

2 CLASSIFY - STEP 2 OPEN DATA PROCESS

Data can only be considered for release under open data after being assessed in terms of information confidentiality, integrity and availability classifications and applicable markings have been applied. Where data meets the requirements of open data, a licence classification using the AusGOAL licence framework is also required to ensure users know how they can use the data. The below discusses these requirements:

2.1 INFORMATION SECURITY CLASSIFICATION

Agencies must assess data to determine if it could be released publicly, based on their information classification procedures as required by the South Australian Government Information Security Management Framework (ISMF). Visit the [Department of Premier and Cabinet website](#) to view the ISMF policies, standards and guidelines. The Information Privacy Principles Instruction (IPPI) also guides information privacy in the Government of South Australia. [View the IPPI](#) (PDF 241Kb).

Data can only be marked as Public if the release of this information does not cause any damage to the state, the government, an agency, commercial entities or members of the public.

Only data that has been assessed and marked as Public can be released as open data. Data that has a Protective Marking or Dissemination Limiting Marker (refer ISMF), for example **Secret, For Official Use Only** or **Sensitive**, cannot be released as-is, however may be re-classified if appropriate de-classification and/or risk mitigation activities are undertaken.

2.1.1 Governance and Accountability

Chief Executives

Agency Chief Executives are accountable for all security matters within their agency. The Chief Executive must authorise the disclosure of all official information (including data and datasets) to the public.

Data Authority

The Data Authority is responsible for ensuring that classification or reclassification of the data is undertaken and that appropriate input is obtained from the data subject experts and information users.

Executive Peer Review

Executive Peer Review provides risk mitigation to ensure the public release of a dataset does not inadvertently put another part of the business security at risk or disclose information that could lead to identification of a person (mosaic effect).

Agencies are encouraged to conduct an Executive Peer Review of all dataset security markings before a dataset that has been assessed and marked as Public, is released. Executive peer review is recommended if data has been manipulated to mitigate risks e.g. when personal information has been de-identified or when information is declassified.

The recommended approach for an Executive Peer Review is to engage your Information Technology Security Adviser or Advocate to circulate open data candidates to all Executives within an agency for review and comment. A sample of the data should be provided and summary of how any data risks that have been mitigated e.g. de-identification or redaction techniques applied.

Agency Information Technology Security Adviser (ITSA)

Agencies should consult with their Information Technology Security Adviser (ITSA) for further advice and guidance on agency specific information classification and marking

procedures and guidelines. Agency ITSAs can assist with data assessment decisions and can provide advice on information security threats and risks.

2.1.2 Security classification and marking decisions

Information security is a complicated area and the Agency ITSA is best suited to both advise on how to properly assess information (including data) in terms of confidentiality, integrity and availability; and how to handle that information once it's been classified and marked in accordance with the ISMF and agency specific policies, standards and guidelines.

For Data to be released through open data, it must firstly be assessed and able to be marked as **Public** then authorised for release by an Agency Chief Executive.

Data marked as **Public** is authorised for unlimited public access and circulation such as agency publications, data download sites and websites.

Government information, including data, is required to be assessed and marked. Where government data is not able to be marked Public, some of this information can have the sensitive aspects redacted, amended or otherwise modified such that it could be re-assessed as meeting the requirements for Public and if still valuable to the public, released under open data.

The Data Authority (typically the business owner of the information) is responsible for re-assessment of information, where appropriate, and should engage their ITSA to assist with these decisions.

Refer to the [Data Security Marking Decision Diagram](#) (Appendix A) and [Open Data Guide to Security Classification](#) (Appendix B) for assistance with security decisions and actions required.

For a full list of the classification markings refer to [ISMF Guideline 8b](#) (PDF).

2.1.3 Privacy of personally identifiable information

In making government information publicly available, agencies must ensure that personally identifiable information regarding citizens is not released in accordance with the [Information Privacy Principles Instruction \(IPPI\)](#). This includes checking metadata for personally identifiable information.

Agencies may need to de-identify personally identifiable information, which means removing anything that can identify a person such as a person's name, address, gender, date of birth, ethnicity etc. Care must be taken to ensure information is properly de-identified and not able to be re-identified by linking with other information sources. Once the data is de-identified it needs to be reclassified.

The security classification assessment and marking decisions detailed above incorporate markings for personally identifiable information.

A number of techniques can be applied to properly de-identify the data and mitigate any risks of identification. Refer to the Privacy Committee of South Australia [Privacy and Open Data Guideline \(PDF\)](#) for more information.

This guideline also provides a [Privacy Risk Assessment Process \(Refer appendix C\)](#).

Executive Peer Review is recommended before release of privacy risk mitigated data. A sample data should be produced and peer reviewed as a means of testing and a summary of how any data risks have been mitigated e.g. de-identification techniques applied.

2.2 RESPONSIBLE INFORMATION SHARING

This process recognises that public information may require high degrees of integrity (accuracy) and availability (and by association that availability requirements may change based on calendar or event driven periods, a notion of 'peak-demand' for certain types of information). The benefits of responsible information sharing include:

- User experience both within and external to government
 - Information is easy to find and can be relied upon
 - Services are interactive and timely
- Elimination of delays
 - Information on demand (i.e. readily shared and accessible when required)
 - Accessible by leveraging the internet, including mobile devices and emerging technologies
- Reduction in costs and increase in organisational efficiencies
 - Lowers cost of service delivery
 - Fosters greater use of 'self-service' capabilities
 - Interactive government is an agile and consultative government
- Accurate and timely information
 - Timely information is achieved because availability requirements have been considered
 - Accurate information is more likely because integrity requirements have been considered
- Maintaining trust and confidence in government as a supplier and custodian of information
 - Reliable and secure information services engender trust and confidence. This is reflected by organisational capability, capacity and communication.

Sharing information with the public may require an adequate level of assurance to business users on the accuracy and availability of the data as per [ISMF Guideline 8a](#).

Some examples of the varying degrees of responsible information sharing are illustrated below:

- Emergency management and crisis response information is generally distributed on a broad public scale, but requiring exceptionally high degrees of accuracy (integrity) and availability in order to inform the community and emergency services personnel in a timely and accurate manner
- Information Sharing Guidelines for Promoting the Safety and Wellbeing of Children, Young People and their Families - these guidelines deal with the legal and practical framework that supports appropriate information sharing for the provision of integrated support to children, young people and their families
- Data delivery through a web service for use in a developers business or application.

2.2.1 Availability Classification

Use the table below to determine how available data needs to be when made publicly accessible. Consult with your agency ITSA for advice to ensure that the data is not over classified and to advise of any additional controls required. A high level of classification (e.g. A4) will increase the costs to manage the data both internally and for public release.

Availability classifications should be discussed with you Data Manager to assist decisions with the approach for release of the data.



Availability Classification

- Information is bound to become unavailable at certain points in time. Whether the information is unavailable due to system outages or planned maintenance windows or as a result of unintended and unplanned events. The determination of a 'tolerable outage' must be accepted by the business.
- Tolerable outages can drive an Availability Classification for the information. Factors to consider include the dependence the business, its customers or the community has on the information and the business impacts a disruption or loss of access to the information may cause. Availability requirements may change at certain times whether event or calendar driven.

Classification	Description
A4	ABSOLUTE requirement, meaning that the business would be crippled by the loss and recovery must be virtually instantaneous (no longer than a few minutes).
A3	HIGH requirement, meaning that loss would cause major disruption to the business and recovery must be achieved within a period measured in hours (typically same business day).
A2	MODERATE requirement, implying the loss would have a significant impact and recovery must be achieved within a period measured in days (typically three business days or less).
A1	LOW requirement, meaning that loss of the data would have only a minor impact on the business for an extended period (i.e. "best-effort" recovery).

2.2.2 Integrity Classification (Quality)

The integrity classification will assist you to determine what level of quality the data needs to be and is generally applied when misinformation could cause risk to a person or risk to the reputation of the government.

The quality of each datasets that will be released to the public may be considered when assessing integrity classification of the dataset. The below discuss elements of quality as per guidance provided by the National Statistical Service (NSS):

High Quality

- Timely, low errors, consistent collection, interpretable, authoritative source

Reasonable Quality

- Low unknown errors, consistent collection, interpretable, source clearly identified.

Poor Quality

- High errors, inconsistent collection, interpretability low and metadata required to interpret the data is not available.

Data that is of poor quality should not be publicly released until quality issues are resolved. If the data is considered to have a high public value, the Data Authority may need to investigate ways to improve the quality of the data in future plans.

2.2.2.1 Characteristics of quality data:

- Timeliness

If there are lengthy delays between the reference period and data availability this can have implications for the currency or reliability of the data. However, historical data is still very valuable to publish, so timeliness in itself would not disqualify data from release.

- Accuracy

Accuracy is an important component of quality as it relates to how well the data portrays reality, which has clear implications for how useful and meaningful the data will be for interpretation or further analysis. The major sources of errors that could cause inaccuracies should be assessed. Consider also the Integrity Classification the information has been marked with.

- Coherence

Coherence is an important component of quality as it provides an indication of whether the dataset can be usefully compared with other sources to enable data compilation and comparison.

- Interpretability

Interpretability is an important component of quality as it enables the information to be understood and utilised appropriately. Check the data to ensure it would have meaning to an external party. Are terms used in the dataset ambiguous, open to interpretation or likely to confuse a user? If so, detailed interpretation metadata with definitions or explanatory notes will need to be provided to ensure the data is understood as intended.

- Authoritative source

Data is more valuable if it is collected at the source or has systems in place to ensure it is the single point of truth or authoritative source of information.

2.2.2.2 Quality Assessment Tool

The ABS through its National Statistical Service (NSS) has developed a free online data quality tool that will help get agencies thinking about what makes data good quality. The Data Quality Statement online tool can be found at <https://www.nss.gov.au/dataquality/>

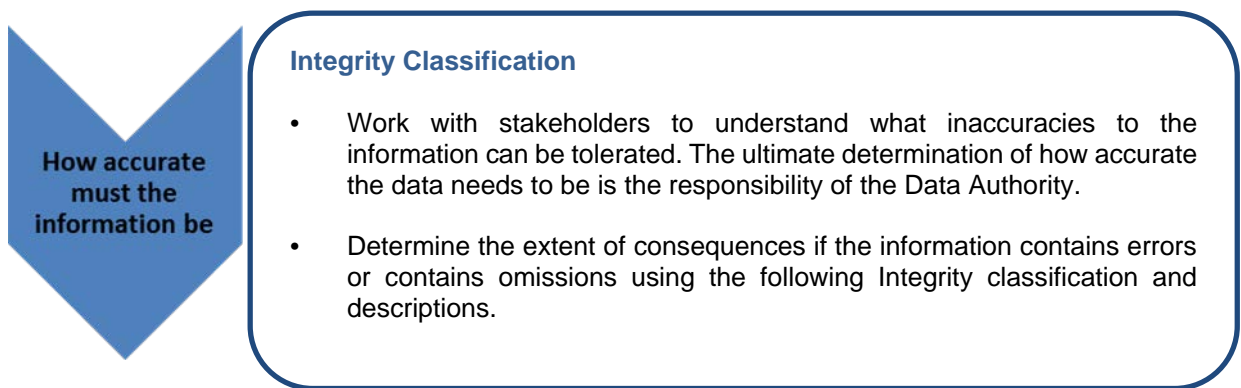
The tool takes you step by step through a series of 7 questions about your dataset.

The output is a data quality statement that can be saved as XML or RFT file and published with the dataset on Data.SA. The template can be saved in progress and revisited when the answers to the questions are available. This is an optional step for agencies, but is a good guide for data quality at a national standard level.

Ensure that personal contact information is not provided with any published version of the statement. If you will release the statement with your data ensure that you include it as an attachment for approval with the dataset.

2.2.2.3 Determine the Integrity Classification

Use the table below to determine the integrity classification of the data. Consult with your agency ITSA for advice to ensure that the data is not over classified and to advise of any additional controls required. A high level of classification (e.g. I4) will increase the costs to manage the data both internally and for public release.



How accurate must the information be

Integrity Classification

- Work with stakeholders to understand what inaccuracies to the information can be tolerated. The ultimate determination of how accurate the data needs to be is the responsibility of the Data Authority.
- Determine the extent of consequences if the information contains errors or contains omissions using the following Integrity classification and descriptions.

Classification	Description
I4	ABSOLUTE requirement, implying that no inaccuracies or omissions can be tolerated
I3	HIGH requirement, meaning that a loss of integrity would cause significant embarrassment and disruption and might be difficult to detect.
I2	MODERATE requirement, meaning that a user would be somewhat affected by a loss of integrity, but the situation could be easily detected and recovered.
I1	LOW requirement, such that there would be minimal impact if the data was inaccurate or incomplete

2.3 LICENCE CLASSIFICATION

The Government of South Australia supports and encourages the dissemination and exchange of public sector information, and endorses the use of the Australian Governments Open Access and Licensing Framework (AusGOAL) by its agencies.

Restrictions on the use of the data will be avoided where possible. Utilising the [Australian Governments Open Access and Licensing Framework](#) (AusGOAL) will support openness and improve usage.

A licence classification is only required for data that has been classified as Public and is intended for release as open data..

2.3.1 *What is AusGOAL*

AusGOAL is the [Australian Governments Open Access and Licensing Framework](#) which provides support and guidance to government and related sectors to facilitate open access to publicly funded information. AusGOAL makes it possible for organisations to manage their risks when publishing information and data in a way that drives innovation and entrepreneurial activities. The AusGOAL licensing framework exists in parallel with our copyrights, which protects our intellectual property, Government of South Australia brand and logo's.






2.3.2 *Creative Commons*

AusGOAL incorporates a suite of licences including the [Creative Commons Version 4.0 International](#). The Government of South Australia also recognises work licensed under the [Creative Commons Version 3.0 Australian licenses](#), which is an older version of the licence used in Australia only.

Key attributes of the Creative Commons V4.0 licences:

- internationally recognised
- defines how data can be used, shared or adapted
- provides clear instructions on the restrictions of using the data
- requires attribution to your Agency, title of work, date sourced and the URL used to access the data
- does not allow data users to make agencies endorse them or their data use
- requires users of the data to indicate if they modified the material
- notifies the user that no warranties are given
- material is offered as-is and as-available and makes no representations or warranties of any kind concerning the licensed material, whether express, implied, statutory or other
- limits the Government of South Australia's liability
- anonymity – can specify no attribution for derivatives
- 30 day period to cure breach before licence terminates.

2.3.3 The Creative Commons Licenses

	<p>Attribution CC BY</p> <p>This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.</p> <p>View License Deed View Legal Code</p>
	<p>Attribution-ShareAlike CC BY-SA</p> <p>This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to “copy left” free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.</p> <p>View License Deed View Legal Code</p>
	<p>Attribution-NonCommercial CC BY-NC</p> <p>This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.</p> <p>View License Deed View Legal Code</p>
	<p>Attribution-NonCommercial-ShareAlike CC BY-NC-SA</p> <p>This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must acknowledge you and be non-commercial, they do not have to license their derivative works on the same terms.</p> <p>View License Deed View Legal Code</p>
	<p>Attribution-NonCommercial-NoDerivs CC BY-NC-ND</p> <p>This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they cannot change them in any way or use them commercially.</p> <p>View License Deed View Legal Code</p>

2.3.4 *The Preferred Licence*

The preferred licence for data is [Creative Commons Attribution 4.0 Licence](#). This licence is the most open with least restrictions. It supports research, analysis, data transformation and enables applications to be created from it. This licence is also recognised internationally.



Attribution CC BY

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

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

This licence must be used unless there is a reason a more restrictive licence should be provided such as:

- data was provided with a term of use
- there is a reason why you will not allow someone to re-use the material for commercial gain
- there is a reason you won't allow a user to adapt the work you require re-licence under identical terms.

2.3.5 *Alterations and additions to a Creative Commons licence*

- **Can I waive license terms or conditions?**

Yes. You may choose to waive some license terms or conditions. Generally a request to waive any conditions will need to be reviewed on a case by case basis.

- **Can I change the license terms or conditions of a CC Licence?**

No if you change the terms and conditions of any Creative Commons license, you must no longer call, label, or describe the license as a “Creative Commons” or “CC” license, nor can you use the Creative Commons logos, buttons, or other trademarks in connection with the modified license or your materials.

Altering terms and conditions is distinct from waiving existing conditions or granting additional permissions than those in the licenses.

If different licence terms are required a separate licence term agreement is required. You will require Crown Solicitor advice to develop new terms of licence. These agreements are discouraged as it creates confusion in expectations and creates additional burden of negotiating permissions on a case by case basis.

If the preferred licence for data ([Creative Commons Attribution 4.0 Licence](#)) does not meet your needs you may require a more restrictive licence.

2.3.6 *When to use a more restrictive licence*

- **Data was provided with a term of use**

When identifying third party property rights you may have identified data that was provided on a set term of use. Often these terms of use will define how you can re-license work. Check the term of use before selecting a licence.

- **There is a reason you won't allow someone to re-use the material for commercial gain**

Restrictions on data for commercial re-use should be avoided as open data could stimulate economic growth in South Australia through using the data to support a commercially viable business. If a commercial licence restriction is in place on the data, then the user cannot even recover the costs of using data.

- **There is a reason you won't allow a user to adapt the data (Derivative)**

Restrictions on the data to adapt the work should be avoided, as this will limit the user's ability to display the data in a different form, innovate, analyse, and create services. Note that all Creative Commons Licenses require the user to indicate if they modified the material and retain an indication of previous modifications.

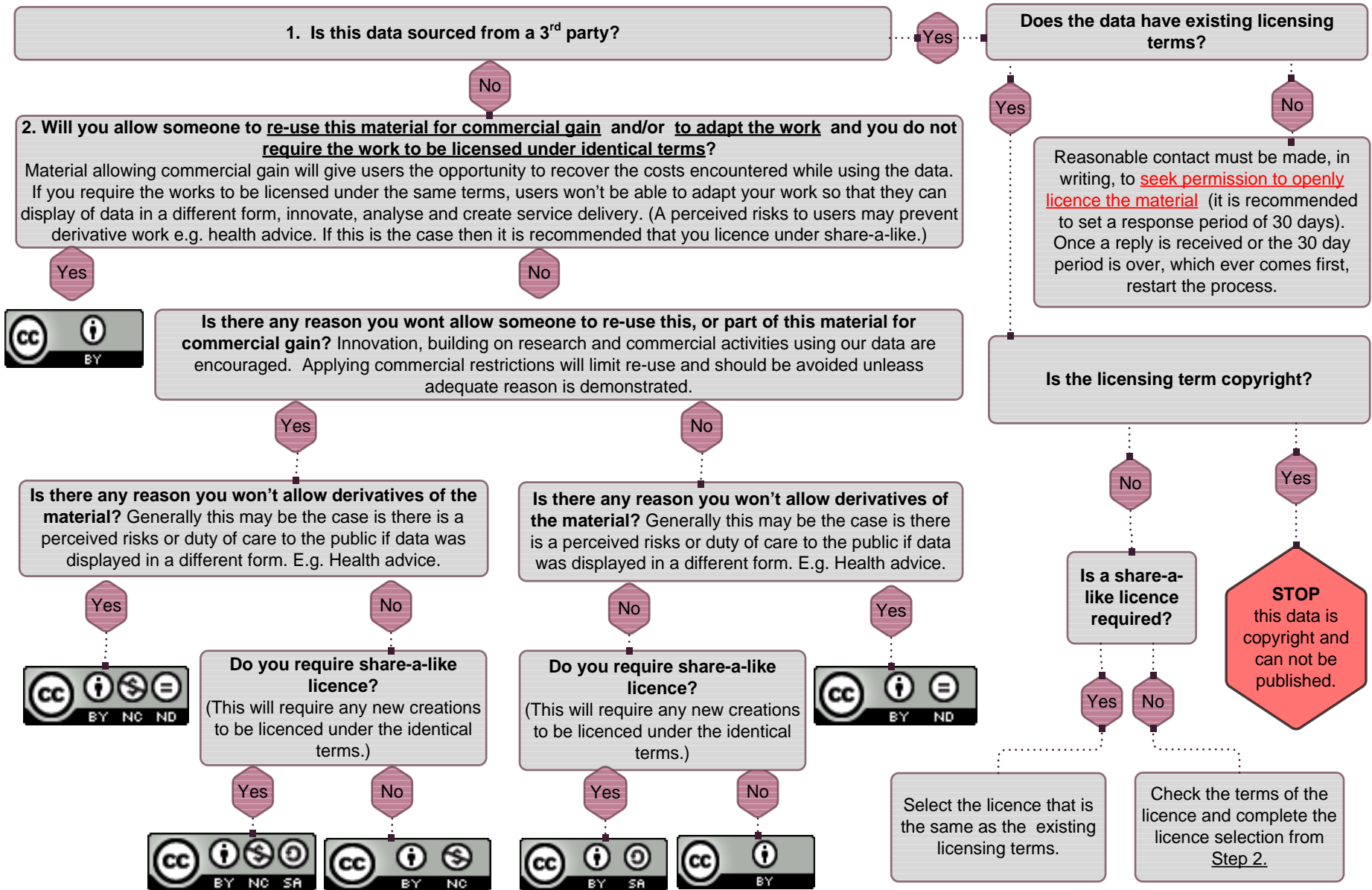
The Data Authority should consider if there are any reasons that derivative work should not be allowed due to a perceived risk or duty of care to citizens if data was used in a way that could cause harm. For example, health advice should not be licensed for derivative reuse, however non-personal health statistics may not have a perceived risk if displayed in a different form or used to advance research. Personal information should never be licensed for derivative use, even if there is a legal requirement to publish information.

- **If you require Re-licence under identical terms (Sharealike)**

If you require users of the data to share the data under an identical licence you can choose a ShareAlike licence. This licence restricts the use of data and requires users of the work to licence any derivative works under the same terms.

If a more restrictive licence is required, use the following [Open Data Licence Decision Diagram](#)

2.3.7 Open Data Licence Decisions Diagram



CLASSIFY SUMMARY

Determine Data Security Marking		
<input type="checkbox"/>	Public	A dataset with a marking of Public can be released as open data.
<input type="checkbox"/>	Reclassification	Some data can be reclassified as Public . The data will require some redactions, amendments or manipulation to ensure sensitive elements are removed so that there is no damage (or potential damage) to the government, business or members of the public. The Data Authority is responsible for reclassification of information and should engage their ITSA to assist with these decisions. An Executive Peer Review is also encouraged.
Determine Responsible Information Sharing		
<i>Availability classification</i>		
<input type="checkbox"/>	A4	ABSOLUTE requirement
<input type="checkbox"/>	A3	HIGH requirement
<input type="checkbox"/>	A2	MODERATE requirement
<input type="checkbox"/>	A1	LOW requirement
<i>Integrity Classification</i>		
<input type="checkbox"/>	Quality	Consider the quality of the data to assist the integrity classification
<input type="checkbox"/>	I4	ABSOLUTE requirement
<input type="checkbox"/>	I3	HIGH requirement
<input type="checkbox"/>	I2	MODERATE requirement
<input type="checkbox"/>	I1	LOW requirement
Determine the Creative Commons Licence		
<input type="checkbox"/>	CC BY	The preferred and most open licence
<input type="checkbox"/>	CC BY-SA	ShareAlike
<input type="checkbox"/>	CC BY-ND	No Derivs
<input type="checkbox"/>	CC BY-NC	Non-Commercial
<input type="checkbox"/>	CC BY-NC-SA	Non-Commercial-ShareAlike
<input type="checkbox"/>	CC BY-NC-ND	Non-Commercial-NoDerivs
Complete Open Data Process Worksheet		
<input type="checkbox"/>	Record decisions	Update the Open Data Process Worksheet.

