



THE UNIVERSITY OF
MELBOURNE

Shifting sensitive data management practices: A delicate balance

Dr Kimberley D'Costa

Research Data Management Program, Petascale Campus
Chancellery (Research and Enterprise)

kim.dcosta@unimelb.edu.au



Management of Sensitive Research Data at UoM

- Sensitive research data has the potential to cause significant harm to participants, researchers and the University if disclosed or accessed by unauthorised parties
- The RDM environment was complex and difficult for researchers to navigate
- Lack of a central position on how sensitive research data should be managed
- Inconsistent RDM practices were commonplace and the University's risk exposure was high



Varying disciplinary norms and research practices



Multiple state, national and international regulations; as well as cybersecurity and technical standards



Multiple teams across the University delivering RDM infrastructure or services



Multiple RDM tools and systems in common usage at UoM



Project Goals

- Help researchers and the University consistently assess the sensitivity of research data
- Empower researchers to manage research data more safely and securely by providing simple, actionable supporting guidance
- Guide researchers to University-provided infrastructure that is appropriate for the collection, storage, analysis and retention of their research data
- Foster effective and compliant research across all disciplines
- Reduce the overall risk for researchers and the University

What is sensitive data?

No globally accepted definition!

Research data that has the potential to cause significant harm to participants, communities, researchers, the University or other organisations when disclosed to or accessed by unintended parties, either accidentally (e.g., through mismanagement) or by malice (e.g., through Cybersecurity attacks)





Research Data Classification Framework

- Published in March 2022
- Classification ranges from Green (least sensitive) to Red (most sensitive)
- Increasing sensitivity relates to the severity of harm to participants, communities, researchers or the University, when data is accessed by unintended parties
- The generic titles were chosen to avoid confusion and eliminate any pre-conceived notions

Data Classification	Description
Green	<p>The green classification applies to information that presents the risk of negligible material, social, reputational, legal, or other harm to individuals, communities, or organisations in the event of accidental or unauthorised access, disclosure, alteration, misuse, loss, or destruction.</p> <p>Examples:</p> <ul style="list-style-type: none">• Information intended for public disclosure/consumption• Published manuscripts or datasets• Data from public websites or social media that does not relate to an identified or identifiable individual
Yellow	<p>The yellow classification applies to information that presents the risk of only limited material, social, reputational, legal, or other harm to individuals, communities, or organisations in the event of accidental or unauthorised access, disclosure, alteration, misuse, loss, or destruction.</p> <p>Examples:</p> <ul style="list-style-type: none">• De-identified or aggregated data that does not relate to an identifiable individual or present any risk of significant harm to a community or group• Unpublished research data and outputs that do not fall into all other categories, e.g., drafts of research publications, novel creative works etc.• Novel analyses or transformations of publicly available data or information• Data generated by instruments, imaging systems or computational models that are not linked to a specific identifiable entity
Orange	<p>The orange classification applies to information that presents the risk of significant material, social, reputational, legal, or other harm to individuals, communities, or organisations in the event of accidental or unauthorised access, disclosure, alteration, misuse, loss, or destruction.</p> <p>Examples:</p> <ul style="list-style-type: none">• Personally identifiable data including name, contact details, financial details, individual medical records, etc.• Genetic or biometric information• Re-identifiable data i.e., when the identity of a specific individual or other sensitive entity can be reasonably ascertained by data linkage or other activities• Culturally and ecologically sensitive data
Red	<p>The red classification applies to information that presents the risk of severe social, psychological, reputational, financial, legal, or other harm to individuals, communities, or organisations in the event of accidental or unauthorised access, disclosure, alteration, misuse, loss, or destruction.</p> <p>Examples:</p> <ul style="list-style-type: none">• Personally identifiable data containing “sensitive” information as defined by Victorian Privacy legislation• Assets and information for defence research, or that have the potential to be adapted for military or ‘dual use’ applications• Data involving Sensitive Security Biological Agents (SSBAs)



Research Data Classification Framework

Supports compliance with data management components of:


- Privacy Act, 1988 (Cth) and Privacy and Data Protection Act, 2014 (Vic)
- Health Records Act, 2001 (Vic)
- Defence Trade Controls Act, 2012
- Environment Protection and Biodiversity Conservation Act, 1999
- National Health Security Act, 2007, National Health Security Regulations, 2018 and SSBA standards
- AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research, 2020
- European Union General Data Protection Regulation 2016/679 (GDPR), as the best practice standard in privacy governance

Data Classification	Description
Green	<p>The green classification applies to information that presents the risk of negligible material, social, reputational, legal, or other harm to individuals, communities, or organisations in the event of accidental or unauthorised access, disclosure, alteration, misuse, loss, or destruction.</p> <p>Examples:</p> <ul style="list-style-type: none">• Information intended for public disclosure/consumption• Published manuscripts or datasets• Data from public websites or social media that does not relate to an identified or identifiable individual
Yellow	<p>The yellow classification applies to information that presents the risk of only limited material, social, reputational, legal, or other harm to individuals, communities, or organisations in the event of accidental or unauthorised access, disclosure, alteration, misuse, loss, or destruction.</p> <p>Examples:</p> <ul style="list-style-type: none">• De-identified or aggregated data that does not relate to an identifiable individual or present any risk of significant harm to a community or group• Unpublished research data and outputs that do not fall into all other categories, e.g., drafts of research publications, novel creative works etc.• Novel analyses or transformations of publicly available data or information• Data generated by instruments, imaging systems or computational models that are not linked to a specific identifiable entity
Orange	<p>The orange classification applies to information that presents the risk of significant material, social, reputational, legal, or other harm to individuals, communities, or organisations in the event of accidental or unauthorised access, disclosure, alteration, misuse, loss, or destruction.</p> <p>Examples:</p> <ul style="list-style-type: none">• Personally identifiable data including name, contact details, financial details, individual medical records, etc.• Genetic or biometric information• Re-identifiable data i.e., when the identity of a specific individual or other sensitive entity can be reasonably ascertained by data linkage or other activities• Culturally and ecologically sensitive data
Red	<p>The red classification applies to information that presents the risk of severe social, psychological, reputational, financial, legal, or other harm to individuals, communities, or organisations in the event of accidental or unauthorised access, disclosure, alteration, misuse, loss, or destruction.</p> <p>Examples:</p> <ul style="list-style-type: none">• Personally identifiable data containing “sensitive” information as defined by Victorian Privacy legislation• Assets and information for defence research, or that have the potential to be adapted for military or ‘dual use’ applications• Data involving Sensitive Security Biological Agents (SSBAs)

Research Data Classification Tool

- A tool to support researchers to determine the level of sensitivity and classify their data into one of the four classification levels
- No identifiable details are captured
- Results are presented as guidance only, and assessments are not currently reviewed or governed
- Built in REDCap, as this meets required security, access and distribution standards, as well as allows for the incorporation of sophisticated branching logic

AAA

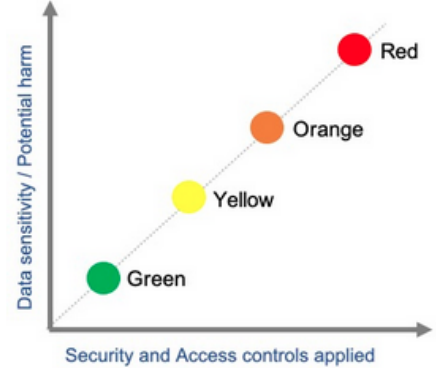


Research Data Classification Tool

The [Research Data Classification Tool](#) is part of the University of Melbourne's [Research Data Classification Framework](#). Using this tool will help you assess the sensitivity of your research data and provide you with guidance on how to manage your research data safely and securely.

After answering a series of multiple-choice questions about your research data or project, your data will be assessed as one of four classifications ranging from 'Green' (least sensitive) to 'Red' (most sensitive).

Each category is defined by the severity of harm that mismanagement or unintended disclosure of data could cause to research participants, researchers, or the University. As sensitivity increases, the security and access controls that should be applied to the research data also increase.



The assessment tool will take approximately 15 minutes to complete, and all responses are anonymous. It is intended as an information resource. You will be able to save a PDF copy of your responses and assessment after clicking "Submit" at the end of the survey.

If you require more specific advice or wish to provide feedback related to the Research Data Classification Framework or this Assessment Tool, please contact the RDM Program team (rdm-program@unimelb.edu.au).

Is your research data already published or publicly accessible*?

* This includes data from public data repositories, websites, social media sites, or open access journals

Yes

No

Does your data fall into any of the following categories?

* Personal information refers to data regarding individuals. This includes personal demographic details, images, audio and video recordings, spatial data of individuals, identifiable survey or interview data, medical records, health or genetic information, personal opinions or information posted on social media etc.

** Dual Use refers to information or technology that is intended for civilian or commercial use, but that could also be used to serve a military purpose. See the [Australian Government - Department of Defence website](#) for further information.

*** For a list of current SSBA's, refer to the [Australian Government](#)

☒
Personal information*
(including health or clinical information)

☐
Ecological information about species or habitats

☐
Assets and information for defence research, or that have any potential military and/or 'dual use' applications**

reset



Guidance provided

- Two new resources have been produced for researchers:
 1. General data handling guidelines recommended for each classification level
 2. Appropriate University-provided or commonly used research data systems/tools for each classification level
- Researchers are also directed to additional support available at the University
- Supplemented by a series of Frequently Asked Questions about the Framework and management of sensitive research data

Research Data Handling Guidelines

Simple practices can help safeguard your data against disclosure or access by unauthorised parties. After assessing the classification of your research data using the [Research Data Classification Tool](#), the following guidelines will help you identify ways to adequately protect your data.

These general Research Data Handling Guidelines should be used together with research data systems that have appropriate technical, security and operational features in place. See the [Recommended Systems for Sensitive Data](#) page for more information.

Recommended for all research data

Recommended for 'Yellow' classification

Recommended for 'Orange' classification

Sharing data and managing access

- Data must not be transmitted through transfer platforms or by providing direct access (e.g., FileSender on Cloudstor).
- Only share data with specific individuals and record of the access provided.
- Ensure that access is only provided to the entire research dataset and only to those who need it.

Recommended Systems for Sensitive Data

Research systems have a combination of technical, security and operational features in place that affect their eligibility to manage sensitive research data. Selecting an appropriate research data system can help you meet recommended security standards more readily.

The University provides a range of [research data systems](#) that have been assessed against best-practice security controls to create the below recommendations.

The assessments represented below are focused on current system capabilities and configurations, and only hold true while data is maintained within the system (e.g., when data stored in OneDrive are synced locally, it is being transferred to a less secure location).

These recommendations should be considered with the [Research Data Handling Guidelines](#), as you must also maintain suitable general data handling practices when using endorsed systems. Systems specific guidance is currently being developed and will be available in July 2022.

University-provided systems

Research Systems	Data Classification Levels			
	Green	Yellow	Orange	Red
<i>General data storage and management systems</i>				
Microsoft OneDrive	✓	✓	✓	✗
Microsoft SharePoint	✓	✓	✓	✗
Network File Share (Research – NAS)	✓	✓	✓	✗



Guidance provided

- **Research Data Handling Guidelines:**
 - Practical steps for the appropriate management of data
 - Minimises risk and likelihood of disclosure to unintended parties
- **Recommended systems informed by an independent, third-party review:**
 - Determine appropriate security controls in alignment with National Institute of Standards and Technology (NIST) Cybersecurity Framework
 - Assess common research data systems against these controls

Research Data Handling Guidelines

Simple practices can help safeguard your data against disclosure or access by unauthorised parties. After assessing the classification of your research data using the [Research Data Classification Tool](#), the following guidelines will help you identify ways to adequately protect your data.

These general Research Data Handling Guidelines should be used together with research data systems that have appropriate technical, security and operational features in place. See the [Recommended Systems for Sensitive Data](#) page for more information.

Recommended for all research data

Recommended for 'Yellow' classification

Recommended for 'Orange' classification

Sharing data and managing access

- Data must not be transmitted through email or file transfer platforms or by providing direct access (e.g., FileSender on Cloudstor).
- Only share data with specific individuals and maintain a record of the access provided.
- Ensure that access is only provided to the entire research dataset and only to those who need it.

Recommended Systems for Sensitive Data

Research systems have a combination of technical, security and operational features in place that affect their eligibility to manage sensitive research data. Selecting an appropriate research data system can help you meet recommended security standards more readily.

The University provides a range of [research data systems](#) that have been assessed against best-practice security controls to create the below recommendations.

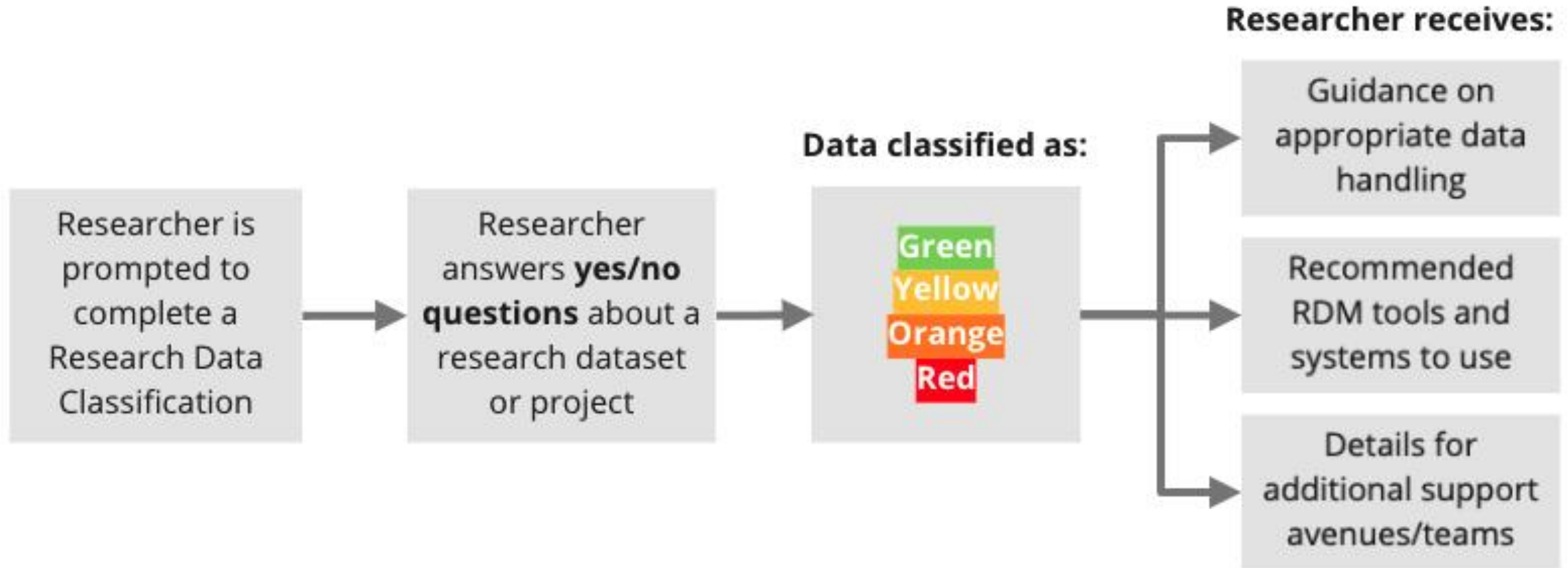
The assessments represented below are focused on current system capabilities and configurations, and only hold true while data is maintained within the system (e.g., when data stored in OneDrive are synced locally, it is being transferred to a less secure location).

These recommendations should be considered with the [Research Data Handling Guidelines](#), as you must also maintain suitable general data handling practices when using endorsed systems. Systems specific guidance is currently being developed and will be available in July 2022.

University-provided systems

Research Systems	Data Classification Levels			
	Green	Yellow	Orange	Red
<i>General data storage and management systems</i>				
Microsoft OneDrive	✓	✓	✓	✗
Microsoft SharePoint	✓	✓	✓	✗
Network File Share (Research – NAS)	✓	✓	✓	✗

Classification Process



Reflections on a Principles-based Approach



Academic-led

- Convened an Advisory Committee and Working Group consisting of both academic and professional experts
- Included representation from a wide range of research disciplines

Researcher-focused

- Not simply a compliance exercise
- Limit additional burden – Add value to the research process and experience
- **Increased coherence in the RDM ecosystem**

Reflections on a Principles-based Approach



Centrally-endorsed position

- Referenced in the University's Research Data Management Policy (MPF1242)
- Ensured buy in from leadership very early on and sought endorsement from the Research Ethics and Integrity Strategy Committee
- Informed other key governance committees at regular intervals, with opportunity to provide feedback

Inclusive of Cybersecurity best practice standards

Benchmarked to other national and international institutions, recommendations



Impact and Future Work

- 173 classification assessments recorded so far
- RDM Service Providers have reported a more efficient and streamlined resolution to queries regarding sensitive data management
- A series of recommendations outlined by the third-party cybersecurity assessment are being implemented to further strengthen the security posture of our research data systems
- Integrate Framework with current University processes (PIA, ethics, contracts etc.)
- Integrate Framework with current training modules
- Evaluate effectiveness of framework with researchers and service providers and iteratively improve resources
- Develop additional resources to supplement provided guidance, as needed



Thank you

For access to these resources or
any other enquiries, please reach
out to us at:

rdm-program@unimelb.edu.au