISO standard. It is flexible enough to work with the ABS's wide range of publication requirements. It is simple enough to be quickly understood by IT professionals that need to work with the format.

The ABS has already invested resources in understanding SDMX and publishing in that format from its internal systems. Evidence of this can be found at a variety of web pages hosted by the ABS including:

- The ABS.Stat SDMX beta web service, with a diverse range of working examples;
- documentation of the ABS.Stat SDMX beta web service; and
- an early SDMX experimental site.

Unfortunately, these web resources contain many broken links, provide very limited information about the future of the web services and provide almost no information about how best to design systems that consume data exposed through those web services. They are also subject to important limitations in terms of the timeliness of the availability of statistical data through the Beta web services. Until the web services expose new releases at the same time as the data is published through other channels, there will be no value for data consumers to invest in connection to the web services.

The ABS needs to publicly commit to SDMX as a publication format for their data. The ABS also need to commit to specific timelines for transition to a production version of their web services, with full SDMX compliance and complete coverage of their statistical publications so that data users can plan their own investments with confidence. The ABS also needs to commit to making data available through the web services at the same time as it is available through other publication channels.

Given existing investments in data extraction from Microsoft Excel® spreadsheets and direct screen scraping of the ABS website, the uptake of web services would be accelerated if the ABS also provided extensive documentation and perhaps simple examples showing how best to design systems that can interpret data in the SDMX format. The SDMX community has created a wide variety of tools for users of the format and the ABS should leverage these resources in their own documentation.

For the ABS to make these commitments, its web-services investment programme needs to be adequately funded in a way that is directly tied to the web-services delivery project rather than out of the overall IT budget. The inadequacy of the current funding approach is evident in the limited progress that has been made toward the publication of web services over the last decade.

Providing access to more detailed data via web services

The specific set of statistical publications released by the ABS, and the specific numbers within them, are the result of decisions attempting to strike a good balance between giving Australia the statistical data required, making sure that the important data is not obscured by a vast amount of related but less important data, and managing the costs of publishing that information.

Today, if statistics users required additional, more detailed data to drill down into the headline numbers, such access needs to be negotiated directly with the ABS.

By publishing fully structured statistics through web-services, the ABS would be able to expose the more detailed data to the small number of interested users at minimal additional cost. By only publishing the additional data through web services, it would not obscure the headline information provided in the existing publication channels (hard copy and Microsoft Excel® spreadsheets).

The limitations on what data is made available by the ABS would then depend upon confidentiality and privacy issues, and data quality and reliability issues rather than the ease of publication.

Recommendations in relation to structured ABS data

1. The ABS should continue publishing all statistical publications as Microsoft Excel® documents to support casual access by people.
2. The ABS should augment its spreadsheet data publications with web services that make all statistics available in a fully structured format.
3. The ABS needs to publicly commit to using the SDMX format for structuring the data exposed by their web services.
4. The ABS needs to commit to a timeline for rolling out a production version of their statistics web services, achieving full SDMX compliance and 100% coverage of their publications and synchronized of publication times with other distribution channels.
5. The ABS should accelerate adoption by providing ABS-specific documentation and examples showing best practices for consumption of their SDMX data.
6. The investment in web services at the ABS should be funded in a way that is tied directly to the web services project. It is likely that the project will require additional funding, on top of the existing ABS budget.
7. The ABS should consider extending the amount of detailed data published through the web services channels, even though that data is not also published through more traditional channels.