[Research Data Planner](https://confluence.csiro.au/display/RDM/Research+Data+Planner)

Check https://confluence.csiro.au/display/RDM/Research+Data+Planner for updates and links to guidance. Contact the Research Data Support (RDS) Team via [researchdatasupport@csiro.au](mailto:researchdatasupport@csiro.au) for any assistance.

**Introduction**

The Research Data Planner is a guide to help you to consider the research data management requirements for your project. Research data management planning will help you to manage your data for your team's use, meet legal, ethical and funder requirements, and allow easier data sharing in the future. Managing and documenting your data throughout its life cycle will assist you to understand your data and support its use, access and appropriate storage. Planning your data management requirements at the beginning of a project and updating it during the project will ensure that appropriate permissions are obtained, infrastructure requirements are there when needed, and risks are minimised. Increasingly, journals are requiring that research data is made available as a condition of publication. Undertaking data management planning will make sure it's easier to publish it when you need to.

Consider from the beginning of the project the intention for the dissemination and/or long term archival storage of all the data, software and other outputs of the project after the project concludes. See [Once the project is complete https://confluence.csiro.au/display/RDM/Once+the+project+is+complete](https://confluence.csiro.au/display/RDM/Once+the+project+is+complete) for guidance.

The factors that you should consider have been divided into 5 broad areas:

1. [Constraints and obligations](#_1._Constraints_and)
2. [Access](#_2._Access)
3. [Description](#_3._Description)
4. [Processes](#_4._Processes)
5. [Storage and compute](#_5._Storage_and)

Every project should have people or roles nominated to have data management responsibilities. The name these roles will have may vary but may be, for example, adata custodian, a data champion, a data manager and/or a data steward. Make sure your planning documents these roles and their responsibilities.

Research data management is for everyone. Share and discuss each factor with your project team to communicate and set in place the processes and guidelines for your team. See [Getting started](https://confluence.csiro.au/display/RDM/Getting+started) <https://confluence.csiro.au/display/RDM/Getting+started> for an exercise to start your conversation. Include your data management plan with your ethics application. Review and update the plan regularly in consultation with your research partners and where appropriate the human research ethics committee during the course of the project.

## About this project:

Project O2D Number: OD-211433 (WBS – R-11964-01)

Project Name: Grainscore Digiscape FSP or AgScores

Business Unit: Agriculture and Food

Project leader: Jaci Brown

Data Custodian (or other role responsible for the data in this project: Neville Herrmann

Data Role: Click here to enter text.

Scope of Responsibilities:Click here to enter text.

## 1. Constraints and obligations

**These aspects will impact your ability to do your research and the way that you intend to publish or share the results. They will influence the answers to the other aspects in your project's data management planning.**

From the beginning of your project you need to consider your intentions for the dissemination or long term archival of the data and software outputs of your project after the project concludes.

* Data produced by CSIRO should be open and published by default in accordance with [Australian Government policy](https://www.pmc.gov.au/sites/default/files/publications/aust_govt_public_data_policy_statement_1.pdf)[[1]](#footnote-1). Data should be published unless there are legal, contractual or privacy requirements that prevent this.
* You need to plan for the constraints and conditions that will have an impact on whether you can publish your data as open access.

**Use this checklist to document your project's constraints and obligations. Choose those that apply to your project**

|  |
| --- |
| 1.1 Using data that you are bringing in from outside the project? |
| This could include "reference data" from other CSIRO projects.  You need to document the ownership and conditions of use.  Create and keep an updated list of data that the project is planning to bring in or acquire. For each item include:   * + Which organisation owns the data? Bureau of Meteorology (ACCESS-S forecast model) / State of Queensland (SILO database)   + Will the acquired data need attribution and how should it be cited?   ACCESS-S Hudson et al, ACCESS-S1: The new Bureau of Meteorology multi-week to seasonal prediction system, 2017, [**http://www.bom.gov.au/jshess/docs/2017/Hudson.pdf**](http://www.bom.gov.au/jshess/docs/2017/Hudson.pdf)  Operational Implementation of ACCESS-S1 Seasonal Prediction System, [**http://www.bom.gov.au/australia/charts/bulletins/apob120\_external**](http://www.bom.gov.au/australia/charts/bulletins/apob120_external)**.pdf**  SILO © The State of Queensland 2019. The Queensland Government supports and encourages the distribution of its material.Unless otherwise noted, all copyright material available on or through this website is licensed under a [**Creative Commons Attribution 4.0 International licence (CC BY 4.0)**](https://creativecommons.org/licenses/by/4.0/)**.**  [Creative Commons Attribution 4.0 International (CC BY 4.0)](https://creativecommons.org/licenses/by/4.0/)  <https://www.longpaddock.qld.gov.au/silo/about.html>   * + Are there intellectual property considerations associated with hosting, using or sharing the data? No   + What are the licence conditions? Document any conditions of use. Do they permit your intended reuse? Are there restrictions on publishing any derived data? When negotiating use conditions for acquired data, try to negotiate open conditions and document the results. Will not be publishing the data in any way. NA   + Will any inherited licences restrict your licence choices on any derived data? No   If you bring in or acquire external data, you should be aware that the data can be requested under the FOI Act. It is important that you keep all related information about the data, including the relevant contract. This information will assist in determining whether the data will have to be released under FOI. If you have concerns please contact the FOI Unit [foi@csiro.au](mailto:foi@csiro.au) |
|  |

|  |
| --- |
| **1.2 Using code libraries, tools or third party code acquired from outside the project? Yes** |
| You need to document the conditions of use.  Keep a list for each item. For each one include:  Python <https://docs.python.org/3/license.html>  R <https://www.r-project.org/Licenses/>  Jupyter <https://opensource.org/licenses/BSD-3-Clause>   * Any conditions of use. Do they permit your intended use and reuse? * All licences and conditions, including any compatibility issues such as sharing and access by third parties. |
|  |

|  |
| --- |
| **1.3 Is this project generating data and software?** |
| You need to plan the ownership and proposed conditions of use.  Include:   * Which of the collaborating organisations for this project will own the data and software generated by this project? CSIRO * Which organisation will host published data and software after the project is finished? Will they assign a persistent identifier? CSIRO * Who will be responsible for releasing software or data? CSIRO * Is there any reason why this data and software will not be made available publicly at the end of the project? NO * What licensing conditions will need to be in place for sharing the generated or derived data or software or code? * What licence is intended to be used for:   + released data? See [Data Deposit checklist](https://confluence.csiro.au/display/daphelp/Data+Deposit+Checklist)[[2]](#footnote-2) no data will be generated   + released software (including code)? See [Software publication release process](https://confluence.csiro.au/display/OSS/Software+Release+Process+Home)[[3]](#footnote-3)   No software will be released at the end of this project but will feed into the larger AgScores project. It will be available on GitHub   * Which people, organisations or groups will be given credit for generated data or software? Consider how you want your generated data and software attributed.   No data will be generated.  Javier Navarro-Garcia and Eva Zinkovsky will be given credit for the software generated.  You may need to consult with:   * [Legal](http://my.csiro.au/Support-Services/Legal.aspx)[[4]](#footnote-4) for Legal advice – no need. Will not be publicly released at the end of this project but will feed into the larger AgScores project. * [Business Development & Commercial](http://my.csiro.au/business-units/development/business-centre)[[5]](#footnote-5) for Contracts advice – no need. No external release. * [Indigenous Cultural & Intellectual Property (ICIP) policy](https://confluence.csiro.au/pages/viewpage.action?pageId=411697886) (work in progress) – N/A |
|  |

|  |
| --- |
| 1.4 Ethics and sensitive data**. NO** |
| Does your project involve animal or human research including social and food and health research? See [Ethics](https://my.csiro.au/Tasks/Science/Research-Ethics-in-CSIRO)   * If so, have you applied for/or received ethics clearances? Document your ethics clearance documentation number. * Document any ethical clearance conditions on your research data. Are there access constraints? * If you are collecting paper based sensitive material also consider who has access to this and how you will store this. * Sensitive Data : document any issues that will require access to be limited or will otherwise affect your ability to share or publish your data due to confidentiality or other issues. Refer to the ANDS [sensitive data guide](http://www.ands.org.au/working-with-data/sensitive-data) for more information. |
|  |

|  |
| --- |
| 1.5 Privacy**. NO** |
| If your project involves personal information in any way (information about an identifiable individual, or who is reasonably identifiable), including   * collecting * reusing * sharing * storing   Then you will be required to complete a threshold [Privacy Impact Assessment checklist](http://my.csiro.au/privacy) to determine if a more detailed PIA is required for the project. A PIA is not is not necessary if the project will not deal with personal information in any way.  Consider the following:   * Are you storing any personal information with a third party? * Will the data need to be de-identified prior to publication? * Will it only be able to be shared with specific parties? |
|  |

|  |
| --- |
| **1.5 Other issues NA** |
| Document other issues that will affect the data and software management for this project. Check for:   * Any other terms in any project contracts that will affect data management that was not covered above. * Any policies relating to any instruments you use. * The data and software sharing or publication requirements of your funding provider. * Any other issues to do with patents, copyright and other IP restrictions. * The submission guidelines of preferred target journals for requirements regarding making data and software available. Any CSIRO or government policies or legislation or policies of collaborating organisations that apply. |
|  |

## 2. Access

**THIS ASPECT WILL INFLUENCE YOUR STORAGE AND TRANSMISSION CHOICES.**

Given the constraints and obligations documented in the previous section consider who will need access to the project research data at each phase of the project.

|  |
| --- |
| **2.1 During the project** |
| Document who will need to be able to access the data storage, computing facility or code repository including CSIRO colleagues and/or external collaborators during the project and their level of access (eg. read only or read/write).   * What level of security is required, for example password protection or encryption for confidential data? * Will a data sharing agreement be required? No   Suggestion: List project members, specify if they are CSIRO or external and specify their required access levels. This will be useful when requesting storage. |
| Eva Zinkovsky - read/write  Javier Navarro-Garcia – read/write |

|  |
| --- |
| **2.2 After the project** |
| Document who will need to be able to access the data storage, computing facility or code repository including CSIRO colleagues and/or external collaborators after the project ceases and their level of access (eg. read only or read/write)   * What level of security will be required if any, for example password protection or encryption for confidential data? No * How will retained versions be managed and preserved after the project has finished, and by whom and where? DAP * If the data is to be made publicly available which repository will be used? When will it be available? Will a persistent identifier be available? GitHub, DAP. Yes, DAP generates a persistent identifier |
| Needs to be accessible to members of the broader AgScore project  Jaci Brown  Eva Zinkovsky  Javier Navarro-Garcia  Chris Sharman  Phil Smethurst  Dean Holzworth  Elizabeth Meier  Mac Coombe |

|  |
| --- |
| **2.3 During any embargo N/A** |
| After the project will any embargo need to be applied?  Document who will need to be able to access the data storage, computing facility or code repository including CSIRO colleagues and/or external collaborators during the embargo period and their level of access (eg. read only or read/write) |
|  |

|  |
| --- |
| **2.4 Other access issues** |
| Note any other access considerations. For example document the location of existing data that you plan to acquire. How will it be transmitted to where it will be used? If not transmitted how will you be accessing it? |
| The climate data is from the Australian Community Climate Earth-System Simulator (ACCESS-S).  It is stored on CSIRO's THREDDS server and accessible at this url:  <https://dcdp.research.csiro.au/access-s?lat=-35.99&lon=142.92&format=csv&start>=  #See instructions in <https://confluence.csiro.au/display/Digiscape/Climate+Forecast+Data+aggregation>  The SILO data can be accessed at this site  <https://www.longpaddock.qld.gov.au/silo/about.html> |

## 3. Description

**This aspect will influence your project ability for each project member to find and use the data during the project and then easily publish or otherwise share your data at the end of a project.**

Your project team will need to decide whether you will be using files and folders for the project's data or some sort of database. Whatever is decided you will need to also plan the description or metadata for the data at some level.

Use the following checklist for metadata planning:

|  |
| --- |
| **3.1 Data collection description: NA** |
| If not using a database how are you going to make your data collection discoverable? Answering this will help you share your data and publish your data later.  If you are using a database how will you make it discoverable?   * Document any decisions relating to description of this data or software as a collection; where known, include any metadata standards that you need to comply with, and/or vocabularies that you need to use. |
|  |

|  |
| --- |
| **3.2 Folder/directory naming:** |
| If not using a database how are you going to help each other find the relevant sections of your data?   * Document decisions regarding folder and directory naming and structures if relevant. * Document any standards that apply. Include any metadata standards that you need to comply with, and/or vocabularies that you need to use. |
| Find out what the fileshare (i.e. basefolder) is for the project (AgScore), then paths will be  basefolder/working/zin005/python/ (etc. e.g. R, data, )  basefolder/working/nav00a/python (etc.)  basefolder/scripts (these are the latest published version of scripts for people to use)  I STILL NEED TO FIND OUT WHERE THE FILES WILL BE STORED – PROJECT LEADER IS AWAY AT THE MOMENT |

|  |
| --- |
| **3.3 File naming:** |
| If not using a database how are you going to help each other find relevant data files and the relevant versions?   * Document decisions regarding file names if relevant. * Document any standards that apply. Include any metadata standards that you need to comply with, and/or vocabularies that you need to use. * Document how versions will be indicated and tracked in your naming conventions (if using). |
| We adopt a mini-convention where we go for more descriptive filenames wherever possible. |

|  |
| --- |
| **3.4 Fields and variables:** |
| How are you going to help each other make your data interoperable?   * Document joint decisions regarding field names and variable formats. * Document any standards that apply. Include any metadata standards that you need to comply with, and/or vocabularies that you need to use. |
| Longer, more descriptive names are better than shorter, non-descriptive names |

|  |
| --- |
| **3.5 File formats:** |
| This will affect your storage and/or service selection and interoperability.   * What file formats will be provided or are intended to be generated– database, jpeg, xls, etc? Take into account the end users of your data, the tools that you and they will be using and any domain norms. * Are there any issues relating to the formats such as special software or hardware requirements – GeoServer, MapServer, OpenDAP, THREDDS? * Are the formats being provided or generated according to relevant standards? |
| Data will be coming from the thredds server but the project will not be generating any new data. Our outputs are summary statistics of an ensemble model compared to summary statistics of SILO data. We will output a summary report in pdf. |

## 4. Processes

**These aspects will influence the reproducibility of your research and your projects outcomes**

|  |
| --- |
| **4.1 Version control:** |
| How will versions of files or databases be managed? A joint understanding of version control processes will aid communication, save time and aid reproducibility. Decisions about managing versions will impact storage and access decisions during and after the project.   * Document how you will distinguish between versions of generated data or software. How will contributors or users know that the data or software has been updated? How will you get updates for acquired data and software? * For software what revision control process will be used? * Document which versions need to be retained and for how long. Do you have ethics clearance compliance factors to consider? * How will retained versions be managed and preserved after the project has finished, and by whom and where? * If there is a service which delivers the data how will this be managed after the project finishes? How will it be funded? |
| A Github repository will be generated for the project |

|  |
| --- |
| **4.2 Reproducibility:** |
| A joint understanding of work flow will aid communication, save time and aid reproducibility. Document:   * What are the proposed data collection and analysis workflows? * How are you going to document the processes used (or the provenance of your worked data)? * How will you document which software and version was used for analysis? Git hub README file * What will be the process for keeping an up to date list of software assets and dependencies? |
| All software will be well annotated with outcomes and analysis in the GitHub repository |

|  |
| --- |
| **4.3 Resourcing:** |
| Document the resources, responsibilities and processes that may be required for data management processes such as:   * storage and management of physical samples, specimens and primary materials * data cleaning acquired data * standardising data from varied project inputs * preparing the data for publication * de-identifying or otherwise processing sensitive data * providing access to any of the following: website, online service, database, data, or software after the project has been completed   If software is a primary output of the project you may need to also consider and document your intended software development processes. See <https://www.software.ac.uk/software-management-plans>.  Estimating these factors early will ensure the project is adequately resourced. |
| The project will produce software for comparative analytics that will be hosted in the Senaps platform. Senaps already manages the entire workflow from uploading the data to providing the summary report. |

## 5. Storage and compute.

**Data loss will severely impact delivering your research.**

It is recommended that all participants in a project use the same storage location for ease of management. Using [CSIRO IMT managed storage](https://confluence.csiro.au/display/RDM/Storage+choices)[[6]](#footnote-6), if suitable, will reduce your data management resourcing requirements. The location of the storage selected needs to be planned and documented so the project participants are informed. The suitability of a storage location will depend on the project stage and will change. Consideration of the destination of the data and software outputs of a project after the project concludes needs to be planned.

|  |
| --- |
| 5.1 Requirements: |
| Document the project's requirements for storage and compute not already covered by the rest of your plan. Include:   * An estimation of the volume of data storage required. * An estimation of its rate of growth. * The type of data storage required. See [Data Categorisation into Storage Block Types](https://confluence.csiro.au/display/SC/Data+Categorisation+into+Storage+Block+Types)[[7]](#footnote-7). * Whether significant compute power (i.e. HPC High Power Computing) will be required. * Any other special needs (e.g. if there will be a lot of small files). * If you will be developing code and so need a project code repository. |
|  |

|  |
| --- |
| 5.2 Existing Storage: |
| Having a joint decision and information about which storage solution to use will save moving data around and losing data. Using IMT managed storage will reduce the potential for data loss.  Do you have **existing managed storage** which fully fits your project's requirements above (taking into account size requirements, access, backups, compute requirements, special format requirements) at all stages of your project?  If Yes. Where? List the locations with path(s) if applicable including project stage (e.g. data acquired from outside the project, raw data, working data, data to be published).   * Document how the data will be stored, backed up and secured if not on managed CSIRO IM&T storage. * Do the proposed storage and compute locations comply with access and governance and format considerations documented in your plan? |
| Yes, tbc. |
| **5.3 Existing Processing:** |
| Thinking about computing resources early will enable those resources to be available when required.  If you require significant computing (HPC) do you already have access to HPC that fully fits your projects requirements?  If yes please specify where. Document available HPC facilities. . |
| NA |

|  |
| --- |
| **5.4 Existing Code Repository:** |
| Deciding on and using a joint code repository will enable good code development practices.  Do you have an existing code repository which fully fits your project's requirements at all stages of your project? If Yes. Where?  If not and you need a code repository contact [atlassiansupport@csiro.au](mailto:atlassiansupport@csiro.au) to create a project in <https://bitbucket.csiro.au/>. |
|  |

|  |
| --- |
| **5.6 If no to any of above- use the decisions in your project's Data Management Plan to request storage and compute via** <https://bowen.it.csiro.au/> |

1. <https://www.pmc.gov.au/sites/default/files/publications/aust_govt_public_data_policy_statement_1.pdf> [↑](#footnote-ref-1)
2. <https://confluence.csiro.au/display/daphelp/Data+Deposit+Checklist> [↑](#footnote-ref-2)
3. <https://confluence.csiro.au/display/OSS/Software+Release+Process+Home> [↑](#footnote-ref-3)
4. <http://my.csiro.au/Support-Services/Legal.aspx> [↑](#footnote-ref-4)
5. <http://my.csiro.au/business-units/development/business-centre> [↑](#footnote-ref-5)
6. <https://confluence.csiro.au/display/RDM/Storage+choices> [↑](#footnote-ref-6)
7. <https://confluence.csiro.au/display/SC/Data+Categorisation+into+Storage+Block+Types> [↑](#footnote-ref-7)