Plan Overview

A Data Management Plan created using DMPonline

Title: Cellular and spatial regulation of Arabidopsis seed germination at control and elevated temperatures

Creator: Inès Marais

Principal Investigator: Mariana Silva Artur, Harm Nijveen, Leónie Bentsink

Data Manager: Mariana Silva Artur, Harm Nijveen, Leónie Bentsink

Project Administrator: Mariana Silva Artur, Harm Nijveen, Leónie Bentsink

Affiliation: Wageningen University and Research (Netherlands)

Template: Data Management Plan | Wageningen University and Research

ORCID iD: 0000-0002-6832-2407

ORCID iD: 0000-0002-9167-4945

ORCID iD: 0000-0001-9510-6059

Project abstract:

Seed germination is tightly regulated and highly sensitive to temperature, yet key regulatory processes remain poorly understood at the cellular, spatial, and post-transcriptional levels under both standard and elevated temperature conditions. This project addresses these gaps by investigating how elevated temperatures during seed development and germination shape gene expression regulation in *Arabidopsis thaliana*. By comparing single-cell and nuclear transcriptomes, the project aims to discover which transcripts are available for translation. Spatial and cell-type-resolved analyses will reveal how different seed tissues coordinate germination, while comparative transcriptomic analyses with tomato and other crop species will assess the conservation of temperature-responsive mechanisms. Together, this work provides a multi-layered view of how seeds regulate germination under changing temperature conditions.

ID: 193069

Start date: 01-07-2025

End date: 30-06-2029

Last modified: 12-12-2025

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Cellular and spatial regulation of Arabidopsis seed germination at control and elevated temperatures

---

A. Describe the research project

1. Name researcher (please, add your full name):

Inès Marais

2. What is the name of your department(s)?

- Plant Sciences

3. What is the name of your chair group(s) or business unit(s)? English name and abbreviation for chair groups from this page; business units from this page (expand to Wageningen Research and keep expanding to find your specific division / group). Examples: Bioprocess Engineering (BPE) or Contract Research Organization (CRO).

Plant Physiology (PPH)
Bioinformatics (BIF)

4. Describe the organisational context of your research project.

DMP version (or date last modified)	2025/12/10
Supervisor / (co-)promotors	Prof. Leónie Bentsink (Promoter) Dr Harm Nijveen (Co-promoter) Dr Mariana Silva Artur (Co-promoter)
Graduate School (WU only)	Experimental Plant Sciences (EPS)
Start date of project	2025/07/01
End date of project	2029/06/30
Project number	3184800154 3184819039
Funding body	NWO-AES project HEAT WUR strategic budget

5. Give a short description of your research project.

Title	Cellular and spatial regulation of Arabidopsis seed germination at control and elevated temperatures
Summary	This PhD project investigates how seed germination is regulated at the transcriptional and post-transcriptional levels under standard and elevated temperature conditions. Using Arabidopsis thaliana as a model, the project combines single-cell and single-nucleus RNA sequencing, spatial transcriptomics (tomo-seq), and degradome sequencing to study gene expression, mRNA availability, and mRNA degradation during germination. By comparing single-cell and nuclear transcriptomes, the project aims to infer which transcripts are available for translation. Spatial and cell-type-resolved analyses will reveal how different seed tissues coordinate germination, while comparative transcriptomic analyses with tomato and other crop species will assess the conservation of temperature-responsive mechanisms. Together, this work provides a multi-layered view of how seeds regulate germination under changing temperature conditions, with relevance for crop resilience in a warming climate.

6. List the individuals responsible for the following data management tasks.

Data collection	Inès Marais, Martijn Jansen (Post-doc)
Data quality	Inès Marais, Martijn Jansen, Harm Nijveen, Leónie Bentsink
Storage and backup	Inès Marais, Harm Nijveen
Data archiving / publishing	Inès Marais, Harm Nijveen, Leónie Bentsink
Data stewardship / support	Judith Risse (BIF data steward)
Any other role [......]	

7. I have requested a review of this data management plan from:

- The (coordinating) data steward of my chair group / business unit.

8. Name of the data management support staff and / or data steward consulted during the preparation of this plan and date of consultation.

Judith Risse (BIF data steward)

B. Describe the data to be collected, software used, file formats and data size.

9. Will you use existing data for this project?

- Yes. Please specify below which data (e.g. DOI, URL, or storage location) and the terms of use (e.g. licence).

Public sequencing data for *Arabidopsis thaliana* will be used; these data are available via the Sequence Read Archive (SRA) hosted by the NCBI. The resulting studies will fairly acknowledge the used data. There are no specific terms of use for the data; the accessibility is via the SRA-NCBI website (Home - SRA - NCBI), and the identification of each dataset will be reported.

10. Will new data be produced?

- Yes.

11. Please describe the data you expect to generate and / or use in the table below. Include reused existing data as well (as these are files that you manage and store).

File contents	Data type	Software	(Open) file format	Estimated size of each file (range)	Estimated number of files (range)
(e.g. lab analysis, gene sequence, interviews, lesion scores, etc.)	(e.g. numerical)	(e.g. Excel)	(e.g. .csv)	(e.g. 20-50 Mb)	(e.g. 50-100)
Gene sequences	snRNA-seq	SRA	.fastq	4-20GB	>10
Gene sequences	scRNA-seq	SRA	.fastq	4-20GB	>10
Gene sequences	snRNA-seq	Sequencing platform	.fastq	4-20GB	50
Tomo-seq	Bulk RNA-seq slices	Sequencing platform	.fastq	4-20GB	10
mRNA co-translational decay	Degradome-seq	Sequencing platform	.fasta/.fastq	4-20GB	10
Scripts for analysis	Scripts	Script programming	Python (.ipynb), snakemake, R scripts	10-40KB	>10
Statistical output, information from the analyses	Statistical values	Excel	.csv or .txt	1-10KB	>5

12. Estimate how much data storage you require in total (e.g. by using the information in the table at question 11).

- >1000 GB

C. Storage of data and data documentation / metadata during research

13. Where will the data,code and accompanying documentation / metadata be stored and backed up during the research project (see the WUR Data Storage Finder)? Include platforms you use to share data, collect data on, or send data to for processing or analysis.

- WUR OneDrive for Business - only when an up to date version of the research data is also safely stored on the W:drive or Yoda.
- Git@WUR (GitLab locally hosted at WUR)
- W:drive Enterprise File Storage (WUR network drive).

D. Structuring your data and information

14. Give a (visual) representation of the folder structure you intend to use.

```
└── Spatial_seed_transcriptomics/
├── Data management
├── WP1_arabidopsis_RNAseq_SRAdata/
│ ├── Metadata (containing information on the SRA accessions used)/
│ ├── Raw data (fastq files are publicly available for snRNAseq and scRNAseq)/
│ ├── Intermediate data and syntax/
│ ├── Final data and syntax/
│ └── Publications/
├── WP2_species_bulkRNAseq_heat/
│ ├── Metadata (Files describing the project & information on the collected data)/
│ ├── Raw data (bulk RNA seq from various closely related species)/
│ ├── Intermediate data and syntax/
│ ├── Final data and syntax/
│ └── Publications
├── WP3_snRNAseq_spatial_transcriptomics/
│ ├── Metadata/
│ ├── Raw data (snRNA-seq data generated and tomo-seq data -- bulk RNA-seq slices)/
│ ├── Intermediate data and syntax/
│ ├── Final data and syntax/
│ └── Publications/
└── WP4_degradome_seq/
├── Metadata/
├── Raw data (degradome seq data of Arabidopsis during seed germination)/
├── Intermediate data and syntax/
├── Final data and syntax/
└── Publications/
```

15. Describe the file naming conventions you intend to use. Please give one or multiple example(s).

The following format is used to keep the naming of the files consistent.
[date_(YYYYMMDD)]_[project(researcher)]_[description]_[version].[extension].

16. How will you distinguish between versions of files (multiple answers possible)?

- We will use Git versioning for code / scripts.

E. Data documentation and data quality

17. Describe below what data documentation and metadata will accompany the data to help make the data findable, understandable, and reproducible.

- Elaborate documentation and notes within scripts / codes.
- The WUR readme file template (see template at https://doi.org/10.5281/zenodo.7701727).

18. Describe what data and analysis quality controls will be used?

- We will perform preliminary (pilot) experiments to validate intended experimental methods.
- Supervisors or peers will review the data and results for any anomalies (e.g. unexpected inconsistencies, outliers, correct labeling of data and / or treatments, correct and consistent coding applied, etc.).
- Statistical model assumptions are adhered to and assessed (e.g. (residual) distribution analysis, outlier analysis, (accounting for) independence, homogeneity of variance, etc.).
- We will use discipline specific community standards for labelling and coding of data. Please specify the community standard used.
- We will use standardised coding and terms of data throughout all experiments so that data descriptions are equal throughout various datasets created.
- We will use standard and validated protocols where appropriate.
- We will use repeated measurements to validate results (e.g. duplicate or triplicate analysis, multiple observer agreement, measurements taken over time, etc.).

F. Working with sensitive data (personal data, ethics), data ownership, sharing and access

19. Who is the (rights)holder of the data (commonly known as the owner of the data)?

- WUR is the (rights)holder of the data.

Contact relevant staff for the management of data ownership.
Officer (Karin Booltink-Polman) and Information Security Officer (Eema Batool)

Created

20. What is the data classification for your project (for example as specified in SmartPIA) taking into account the (privacy) sensitivity of the data?

- Negligible.

21. Is this project registered in SmartPIA?

- No. Please register in SmartPIA in the case (privacy) sensitive data is collected (when applicable: via your supervisor, the project manager, see guidance).

We will also contact relevant staff if needed.
Officer (Karin Booltink-Polman) and Information Security Officer (Eema Batool)

22. Please specify the (sensitive) data and privacy protection measures. Note that any measures undertaken should be consulted with the Information Security Officer (ISO) and Privacy Officer (PO).

- We will consult with the ISO and PO for appropriate measures to undertake.

23. Are there other ethical issues that need to be taken into account which may include approval from ethical committees?

- No.

24. Will there be any intellectual property (IP) rights or alternative applications or routes to impact (such as commercial interests) associated with the data?

- No.

G. Data archiving and publishing

25. Are there reasons to restrict access to the data or limit which data will be made publicly available?

- No.

26. Describe what data from question 11 will be archived internally (e.g. WUR network drive / Yoda@WUR) and not published, for a minimum of 10 years? Include the exact name for the storage medium chosen (see the WUR Data Storage Finder).

- Not applicable as data will be published.

Local machine, WUR proprietary storage systems, WUR git and public git repositories.

27. What data will be published and made available for reuse via a data repository?

- Data underlying publications or reports. Please specify below which data listed in question 11.

Sequencing data of *Arabidopsis thaliana* generated during the analyses (fastq format), as well as the metadata (csv format). Tomo-seq sequencing data (fastq format) generated and all integration data.

28. When will the data be available for reuse, and for how long will the data be available?

- Data will be available for at least 10 years as soon as the article or report is published and not required for any other article publication.

29. Which data repository do you intend to use to make the data findable and accessible (see the WUR Repository Finder)?

- Other, please specify below.
- Zenodo.

SRA and GEO

30. Which metadata standard will be used to describe the data during internal archiving and / or depositing in a data repository?

- A discipline-specific metadata standard. Please specify below.

According to SRA standards.

31. Which licence/terms of use will be applied to the data?

- Open access (Creative Commons Attribution licence (CC BY); anyone can access and reuse with attribution).

H. Data management costs

33. What resources (in time and / or money) will be dedicated to data management, data archiving or publication, and ensuring that data is reusable? Indicate as well how these costs will be covered.

- All costs for 10 year data storage and access management to that data after journal publication or report are covered by the research group / project.