Managing Climate Change Adaptation Data and Information

A Reference Guide for Element 2, Stream 2 Projects

Version 1.1 (12 February 2014)

The NRM Climate Change Adaptation Information Management Support (CCAIMS) project is funded by the Australian Commonwealth Department of Environment and implemented by Griffith University.

Australian Government
Department of the Environment

Griffith UNIVERSITY

qcif

Enquiries

Sam Mackay, Project Manager
Griffith University NRM Climate Change Adaptation Information Management Support
Nathan Campus, Griffith University
Tel: (07) 373 57599
Email: nrm-informationsupport@griffith.edu.au

Licensing

Acknowledgements

These guidelines are informed by information provided under open licences by other organisations, including:

Australian National Data Service. ANDS guides and other resources. Available at: http://ands.org.au/guides/

Digital Curation Centre. How-to guides. Available at: http://www.dcc.ac.uk/resources/how-guides

Griffith University. Best practice guidelines for researchers: Managing research data and primary materials. Available at: http://www.griffith.edu.au/__data/assets/pdf_file/0009/528993/Best_Practice_Guidelines.pdf

Monash University Library. Research data management guidelines. Available at: http://monash.edu/library/researchdata/guidelines/

UK Data Archive. Available at: http://data-archive.ac.uk/ and as the publication: Van den Eynden, Veerle, et al (2011). *Managing and sharing data: Best practice for researchers.* Colchester: UK Data Archive. Available at: http://data-archive.ac.uk/media/2894/managingsharing.pdf

Disclaimer

Table of Contents

Introduction

The Commonwealth Department of the Environment (DoE) has engaged Griffith University to support the nine project teams funded under Element 2, Stream 2 of the Regional Natural Resource Management (NRM) Planning for Climate Change Fund to integrate best practice in climate change adaptation data and information management into the development and delivery of their climate change adaption information and products. The Griffith University NRM Climate Change Adaptation Information Management Support (CCAIMS) project commenced in July 2013 and is due for completion in June 2014.

The information management protocols contained in this document are a key deliverable for the Griffith University CCAIMS project, and aim to provide the nine NRM Element 2 project teams with guidance and resources on managing the information and data developed throughout their project activities. In order to support the Element 2 project teams in the implementation of these protocols, supplementary awareness materials on information management topics that are identified as high priority will be developed and provided to the Element 2 project teams.

The quality and effectiveness of future NRM adaptation planning and responses will be enhanced if the outputs of Element 2 projects are effectively managed, discoverable, accessible and reusable (by NRM planners, and by other key research, policy and planning groups involved in climate change adaptation). Implementing good data and information management practices into Element 2 project activities will both save time and ensure that important data and information is available beyond the life of the project.

For data and information to be managed effectively, the Element 2 project teams need to implement a planned approach to information management, which instils a shared understanding of how projects will:

- meet the requirements of multiple stakeholders, including: (i) NRM groups (as the primary beneficiary); (ii) DoE (as the funder); and (iii) research institutions (as the implementing agencies).
- consider professional obligations to the broader climate change adaptation community and the general public
- maximise the impact of the research through publication and/or other modes of dissemination, ensuring that outputs are accessible and citable beyond the lifetime of the projects
- manage intellectual property, including establishing clarity around the terms and conditions of re-use of third party datasets, and around data ownership and licensing of project outputs
- fulfil ethical requirements, where applicable
- ensure data is stored and transferred securely, and that technical issues relating to digital file formats and hardware/software dependencies are addressed
- organise and document data, including production of metadata (structured information about the project outputs)

Implementing institutions and other agencies such as the Australian National Data Service (ANDS), can provide infrastructure (institutional, regional, national and international), opportunities to develop professional skills, and access to advice and expertise that enable researchers to apply these protocols. Where possible, this document has identified further resources that could be useful to the Element 2 project teams.

Recommended information management practices

Topic	Summary advice	Protocols section
Planning documentation	Develop and maintain a data management plan, using a template provided by the lead institution or the checklist from the Digital Curation Centre (UK) (see page 10). All project partners should be involved in data planning.	2
Re-using information and data from other sources	Establish as early as possible who owns the rights in third party material and what terms and conditions are associated with their re-use. If 'express permission' (e.g. a copy of a licence or data re-use agreement) is not available, seek permission in writing directly from the rights holders. Avoid verbal agreements and informal written agreements (e.g. personal emails).	3.1
Licensing outputs for re-use	All partners should have a shared understanding of who owns the rights in project outputs. Assume that copyright subsists in all outputs. To encourage maximum re-use of NRM materials, assign the least restrictive open licence that is applicable. The Creative Commons Attribution (CC-BY) licence is recommended. Restricted licences, custom re-use agreements and copyright waivers should be avoided. Not assigning a licence and relying on the Copyright Act alone will limit the re-use that NRM networks can make of outputs from the projects.	3.3
Ethics and consent	Comply with privacy legislation from each State and the National Statement on Ethical Conduct in Research Involving Humans. Consult AIATSIS's Guidelines for Ethical Research in Australian Indigenous Studies (section 13 applies) for advice on projects involving indigenous people.	4
Choosing durable digital file formats	Choose file formats approved by standards bodies and in wide use within the relevant user communities. Identify and manage risks associated with dependencies on	5

Topic	Summary advice	Protocols section
	particular hardware and software.	
Secure storage during the project	Only store master copies on enterprise network drives and on cloud storage services that have been assessed and approved by your organisation. Identify the risks associated with non-approved cloud storage providers (e.g. Dropbox), desktop and laptop computers and removable media, and put strategies in place to manage these risks (e.g. backup and synchronisation, physical security, password protection and encryption, and good handling practices).	6
Secure data transfer	Transfer data using AARNet's Cloudstor service, in preference to email and non-approved cloud services like Dropbox. Cloudstor encrypts your data and accommodates data transfers up to 100GB.	7
Documentation and metadata	Develop agreed strategies early in the project for folder structures and file naming. Assign unique and persistent identifiers. Assess the metadata requirements of potential repositories and ensure that all useful contextual information is captured throughout the project. Where possible, use existing sources of search terms to ensure consistency and ease of discovery.	8.1-8.4
Versioning	Develop a version control strategy early in the project. This could include systematic naming conventions, metadata practices, location tracking, synchronisation, version control facilities in existing software, versioning software, and controlling editing rights.	8.5
Sharing data through a repository	Deposit works in domain-specific (TerraNova) and/or institutional repositories to ensure long-term access and discoverability. Develop an early understanding of your repository of choice's target audiences, and its approaches to preferred file formats, metadata standards, rights management and syndication.	9

Information management benefits

The implementation of a coordinated approach to information and data management will benefit Element 2 project teams, DoE, the element 2 project implementing institutions and the broader climate change adaptation community.

Stakeholder group	Benefits
Element 2 project researchers	• Save time if data is better organised and easier to find • Reduce risks associated with the theft, loss or misuse of data • Gain easier access to data needed for future research • Increase researcher profiles and potentially find new audiences and collaborators through dissemination, citation and re-use of research outputs
NRM groups	• Have long-term access to outputs beyond the life of the Element 2 projects, because storage and technical issues around obsolescence will have been addressed • More effectively find and use outputs, as high quality metadata will be available for discovery and to provide valuable context • Re-use materials appropriately, as ownership and licensing will be as transparent as possible
DoE	• Maximise the return on investment by ensuring outputs are freely available for re-use • Reduce the risk of "reinventing the wheel" with future projects
Implementing institutions	• Identify more research outputs • Increase compliance with government and funding agency requirements, including the *Australian Code for Responsible Conduct of Research* (2007)
The broader climate change adaptation community	• Has enhanced access to NRM research outputs for a variety of purposes: industry, government agencies (outside the NRM planning context), research institutions, and not-for-profit organisations could potentially re-use data and information to support a range of activities.

Protocols for information and data management

1. Legal requirements, policies and professional obligations

In addition to meeting contractual requirements from DoE, Element 2 project researchers are required to manage their data in accordance with a range of standards, including legislation, policies, funding agency requirements, technical protocols, audit and accreditation processes, discipline norms and the expectations of the broader climate change adaptation community.

Researchers should be aware of obligations under the following agreements and ensure that information management practices comply.

Requirements	Description
Code for Responsible Conduct of Research	The *Australian Code for the Responsible Conduct of Research* (2007) assigns researchers and their institutions, especially universities, a shared responsibility to manage research data and primary materials well, by addressing aspects of ownership, storage and retention, and accessibility.
Implementing institution policies	• Intellectual property • Data management • Reporting • Open access
Professional obligations	All researchers should consider the expectations of researchers in their discipline and from other disciplines and how these might affect how information and data should be managed (including sharing, if possible).
Climate change adaptation community expectations	Climate change adaptation is an area of high public interest. Citizens increasingly expect access to the information generated by publicly funded research. The Australian Government's 2010 declaration on open government stated the intention to "make more government information available to the public online" to encourage reuse of that information. The benefits of open access to data include greater engagement by the public in the formation of policy and delivery of services that impact upon them.

2. Planning documentation

DoE does not require a formal data management plan (DMP), but it is strongly recommended that each project complete a DMP.

If the lead institution for the project does not have a DMP template to follow, the template provided by the Digital Curation Centre (UK) provides a good basis:

> DCC. (2013). Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans

All project partners should be involved in the development and signoff of a DMP.

Further resources

Data Management Planning (Australian National Data Service)

---

3. Intellectual property

Note : These protocols contain general information only. This information is not a substitute for obtaining legal advice from your organisation.

3.1 Re-using information and data from other sources

In the context of Element 2 projects, it is important to have permission for the broadest possible re-use. Most Element 2 projects plan to produce work that will be a compilation or derivative of third party materials. The ability to the project outputs for maximum access and re-use (see section 4.3 below) will potentially be impeded by the terms and conditions under which access to source datasets is sought. Projects must therefore:

- Establish who holds the rights in the material being re-used, and
- Establish the terms and conditions of re-use granted by the rights holder/s and assess whether the proposed re-use fits within these. If not, new terms and conditions may need to be negotiated.

In order to do this, project staff will need to follow one of the three steps outlined in Table 1 (see below).

Further resources

Practical Data Management: A Legal and Policy Guide (OAKLaw Project, Queensland University of Technology)

Note: Further guidance on intellectual property will be provided as part of the awareness materials developed by the Griffith CCAIMS team.

a) **Find and keep a copy of any 'express permission' that the rights holder has given**	**OR** b) **If no express permission is given, or if the express permission does not allow for desired use, seek permission *in writing* from the rights holder**	**OR** c) **Re-use material only within the bounds of what is permitted under the Copyright Act**

This will usually be a licence or a set of standard terms and conditions that apply to the process by which the project has obtained the data, such as downloading from websites and online data archives.
 	
Note: When accessing datasets from a source that aggregates information and data – for example, the Atlas of Living Australia – different licenses are likely to be applied to different items by the depositors.
 	
Examples of express permissions for data re-use:
• Lockyer Valley Crop mapping: this dataset, which can be downloaded via the CSIRO Data Access Portal is made available under a CSIRO Data Licence.
• Australian Bureau of Statistics copyright page, which outlines the express permission granted for use of ABS datasets (a Creative Commons Attribution 2.5 Australia licence) and provides guidance on attributing material.
	The permission letter should be dated and should include:
	• Name of the copyright holder
	• Title of the work/works that the project intends to re-use, and where they are located (e.g. web address)
	• A detailed description of the proposed re-use, including
	  ○ how much of the material will be used (all of it? or just some parts?)
	  ○ how it will be used (e.g. reproduced in its entirety; integrated with other datasets)
	  ○ how it will be distributed (e.g. reproduced as part of printed learning materials packs; delivered via an open access repository)
	  ○ whether any commercial use will be made of the material
	• Details of who is making the request: contact person, position, organisation and contact details.
	
	Where possible, verbal agreements and informal written agreements (e.g. personal emails) should be avoided in favour of a standardised permissions process.
		Copyright owners have exclusive rights, including the right to reproduce copyright works, and to publish or communicate the work (e.g. via the web).
		
		If your re-use is not covered by an express permission such as a licence or by permission you have sought directly from the rights holder, you can only use the material in certain ways, such as:
		
		• linking to copyright material available on a public website
		• copying up to 10% of the material to support your work throughout the research process (but not publishing it)
		• reproducing small amounts of the work (e.g. as quotations) for the purposes of research or criticism.
		
		For many Element 2 projects, the uses permitted under the Copyright Act will be insufficient. Integrating different data sources, re-packaging data sources with explanatory material, and re-purposing work as part of training materials are all activities that would *not* be covered under the Copyright Act.

Table 1: Options for re-use of existing information and data

3.2 Data being generated by the projects

It is recommended that the Element 2 project teams take the following steps when creating or collecting data and information.

Requirements	Description
a) Determine what rights, including copyright, will subsist in the information and data produced by the project	Recent case law developments in Australia have led to the situation where it is unclear which data is subject to copyright. In this situation, the Australian Government Open Access Licensing initiative (AusGOAL) has advised a pragmatic approach: AusGOAL suggests that researchers assume copyright subsists in research data, in the same way that it applies to written works like books, journal articles and reports, and is licensed accordingly.
b) Establish who will be the rights holder/s for the data	Ownership of copyright and other IP should be explicitly agreed to by all the parties involved in the research. **Staff members** As a staff member of an implementing institution most data you produce will be legally owned by your organisation. Intellectual property policies can differ from institution to institution though, so be sure to check your organisation's intellectual property policy and (if available) data management policies or guidelines. If the employing institution is a government department or agency, the Commonwealth Government (or a State or Territory Government) will own copyright. **Research students** Higher degree by research (HDR) students usually own the copyright in all materials generated in the course of their studies, including research data. The supervisors of any research students working on the cluster projects *must* discuss copyright and other IP issues with the student prior to their commencement. It may be necessary to complete the relevant IP assignment notification form or deed to ensure that obligations to DoE can be met. **(Sub-)Contractors**

		Ownership of IP, including copyright, should be covered in the contract for services between grant awardees and any sub-contractors. If the contract is not explicit, then the sub-contractor will usually own the copyright and other IP in any outputs they create or compile. This could lead to difficulties for the lead organisation in terms of ensuring that obligations to DoE can be met.
c)	**Determine how the project will manage copyright to permit others to use the project outputs as intended.**	As a copyright owner if you wish to deposit information or data into a repository or database, share it online, and allow others to access and use your data, you will need to manage your copyright to permit these uses. The following section on licensing covers common methods for managing your copyright to achieve these goals.

Further resources

Copyright and Research Data (Australian National Data Service)

Research Data Licensing FAQs (Australian Government Open Access Licensing)

Practical Data Management: A Legal and Policy Guide (OAKLaw Project, QUT)

Note: Further guidance on intellectual property will be provided as part of the awareness materials developed by the Griffith CCAIMS team.

3.3 Licensing for re-use

Where data that is owned or managed by an Element 2 project team member is disseminated, it is recommended that consideration is given to how you want others to re-use it. It is the data owners' responsibility to communicate clearly the terms and conditions for re-using their data.

For each Element 2 project output, project teams will need to choose one of the four options outlined in Table 2 (see below).

Further resources

How to Licence Research Data (Digital Curation Centre, UK)

Research Data Licensing FAQs (Australian Government Open Access Licensing)

Creative Commons and Data (Australian National Data Service)

Creative Commons Australia (website)

Note: Further guidance on intellectual property will be provided as part of the awareness materials developed by the Griffith CCAIMS team.

Some rights reserved: standard open licences *** RECOMMENDED ***	Some rights reserved: restricted licences and custom re-use agreements	No rights reserved: copyright waivers and public domain dedications	All rights reserved: relying on the Copyright Act
For openly accessible data, a standard licence is the most effective way of ensuring appropriate re-use. An open licence lets you reserve some rights as the owner of the material, but grant re-users more rights than they would have just under copyright legislation. Adopting a standard licence is often a pre-condition to depositing in a repository or archive, but licences can also be applied to resources disseminated via the web or other means. Licences enable clearly indicate to others your wishes about how the data can be re-used and how you want to be attributed. Element 2 projects should apply the AusGOAL principle of *using the least restrictive licence that is applicable to the material being licensed*. If you want your material to be as widely used as possible, the Creative Commons Attribution Only licence (CC-BY) would be the most useful for that aim.	If you would like to make data available only under certain conditions or by negotiation, you can use a restrictive licence or other written agreement (such as a Data Transfer Agreement). Agreements of this kind could be constructed from a model template or developed for you especially to meet the requirements of a specific project. A restricted licence provides you with more protection and enables you to be specific about terms and conditions, but it can also be time-consuming and, if legal advice is required, expensive. Element 2 projects should only consider this approach when an open licence is not suitable, e.g. because data contains personal or other confidential information, or if there is a requirement to impose some other condition such as a time limit on use or payment.	Some licences or agreements – such as Creative Commons "No Rights Reserved" (CC0) and the Open Data Commons – Public Domain Dedication - allow rights holder to place work in the public domain. When apply these to your work, you waive all rights and the protections offered by copyright, including the right to be credited as the creator. Element 2 projects should think carefully before using a 'No rights reserved' licence. Waiving all rights means that neither individual researchers nor their institutions must be credited if data is re-used. If you are required by an archive or repository to use a copyright waiver or public domain dedication, you should find out whether any "community norms" statements can be applied: these will not be legally binding but can signal your wishes to potential re-users, where this is practical.	If you do not apply a licence or other terms and conditions, you will reserve all your rights under the Copyright Act. This means people can view and download a copy of your information or data for private research and study only. They must credit you as the creator, and potential re-users would need to seek your permission for any other type of activity, including re-publishing. In the case of information and data arising from the NRM cluster projects, reserving all rights could limit the impact of your work by restricting NRM planners and other audiences from undertaking common activities such as deriving data or aggregating your data with other datasets. As the goals of the Element 2 projects are to disseminate outputs widely and to facilitate the greatest re-use possible for NRM planning processes, applying an open licence will be more effective than relying on copyright legislation.

Table 2: Options for licensing outputs for re-use

4. Ethics and consent

A number of ethical requirements apply to the management of research data, particularly where the research involves human subjects (e.g. surveys, interviews, and analyses of personal data).

Researchers must be aware of their obligations in the following areas, and address these obligations through the existing ethics application processes in your organisation.

Area	Comments
Privacy	Information must be managed in accordance with the relevant privacy legislation in each State, e.g. *Information Privacy Act 2000 (Vic); Information Privacy Act 2009 (Qld)*. You should seek further information as needed from your organisation's privacy officer and ensure that you comply with local policies and guidelines.
Confidentiality and consent	The National Statement on Ethical Conduct in Research Involving Humans, particularly Chapter 3.2 on Databanks, outlines researchers' responsibilities in the areas of collecting, storing, using and disclosing data. You must respect any confidentiality agreement about stored data that has been made with participants. You must retain records of these agreements and ensure that data will not become available for uses to which participants did not consent.
Cultural sensitivity	Research involving indigenous people may have special data management requirements: decision making and communicating about methods of collecting information, including where data is stored and how it is accessed, may need special attention. There may be a need to take into account indigenous intellectual and cultural property rights, in addition to other ownership and use of copyright and intellectual property. The Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) provides resources for learning more about these issues and how to address them.

Researchers need to understand and manage potential tensions between ethical requirements and other requirements for long-term accessibility to research outputs, for the purposes of validating the research and furthering knowledge.

You will need to be explicit in any ethics application about plans you have to make data available, and to describe your strategies for protecting privacy and confidentiality, e.g. by ensuring that

- participants will not be identifiable, and/or
- informed consent will be sought from participants for the proposed data re-use, and/or
- access controls or appropriate re-use agreements will be in place.

Be explicit in your consent forms about any plans to make data available, who will be able to access the data, and how the data would be accessed and potentially re-used.

You will enhance your ability to share data later if you identify broad types of access rather than specific services that may be unsuitable or unavailable in future. For example, saying that you will publish data 'through web-based institutional or subject archives or repositories' will give you more flexibility than if you specify a single repository or archive.

Do not specify in your ethics application that you will destroy data at a certain time, until you have considered the significance of the data and the possibility that it may be used in follow-up work by NRM groups and by other researchers.

Data does not have to be openly accessible to be shared for the benefit of future researchers and other interested groups. In some cases, mediated or restricted access could be more appropriate. Even if your data is confidential, you may be able to share it under certain circumstances, for example:

- if users have to register
- if an application is made to an ethics committee or project board, or
- if a formal re-use agreement protecting participant confidentiality is in place.

Further resources

Ethics, consent and data sharing (ANDS)

Consent and Ethics (UKDA) especially subsections on Consent, Anonymisation, and Access control

Guidelines for Ethical Research in Australian Indigenous Studies (AIATSIS, section 13 applies)

---

5. Choosing durable digital formats

You can assess the durability of the file formats you will use by considering if the format is:

- endorsed and published by standards agencies such as Standards Australia or the International Standards Organisation (ISO)
- publicly documented, i.e. complete authoritative specifications are available
- the product of collaborative development and consultative processes
- widely used and accepted as best practice within your discipline or other user communities.

You should also assess the long-term accessibility your outputs in terms of any hardware and software used to create and manipulate research data.

Considerations include:

- the likely time that the hardware and software will be available
- the size and level of activity of the communities of developers and users associated with the hardware and software
- the level of technical support that is available now and in the future.

If you are developing software as part of your project, follow available best practice guidelines for developing, releasing and licensing your software, such as those provided by the Software Sustainability Institute (UK).

Further resources

File formats: awareness level **(ANDS)**

File formats: working level **(ANDS)**

Tips and tricks for sustainable software and development **(Software Sustainability Institute)**

---

6. Secure storage during the project

Network storage *** RECOMMENDED ***	Cloud services	Desktop and laptop computers	Removable media
Storing your data on enterprise network drives at your institution has the following benefits: • data is readily available to you and other authorised users • standard security and access controls can prevent loss, theft or unauthorised use • automated systems can be put in place for back up, replication and integrity checks.	Cloud storage services – i.e. where data is hosted by third party infrastructure outside your institution are emerging for the research sector but are currently at an early stage of development. Consult with your IT providers about *approved* cloud storage options that may be available to you. Other cloud providers outside of your organisation (e.g. personal Google Drive accounts, Dropbox) should only be used to store additional copies that are not critical, and that contain no private, confidential, or sensitive information. Read the terms and conditions of use of any external service carefully, and assess the risk associated with using it. In particular you should ask yourself: **Who actually has my information and data? What are their policies on intellectual property?** Some cloud services assert their ownership of the intellectual property in anything that is uploaded by users. **Where is the service located? If the organisation delivering the service is in a different jurisdiction, are there any legal implications?** In many cases, storage of material that contains personal information outside Australia will be a breach of national or state privacy legislation.	You should not store master copies of digital data on individual desktop or laptop computers. You should treat these as convenient working areas but not as primary stores. Local drives fail and are often not backed-up. Local machines are regularly replaced, upgraded, and allocated to other people – information and data on those machines is at risk of being lost or inappropriately accessed. If you store working copies on local computers, schedule automatic synchronisation and/or backups and be sure to password-protect and physically secure the machines (e.g. with a laptop lock).	Removable media like CDs and DVDs, flash memory devices (i.e. USB sticks), and portable hard drives are: • not always long-lasting, especially if they are not stored correctly • easy to damage physically (e.g. through magnetism or shocks) • a risk in terms of data security – they are easy to misplace or lose, and often the data does not have access controls. In the case of USB drives, they are also an easy target for viruses and malware. If you store additional working copies on removable media, schedule automatic synchronisation and/or regular backups. You should password-protect

Network storage *** RECOMMENDED ***	Cloud services	Desktop and laptop computers	Removable media
	Who has access to my data and what controls are in place to ensure that they will not misuse my data? What happens if my data is lost or becomes corrupted? What happens to my data if I stop using the service? The Terms and Conditions of some cloud storage services state that they will take no responsibility for data loss and can withdraw the service at any time. If you store additional working copies of non-critical and non-sensitive data on a cloud service, make sure that you have scheduled backups in place, that your account is secure, and that your use of the service is not in breach of any legal or contractual requirements.		and encrypt the media and ensure they are as physically secure as possible. Choose high quality products, and follow the instructions provided by the manufacturer for care and handling, including environmental conditions and labelling. Regularly check media to make sure that they are not failing, and periodically 'refresh' the data (i.e. copy to a new disk, USB stick, or portable drive).
	Further resources Cloud computing and the privacy principles (Office of the Information Commissioner Queensland) – *Note: Refers to Queensland legislation only – similar guidelines may be available in other jurisdictions.*	*Further resources* Set and use strong passwords (Australian Government Department of Communications)	*Further resources* Caring for CDs and DVDs (UK Preservation Office) Using Optical Media for Digital Preservation (UK JISC Digital Media) Recommended encrypted USB memory sticks and external hard drives (University of Exeter)

Table 3: Options for working storage during the project

7. Safe and secure data transfer

Cloudstor *** RECOMMENDED ***	Data transfer using cloud services	Email
Australian researchers can use the Cloudstor service run by AARNET (Australia's Academic and Research Network) to transfer research data, particularly that containing sensitive or personal information. Cloudstor: • can be used to transfer files to/from collaborators at a range of subscriber organisations • encrypts your data before submission • is accessible using your institutional username and password for staff of subscriber organisations • can accommodate large files (up to 100GB in a single transfer). *Important note: Cloudstor is a data transfer service only and does NOT support long-term shared storage of files. Your data is deleted after **20 days.***	While other services provide functionality that is very attractive, you should carefully consider their use. Dropbox, for example, has experienced at least three major security breaches since 2011, including an incident in which all users' files were publicly accessible for several hours. Cloudstor provides a research sector specific alternative that is safer and more secure than other web-based services.	You should avoid using email for data transfer. Some of the limitations of email include: • size restrictions - most institutions have limits on the size of emails and attachments • security risks - particularly if you are working with data that is personally or commercially sensitive and/or utilising personal accounts that may not meet legal and ethical requirements around privacy and confidentiality, and • version control issues.

Table 4: Options for data transfer during the project

8. Organising and documenting data

It is recommended that Element 2 project teams create and maintain sufficient documentation and/or metadata (i.e. structured information about the data) to enable research data to be identified, discovered, associated with its owners and creators, linked to other related data or publications, contextualised in time and space, and to have the quality of the data assessed and research results validated.

If Element 2 project information and data is poorly documented and described, it will be very difficult to find it and manage it in the longer term.

Although information management practices may differ among the element 2 project teams due to the unique regional and sectoral issues facing each project, it is recommended that, consistent documentation and metadata standards are implemented as early as possible.

8.1 Folder structure and file naming

It is recommended that each Element 2 project team develop clear file names for their data and information. Good file names provide useful cues to the content and status of a file, can uniquely identify a file and can help in classifying files. Structuring files in an orderly folder structure also helps to locate and organise files and versions.

File names might contain project codes, researchers' initials, a version number, file status and date. The best hierarchy for files could be deep or shallow, depending on the research. The most important thing is to develop folder structure hierarchies and file naming conventions early in a research project, and ideally to agree on these with colleagues and collaborators before data is created.

In all cases, filenames should be unique, persistent and consistently applied, if they are to be useful for finding and retrieving data.

Best practice is to:

- create meaningful but brief names
- use file names to classify broad types of files
- avoid using spaces and special characters
- avoid very long file names

Further resources

Organising data: file naming (UK Data Archive)

8.2 Metadata standards

Some common descriptive standards are available that work for many different kinds of material and across disciplines.

In many disciplines, you will find existing standards specifically designed for describing and sharing data for that community. The UK's Digital Curation Centre provides a useful list of metadata standards suitable for different disciplines.

Appendix A contains a list of metadata elements that are recommended for use by the Element 2 project teams, with a suggested requirement level (mandatory / recommended / optional).

These elements align with three common standards used in Australia:

- Dublin Core: This is one of the most widely used general metadata schemas and can be used to describe many different types of content.
- ANZLIC Metadata Profile: The current ANZLIC Metadata Profile adopts established Australian / New Zealand and International Standards for the description of spatial data and provides a consistent basis for communicating information about resources. Metadata requirements for the Terra Nova climate change adaptation information hub were developed with reference to ANZLIC for geospatial aspects of descriptions.
- Registry Interchange Format - Collections and Services (RIF-CS): This schema was developed as a data interchange format for supporting the electronic exchange of collection descriptions. It organises information about collections and services into the format required by the Australian National Data Service (ANDS) for Research Data Australia.

Not all repositories will offer all of these elements; sometimes, terminology may differ between repositories.

Further resources

Metadata (awareness level) and Metadata (working level) (Australian National Data Service)

Disciplinary metadata (Digital Curation Centre, UK)

Note: More detailed guidance on metadata creation will be provided as part of the awareness materials developed by the Griffith CCAIMS team.

8.3 Controlled vocabularies

A vocabulary sets out the common language a discipline has agreed to use to refer to concepts of interest in that discipline. It models of the concepts in a discipline, by applying labels to the concepts and relating the concepts to each other in a formal structure.

Vocabularies take many forms. They include glossaries, dictionaries, gazetteers, code lists, taxonomies, subject headings, thesauri, semantic networks and ontologies.

Wherever possible, it is recommended that Element 2 project teams should use an existing controlled vocabulary. Even if you need to adapt or customise an existing standard, this is preferable to creating something from scratch.

For the Element 2 projects, recommended vocabularies, particularly where these are readily available through repository deposit interfaces, include:

- Australian Bureau of Statistics Australian and New Zealand Standard Research Classification (Field of Research codes)
- the Search Words and Qualifiers outlined in the ANZLIC Metadata Profile (see Appendix 2)
- geospatial vocabularies for:
    - place names
    - Natural Resource Management (NRM) regions
    - Local Government Authority (LGA) regions
    - Interim Biogeographic Regionalisation for Australia (IBRA) regions.

Where more specific search terms are desirable, these can be:

- added as keywords or tags (depending on the metadata creation tools available)
- included in the titles and descriptions of the works being described.

Further resources

Vocabularies and research data (awareness level) (Australian National Data Service)

8.4 Identifiers

An identifier is a reference number or name for a data object and forms a key part of your documentation and metadata. To be useful over the long-term, identifiers need to be:

- unique - globally if possible, but at the very least unique within your particular systems and processes, and
- persistent - the identifier should not change over time.

The emerging identifier standard for publicly available datasets is Digital Object Identifier names (DOIs). Although DOIs have been traditionally used for electronically published journal articles, they can now be assigned to datasets.

DOIs can only be assigned through machine-to-machine interfaces via a registration agency. In Australia, the registration agency for datasets and 'grey literature' (reports, working papers, theses etc) is ANDS. DOIs can only be assigned to datasets by depositing your data in a repository that has infrastructure and agreements in place with ANDS for the minting of DOIs. It is not possible to mint DOIs manually as a dataset owner.

Similarly DOIs for publications are minted via publishers that have a subscription with the international registration agency for DOIs for publications, CrossRef. It is not possible to mint DOIs manually for publications that you have authored and published.

Further resources

Persistent identifiers (awareness level), Persistent identifiers (working level) and Persistent identifiers (expert level) (Australian National Data Service)

The Digital Object Identifier and DOI Names (Australian National Data Service)

8.5 Version control

A version control strategy depends on whether files are used by single or multiple users, in one or multiple locations and whether or not versions across users or locations need to be synchronised or not.

Good practices to consider include:

- Decide upfront how many versions of a file to keep, which versions to keep, for how long and how to organise versions
- Identify important versions of files to keep at certain points in time (a 'milestone' version)
- Develop a systematic naming convention, e.g. including at the end of the filename
    - consecutive numbers (version 1 for the first draft, version 2 for the next), or
    - version numbers with points to reflect major and minor changes, such as version 1.1 (first version), version 2.1 (second version with a major change), version 2.2 (third version with a minor change).
- Record version numbers and or version status (draft, final):
    - in the work itself, e.g.
        - in the footer of a draft report
        - embedded in File Properties where this is possible with the software generating the file
        - in a version control table that contains information about the changes made, who made them, and when they were made
    - as part of a descriptive metadata record associated with the work.
- Track the location of files if they are stored in a variety of locations
- Regularly synchronise files in different locations
- Maintain single master files in a suitable file format to avoid version control problems associated with multiple working versions of files being developed in parallel
- Identify a single location for the storage of milestone and master versions of files

If you have version control requirements that cannot be met using file naming and embedded metadata techniques, you may need to consult your IT specialists on other techniques such as:

- version control facilities within the software you are using
- versioning software, e.g. Subversion (SVN)
- controlling rights to file editing
- manual merging of entries or edits by multiple users

Further resources

Version control and authenticity (UK Data Archive)

8.6 Authenticity

Element 2 project teams may need to demonstrate the authenticity of data and information by being able to prevent unauthorised changes.

Recommended techniques for maintaining authenticity include:

- keep a single master file of data
- assign responsibility for master files to a single project team member
- control 'write' access to master files
- record all changes to master files (see versioning techniques above)
- archive copies of master files at regular intervals ('milestone' copies)
- utilise infrastructure that offers authenticity checks such as checksums and provides reports of file corruptions and instances of unauthorised access.

Further resources

Version control and authenticity (UK Data Archive)

---

9. Sharing data through a repository

By depositing data and information in a repository or archive, you can make sure that your material can be accessed and cited in the long term.

Before depositing, you should consider the implications of doing so, in terms of ownership of intellectual property, and ethical requirements like privacy and confidentiality.

Repositories differ in their focus and the types of content that they accept. It is common for repositories to specify some or all of the following:

- preferred file formats that facilitate long-term access and preservation
- minimum standards for documentation and metadata that enhance the discoverability and usability of the material being deposited
- assurances from you, as the depositor, that storing the work and making it available will not infringe upon the rights of others, and
- your assignment of a licence that makes clear what rights re-users are granted.

In choosing a repository to deposit in, it is recommended that Element 2 project teams consider a number of things, including:

- the sustainability of the service (in terms of staffing, funding arrangements, and support from its implementing institution)
- its level of support for and within your discipline
- its reach within the audiences that you are trying to engage
- the extent to which it offers metadata for 'harvest' by other discovery services ('syndication')
- whether the metadata standards used will accommodate search and browse strategies likely to be used with information of the type that you are creating; for example, if location is an important feature of your data, does the repository metadata include locations in appropriately standardised formats such as coordinates?

In some cases it may make sense to deposit your work in multiple repositories, or to deposit in one place with metadata records from other repositories linking to the deposited object in the original repository. While this may create a little extra work, it may be necessary if you are trying to reach different audiences and/or need to comply with different requirements (e.g. from a funder vs. your institution). Automated sharing of metadata between repositories can be investigated to make this process easier.

Further resources

Digital repositories (Digital Curation Centre, UK)

9.1 Terra Nova

Terra Nova is an online repository where researchers and decision-makers can discover, store, share and make use of climate change adaptation information and data. Users of Terra Nova can:

- Search for climate change adaptation information and data
- Upload and store adaptation information and data on Terra Nova, with a metadata schema that links to Research Data Australia (RDA)
- Use selected tools that help overcome common barriers to developing and implementing adaptation responses.

Terra Nova is the *recommended domain-specific repository* for the Element 2 project teams for reasons including:

- It is specifically targeted at the community of users that are interested in climate change adaptation and its profile is likely to grow over the lifetime of the Element 2 projects.
- Terra Nova accepts a broad range of content, including datasets and reports.
- The metadata can accommodate rich information about geospatial locations that is important for NRM planning.
- It has a sustainability plan in place and Griffith University is committed to maintaining, enhancing and promoting the service.
- Metadata for collections is syndicated to Research Data Australia and all content is easily accessible to search engines like Google.
- Terra Nova can provide feeds for live updates of content that can be added to other websites.
- Terra Nova enables other repositories such as institutional data repositories (see below) to harvest collection level records.

Constraints include:

- Metadata harvesting by other services is only supported at the collection level not the item level.
- Terra Nova provides full open access only to metadata and content. There is no ability to place embargoes on material or to restricted access to metadata or content.
- Some limitations on size due to web upload – not optimised for extremely large file uploads.

9.2 Institutional repositories for publications and data

All Australian Universities and some research institutes maintain an institutional repository containing metadata and, where possible, full text content of publication outputs from the organisation. Publication repositories support reporting and quality assessment exercises such as the annual Higher Education Research Data Collection (HERDC) and the 2-3 yearly Excellence in Research for Australia (ERA) initiative.

Many institutions now include datasets within their publication repositories and/or maintain separate data repositories where researchers can store and make available research data.

The benefits of institutional repositories include:

- Sustainability – publication repositories are well-developed core parts of the institutional infrastructure and usually have some dedicated staff resources associated with them.
- Staff can provide local advice about metadata creation.
- IRs may accommodate a range of content, including datasets and reports.
- Metadata is syndicated to Research Data Australia (for collections) and other discovery services such as National Library's Trove (for items).
- IRs in universities enable institutions and researchers to meet their reporting obligations and contribute to research quality assessment.
- IRS may offer the ability to link research outputs with information about the grants held in research management systems.
- IRs may offer more access control, e.g. the ability to place embargoes on material or to restrict access to certain groups of users.

Some constraints relating to institutional repositories include:

- IRs may not accept all types of content, e.g. datasets may be excluded or there may be restrictions about including reports and working papers that are not peer-reviewed.
- Not all content types may be included in the same repository; data may be separate from publications, and there may not be good links between them.
- Metadata may be too generic to meet all the needs of specific communities – for example, locations may only be able to be entered as placenames, not as geospatial coordinates
- Discovery interfaces are generic – for example, few IRs offer searches limited by geographic region.

Appendix A: Suggested metadata elements for Element 2 projects

Element 2 recommended core metadata elements	Level of requirement	Notes	Terra Nova mapping	Dublin Core mapping
Title	Mandatory		Title	title
Description	Mandatory		Brief description Detailed summary	abstract description
Contributors	Mandatory (at least one)		Contributors	creator
Point of contact	Recommended		Custodian	contributor
Publisher	Mandatory		Owning organisation Logo	publisher rightsHolder
Content type	Recommended		Content type	type
Placename	Recommended		Placename	coverage
Coordinates	Recommended		Coordinates	spatial
LGA region	Optional	Controlled list available in Terra Nova	LGA region	spatial
NRM region	Recommended	Controlled list available in Terra Nova	NRM region	spatial
IBRA region	Optional	Controlled list available in Terra Nova	IBRA region	spatial
Temporal coverage	Optional		Temporal Coverage	temporal
Date created	Recommended		Creation / publication date	created

Element 2 recommended core metadata elements	Level of requirement	Notes	Terra Nova mapping	Dublin Core mapping
Date uploaded	Mandatory	Usually system-generated	Upload Date	dateSubmitted
Expiration date	Optional		Expiration date	valid
Keywords	Recommended	Can be selected from ANZLIC search words and qualifiers	Thematic Tags	subject
Sector focus	Recommended if available	Controlled list available in Terra Nova	Sector Focus	subject
Fields of Research	Recommended	Controlled list available in Terra Nova and many institutional repositories	Fields of Research	subject
Identifier	Recommended	Required for Terra Nova if linking to external content	External identifier/DOI	identifier
Rights statement	Mandatory		Rights statement	rights
Licence	Recommended		*[Can be included in Rights statement]*	license
Access statement	Optional		Attributions and Constraints	accessRights
Source	Recommended, if applicable		*[Can be included in Rights statement]*	source
Related	Optional		Related Website	relation
Citation	Optional		Citation	bibliographicCitation

Appendix B: ANZLIC Search Words and Qualifiers

The following list of ANZLIC search words is provided as a recommended source of keywords that might be relevant to the Element 2 projects. These can be chosen from a controlled list (where this functionality is available in your repository of choice), or alternatively, used in titles, descriptions and added as keywords or tags to facilitate discovery.

AGRICULTURE
AGRICULTURE Crops
AGRICULTURE Livestock
AGRICULTURE Horticulture
AGRICULTURE Irrigation
ATMOSPHERE
ATMOSPHERE Air Quality
ATMOSPHERE Ozone
ATMOSPHERE Greenhouse
ATMOSPHERE Pressure
BOUNDARIES
BOUNDARIES Administrative
BOUNDARIES Biophysical
BOUNDARIES Cultural
CLIMATE AND WEATHER
CLIMATE AND WEATHER Meteorology
CLIMATE AND WEATHER Climate change
CLIMATE AND WEATHER Drought
CLIMATE AND WEATHER El Nino
CLIMATE AND WEATHER Extreme weather events
CLIMATE AND WEATHER Radiation
CLIMATE AND WEATHER Rainfall
CLIMATE AND WEATHER Temperature
DEMOGRAPHY
DISEASE
ECOLOGY
ECOLOGY Community
ECOLOGY Ecosystem
ECOLOGY Habitat
ECOLOGY Landscape
ENERGY
ENERGY Coal
ENERGY Electricity
ENERGY Petroleum
ENERGY Renewable
ENERGY Use
FAUNA
FAUNA Exotic
FAUNA Insects
FAUNA Invertebrates
FAUNA Native
FAUNA Vertebrates
FISHERIES
FISHERIES Aquaculture
FISHERIES Freshwater
FISHERIES Marine
FISHERIES Recreational
FLORA
FLORA Exotic
FLORA Native
FORESTS
FORESTS Agriforestry
FORESTS Natural
FORESTS Plantation
GEOSCIENCES
GEOSCIENCES Hydrogeology
GEOSCIENCES Geochemistry
GEOSCIENCES Geology
GEOSCIENCES Geomorphology
GEOSCIENCES Geophysics
HAZARDS
HAZARDS Cyclones
HAZARDS Drought
HAZARDS Earthquake
HAZARDS Fire
HAZARDS Flood
HAZARDS Landslip
HAZARDS Manmade
HAZARDS Pests
HAZARDS Severe local storms
HAZARDS Tsunamis
HEALTH
HERITAGE
HERITAGE Aboriginal
HERITAGE Architectural
HERITAGE Natural
HERITAGE World
HUMAN ENVIRONMENT
HUMAN ENVIRONMENT Economics
HUMAN ENVIRONMENT Housing
HUMAN ENVIRONMENT Livability
HUMAN ENVIRONMENT Planning

HUMAN ENVIRONMENT Structures and Facilities
HUMAN ENVIRONMENT Urban Design
INDUSTRY
INDUSTRY Manufacturing
INDUSTRY Mining
INDUSTRY Primary
INDUSTRY Service
INDUSTRY Other
LAND
LAND Cadastre
LAND Cover
LAND Geodesy
LAND Geography
LAND Ownership
LAND Topography
LAND Use
LAND Valuation
MARINE
MARINE Biology
MARINE Coasts
MARINE Estuaries
MARINE Geology and Geophysics
MARINE Reefs
MARINE Human Impacts
MARINE Meteorology
MINERALS
MOLECULAR BIOLOGY
MOLECULAR BIOLOGY Genetics
OCEANOGRAPHY
OCEANOGRAPHY Physical
OCEANOGRAPHY Chemical
PHOTOGRAPHY AND IMAGERY
PHOTOGRAPHY AND IMAGERY Aerial
PHOTOGRAPHY AND IMAGERY Remote Sensing
PHOTOGRAPHY AND IMAGERY Satellite
POLLUTION
POLLUTION Air
POLLUTION Noise
POLLUTION Soil
POLLUTION Water
SOIL
SOIL Erosion
SOIL Biology
SOIL Chemistry
SOIL Physics
TRANSPORTATION
TRANSPORTATION Air
TRANSPORTATION Land
TRANSPORTATION Marine
UTILITIES
VEGETATION
VEGETATION Floristic
VEGETATION Structural
WASTE
WASTE Liquid
WASTE Solid
WASTE Toxic
WASTE Sewage
WASTE Greenhouse gas
WASTE Heat
WATER
WATER Groundwater
WATER Hydrology
WATER Hydrochemistry
WATER Lakes
WATER Rivers
WATER Salinity
WATER Supply
WATER Surface
WATER Quality
WATER Wetlands

ANZLIC also provides qualifier words that might also be relevant as a source of keywords for the Element 2 projects:

Biodiversity
Classification
Conservation
Distribution
Exploration
Indicators
Inventory
Management
Mapping
Maps
Models
Monitoring
Networks
Planning
Production
Reference
Reports
Research
Reserve
Resources
Statistics
Surveys
Sustainability