

1 IDENTIFY – STEP 1 OPEN DATA PROCESS

With the Declaration of Open Data, agencies need to consider what datasets are available for release. Agencies create, collect, process, preserve, maintain and disseminate all sorts of data in the course of doing business.

There are several activities that need to occur through the identify step of the open data process. Including identify data, confirm who is the Data Authority; establishing if there are third party rights; and considering the value and quality of the data identified.

1.1 IDENTIFY DATA

Agency data may be held in databases, reports, photographs, journals, videos, visualisations or exist in 3D, geospatial and numerical form. There are a number of ways agencies can identify existing and potential datasets for release.

1.1.1 *Public Facing Information*

Most agencies already publicly release information e.g. on their websites. This type of data can be considered first as minimal effort may be required to classify, mark and release the data in an open format and open licence.

Agencies should review and consider existing systems and public facing mediums:

- website content
- publications such as annual reports
- data published in tables
- public service announcements.

1.1.2 *Inventory Method*

Agencies should review their Information Assets for open data candidates; agencies are required under the ISMF to hold an inventory of Information Assets. Agencies may consider expanding this inventory to record open data process decisions. Agencies should also review other sources for datasets, which may include:

- KPI's and performance statistics
- research or administrative data records
- reports or data extracts generated from information systems
- reports or data used to make administrative or policy decisions
- data from commissioned research or consultation
- data supplied to a third party, including Commonwealth reporting
- databases or data catalogues.

1.1.3 *Proactive Disclosure*

Freedom of Information (FOI) requests, regular commonwealth reported data and across jurisdictional reported data should be considered as open data candidates. Resource efficiencies and significant savings can be achieved through the proactive disclosure of reported or requested data. Improved ability to make decisions can also be achieved through early access of data.

Agencies are encouraged to complete an annual review for dataset FOI requests to identify frequent requests for data. Significant resource efficiencies and expenditure savings could be realised through the proactive release of regular requested FOI data. Advice from your Agency FOI Officer must be received before considering this data for possible release.

Information about FOI is also available through the [Archives SA website](#).

1.1.4 Business Planning Method

Data can be identified through a business planning method where the agency identifies a specific goal or outcome to achieve through their strategic planning, business planning, or information communication and technology (ICT) planning processes. Agencies should consider the following through their planning process:

- business projects that are enabled by ICT should consider open data principles and plans. Business requirements documentation should identify data collection, consider the approach for release of data and automated release opportunities
- identify opportunities where third parties could add value to services if data was available for re-use e.g. could someone create an application from your data
- consider planned legislative or policy changes for open data opportunities
- review strategic programs to align activities with open data projects
- consider a problem you have that your agency does not have resources, funding and expertise to solve.

1.1.5 Community Demand

Agencies are encouraged to engage and collaborate with citizens and industry as per the [Better Together Principles for Engagement](#). Channels such as [YourSAy](#) and [Data.SA](#) enable the community to request data. Demand from the community should be given a high priority if the release of data could in some way provide public benefit either directly or when value added by a commercial entity. Data that cannot be released immediately should be considered in future plans.

1.2 IDENTIFY THE DATA AUTHORITY

Once a dataset is identified, the Data Authority must be assigned. The Data Authority has administrative or legislative responsibility for a specific government business function to which the data is integral. The Data Authority has responsibility for the existence, protection, and use of this dataset. In most cases, the Data Authority will be a Director or Executive.

If data has been sourced from another State or Commonwealth Government agency, direction and approval to re-classify, re-mark and openly release the data must be received in writing from their Data Authority or information owner. Agencies are not permitted to alter the classification or markings of information they do not own.

In many cases the Data Authority will already be known, if you are unsure consider the following sources to identify the Data Authority:

Legislative Delegations	Review legislative delegations approved for the administration of the legislation that empowers the SA Government to collect the data.
Public Sector Delegations	Delegated responsibility for an operational function as per the Public Sector Delegations.
Policy and Accountability	Public sector policies or accountability mechanisms may define who is responsible for the data.
Agreements or contracts	Review any agreements or contracts with third parties during the collection of data to determine what authority exists for the publishing of data e.g. memorandum of understanding, terms and conditions of collecting data.

1.3 IDENTIFY THIRD PARTY RIGHTS

Data should be checked for third party rights. Third party rights are derived from:

- copyright of intellectual property
- terms and conditions of use specified when collecting data.

1.3.1 Intellectual Property and Copyright

Agencies must not make datasets available that contain Intellectual Property (IP) rights of a third party where the SA Government does not have permission from the copyright owner or term of use licence is not provided. This includes circumstances where an agency co-produces a dataset with another government agency (state or federal level).

Under the *Copyright Act 1968*, the owner of copyright has exclusive rights to reproduce, publish and communicate the work and to license others to do so. The government generally holds copyright in works created by its employees, but not in any third party material that is incorporated into those works unless otherwise specified under contract.

Copyright generally exists for at least 50 years. Accordingly, data, records or images older than this should be considered for release.

View [The Government of South Australia Intellectual Property Policy \(PDF\)](#). Agency specific IP policies may also apply.

It is recommended that agencies consider the following actions when using data from third parties:

- identify and document the input sources used in the creation of the work
- identify and review publishing restrictions that apply to the additional inputs
- obtain an agreement or terms of use license for using the data
- the source of original data is Attributed in works created
- any use, re-use or manipulation of external data should be documented
- when in doubt seek permission from the third party or seek legal advice.

Where a dataset has been developed by a third party under a contract or agreement, copyright ownership may be dealt with specifically and should be checked.

Existing contracts and agreements

Agencies may have entered into contracts or agreements with third parties who have sourced the data that specifically determine whether datasets can be made available and who owns copyright on material created under the arrangement. These contracts or agreements must be checked for restrictions before releasing datasets.

Legal rights transferred

Where the legal rights have been transferred through the terms of a contract, the assignment of copyright must be in writing. This is essential as it removes ownership of copyright and enables agencies to own the data and openly licence and release it where appropriate. If you are unable to determine if rights have been transferred, particularly in the case of historical data, seek to obtain agreement of terms of use from the third party.

Terms of use

In some cases the third party is the copyright owner but provides a 'terms of use licence'. These licenses may have conditions on how the information can be used.

A term of use license may exist in the contract of agreement for use or issued under a commonly accepted open licence that is published with datasets or information. For example:

- [Attribution CC-BY licence](#)
- [Open Data Commons Open Data License \(ODbl\)](#).

Generally a Terms-of-Use licence will require attribution for the work to the third party (IP holder). If you reuse this work ensure this attribution is published with the work or derivative of the work.

How to attribute data from a third person

If you need to attribute works from a third party, you may need to check the attribution requirements they have requested. Standard attribution required:

- organisation name
- title of works
- URL link to published location
- date sourced.

Attribution for third party material that is reused must be published with the work.

Unknown third party rights

If the rights to reuse data provided by a third party are not clear you must obtain agreement of terms of use.

The preferred method to seek agreement is a standard letter with the following elements:

- details of the data in question
- request an agreement for terms of use under an CC-BY licence
- detail why you will release the data (e.g. Public Service)
- provide a reasonable period for the person to respond and a due date
- disclose that a non-response will be taken as acceptance of CC-BY licence
- provide appropriate means to respond and/or discuss.

Creation of new datasets

New intellectual property may be created if:

- third party data are manipulated
- third party data are merged with other data or additional data is collected
- different data structures are applied to the data.

Agencies should deal with third party intellectual property in a manner that avoids infringing the IP rights of others and complies with the law. Agencies may only consider this new data for open data if you have clear rights to create derivative works from the data and authority to use.

New contracts and agreements to acquire data rights

Agencies are encouraged to consider open data principles when establishing contracts and agreements for research or consultation reports.

Data is often required and publicly funded in order to develop these works. This Data is often considering a background IP. Where IP inputs (Background IP and/or Third Party IP) will be brought to a procurement agreement, the agreement must address South Australian Governments use of those inputs. The default position is that:

- the Background IP of each party remains the property of that party, along with any improvements to such Background IP
- The contractor must grant the agency a licence over the Background IP and any Third Party IP to the extent needed to allow the agency to enjoy the full benefit of the agreement.

While the licence over Project IP must be restricted in scope to the purposes of the procurement, agencies should carefully consider what these purposes are and what they may be in the future. Depending on the circumstances, the licence may need to allow:

- other agencies (not involved in the specific procurement) to use the IP
- the IP to be used and adapted for subsequent projects
- the IP to be adapted later by another service provider engaged by government
- the IP or background IP to be released publicly.

Grants for research and data collection should encourage the open release of the background IP data.

The commercial impacts of openly releasing purchased data from commercial entities should also be considered. Data that is marked as Sensitive: Commercial, is information, whose compromise could affect the competitive process and provide the opportunity for unfair advantage including:

- information concerning the trade secrets of any person
- information (other than trade secrets) that has a commercial value to any person
- likely to harm the business's commercial advantage in the marketplace.

Agencies are encouraged to see legal advice in these circumstances.

For more information on how to deal with copyright when establishing new contracts and agreements [View the Government of South Australia Intellectual Property Policy](#) (under review).

1.3.2 Terms and conditions of use specified when collecting data.

Data collected through applications and information systems may have terms and conditions which let the user know how the data they provide will be used. The restrictions detailed in the 'Use of Information Disclosure' may prevent you from releasing the data.

Agencies should review existing terms and conditions and seek Crown Solicitor Advice when amending or creating new terms and conditions.

This is particularly important with the increase to digital services offered to the public and businesses. Upfront disclosure will ensure users of services are aware of how data will be used and a statement, which indicates the highest standards of privacy, security, confidentiality and integrity with respect with the data we collect.

Establishing new collection Terms and Conditions

If establishing a new data collection service you should ensure that the Information Use Disclosure in your terms and conditions, detail the following:

- your agency and contact information is provided
- how he or she is able to gain access to the information
- the purpose for which information is collected
- any laws that require particular information to be collected
- any information collected that will be protected (e.g. personally identifiable information or commercially sensitive); and
- how any information collected may be disclosed to other parties (e.g. under a licence for reuse with exclusion to personally identifiable information).

A privacy statement and disclaimer are also required when creating any terms and conditions. It is recommended that you seek Crown Solicitor Advice before finalising any terms and conditions.

1.4 DATA VALUE

Agencies must consider the value of data if released, as this may help determine the priority or approach for releasing the data or help determine if it is able to be released. Valuable data when re-used contributes to economic, social, cultural and environmental growth, illustrates government's performance and contributes to greater government efficiencies through improved information sharing.

There are 3 types of value that must be considered before releasing data:

- value to the Government – what is it worth to us? How could the release of data create value for the government? What protection and due care is required to prevent damage to government administration or reputation if sensitive information was released inappropriately or accidentally?
- value to criminals and perpetrators of terrorism – could the data be used by criminals for illegal profit or intelligence for harm?
- value to the consumer - Data can be useful in unexpected and unpredictable ways, which means the potential usefulness of any data set should not be pre-judged.

It is also important to consider the aggregate value of information; datasets (small or large) may be independently harmless, however when aggregated with additional datasets from the same agency, or datasets from other agencies (state or commonwealth) as a whole the information may have an entirely different value.

The following characteristics should be considered when considering the value of a dataset for release:

1.4.1 Publicly Available Information or Provides a Public Service

Publicly released information in tables, spreadsheets or reports on websites should be released as open data. Agencies should consider this data as a priority for release as the public need is already identified and minimal effort may be required to classify, mark and transform the data and release it in an open format and open licence.

Information that provides a public service is valuable and should be given a high priority to release in an automated and open format (e.g. Rich Site Summary (RSS) feed) to enable a more informed community. This will disseminate your information to a wider audience with the potential for new information services to be delivered by the digital community.

1.4.2 Supports Strategic Goals of Government

Open data that could support or aligns with a strategic goal should be considered as a priority as this has value to the Government e.g.

- [South Australia's Strategic Plan](#)
- [A Stronger South Australia](#)
- [Modern Public Service](#)
- [SA Connected](#)
- Aligns with Agency strategic plans.

1.4.3 Stimulate Development, Industry or Services

Data that has the potential to spur economic growth, innovate or create new products or services is highly valuable. South Australian industry has the skills to transform data into new digital services. The entrepreneurial community is agile and can take advantage of emerging technologies and trends and as such are responsive to the needs of South Australians. Often these services enabled by data are not the core business of government; however the data is valuable asset that can support the new services.

1.4.4 Empower Business or Citizens to Make Informed Decisions

If data can empower business or citizens to make more informed decisions it should be perceived as high value. Data on housing, services, safety, crime, transport, travel, sports, recreation and events could assist citizens make more informed decisions about their lives.

Data used to make industry, infrastructure or planning related policy is also likely to help business make decisions. Improved business decisions can lead to economic growth in our state.

1.4.5 Research or Industry Efficiencies

Statistical information that could assist researchers should be released as a high priority. Examples of research study fields include environment, agriculture, humanities, social sciences and health.

By releasing research data openly we allow others to build on data to further advance analysis and research without the need to fund duplicate data collection. Open release of research data can facilitate debate and communication of ideas to a broader audience.

Open data can reduce industry red tape and provide efficiencies through direct access to data collected from the source and reducing the need for agencies to collect or estimate value information required to make business decisions.

1.4.6 Comparable Across-Jurisdictional Open Data

Data that can be compared with other jurisdictions is beneficial for trend analysis and can improve business decision making. The business viability of a data enabled application (app) is also increased if the market for the app is national or multi-jurisdictional. Agencies can check what comparable data has been released by visiting the other Australian Government Open Data Portals:

- [Australian Government data portal data.gov.au](#)
- [Queensland Government data portal data.qld.gov.au](#)
- [Queensland Government spatial data globe](#)
- [New South Wales Government data portal](#)
- [New South Wales Government spatial data portal](#)
- [Australian Capital Territory Government data portal](#)

- [Victorian Government data portal](#)

1.4.7 Government Sharing Efficiencies and Policy Development

Data that is often shared between government agencies or reported to other jurisdictions can create efficiencies if the data is released openly. Often data is reported to several different parties, at different times and in different formats. Openly releasing this data allows the users to access the required information when they require it and transforms the data into a format that suits their needs. Value to the Government can be achieved through reduced resources to process Freedom of information request. Open data can lead to improved public sector productivity or lower service delivery cost.

Open data can lead to improved data to decision policy making across the Government through improved access and currency of data required to make informed decisions.

Providing data openly to support the rationale for policy particularly through community engagement processes can improve the development and implementation of policies.

IDENTIFY SUMMARY

Identify Data		
<input type="checkbox"/>	Public Facing Mediums	Review existing public facing data via agency websites, notifications & publications.
<input type="checkbox"/>	Inventories	Review information assets and data inventories for possible datasets.
<input type="checkbox"/>	Proactive Disclosure	Consider releasing data from frequent Freedom of Information Requests or regular across jurisdictional reporting.
<input type="checkbox"/>	Business Planning	Identify opportunities for obtaining data in the future and obtaining intellectual property rights.
<input type="checkbox"/>	Community Demand	Engage with the community and respond to requests for data.
Do Third Party Rights Exist?		
<input type="checkbox"/>	Intellectual Property	Who has intellectual property rights of the data?.
<input type="checkbox"/>	Copyright	Does a third party hold copyrights of the data?
<input type="checkbox"/>	<i>If Yes, seek to Obtain Rights or else dataset is not eligible for release</i>	
Are there Terms and Conditions?		
<input type="checkbox"/>	Use of Information Disclosure	Are there terms and conditions which prevent data collected from being released.
Value		
<input type="checkbox"/>	Value	Consider the value of the data to the Government, Criminals and Public.
<input type="checkbox"/>	Public Value	Consider whether a potential dataset will meet strategic goals, improve services, stimulate and empower business and industry, create efficiencies and improve public policy and enable more research or across jurisdictional comparisons.
Commence the Open Data Process Worksheet		
<input type="checkbox"/>	Start a new worksheet for each dataset.	Record all decisions made about the dataset on the worksheet.
<input type="checkbox"/>	Data Authority	Identify the Data Authority (usually a Director or Executive) who will ultimately approve the dataset release. Datasets may have a different data authority.