



D7.2 Germinate database dedicated to BEST-CROP

(Version 0.3, 13th May 2025)

THE JAMES HUTTON INSTITUTE (JHI) This report is public Deliverable description

DELIVERABLE:
D7.2 Germinate database dedicated to BEST-CROP
WORK PACKAGE:
WP7. Data management and biostatistics/bioinformatics support
AUTHOR(S):
Paul Shaw and Sebastian Raubach, The James Hutton Institute (JHI)
DUE DATE:
31/12/2024
ACTUAL SUBMISSION DATE:
13/05/2025
DISSEMINATION LEVEL Select the proper one and delete the remaining
[x] PU: Public (must be available on the website)
GRANT AGREEMENT No:
101082091
PROJECT STARTING DATE:
01/07/2023
PROJECT DURATION:
60 months

Quality of information - Disclaimer according to the Art. 17.3 of GA

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EUROPEAN RESEARCH EXECUTIVE AGENCY (REA). Neither the European Union nor the granting authority can be held responsible for them.

REVISION HISTORY				
Version	Date	Modified by	Comments	
0.1	09/12/2024	JHI	Initial submission	
0.2	29/12/2024	UMIL	Final revision by the coordinator	
0.3	13/05/2025	UMIL	Second submitted version:	
			Deliverable title aligned with Sygma	
0.4				
0.5				

Table of Contents

Blossary of terms	. 4
1. Introduction	5
2. Background	5
Why Germinate?	6
Why not develop from scratch?	6
What types of data will we store?	6
Where will the data be stored?	6
Other benefits	6
3. What we have implemented	7
4. What's next	7
5. Resources	7

Glossary of terms

Germinate The database that will hold most of the BEST-CROP experimental data and provide mechanisms for allowing us to make data available to both project partners and with worldwide plant community.

GitHub	A Microsoft owned online tool for the version control of computer code.		
FAIR	Findable, Accessible, Interoperable and Reusable.		
DOI	Data Object Identifier. An ID used to uniquely identify a digital resource.		
PGR	Plant Genetic Resources		
CC BY	Unrestrictive open source licence that requires attribution to BEST-CROP.		
C&E	Characterisation and evaluation data.		
UKRI	UK Research and Innovation (<u>https://www.ukri.org/</u>)		
FAO GLIS	Global Information System of the FAO		
EURISCO	European Search Catalogue for Plant Genetic Resources		
Genesys genebanks	Worldwide Plant Genetic Resources for Food and Agriculture conserved i		

1. Introduction

The aim of this document is to provide an update on how experimental data from the BEST-CROP project will be made available to both project partners and the wider scientific community. It follows on from, and builds upon, the guidelines and foundations put in place from D9.1 (Data management plan) and D7.1 (Germinate and GridScore training) and lays the foundations for contributions to to sections 1.3.2.8 Open science and 1.3.2.9 Data Management in the project proposal, Tasks 7.1 and 7.2 in WP7, deliverables D7.1, D7.2 and D7.3 and finally milestone M7.1. The solutions we put in place under this deliverable will ensure that experimental data generated under BEST-CROP can be efficiently uploaded into a single public facing repository based on Germinate (Shaw et al. 2017; Raubach et al. 2021). Germinate ensures that data from BEST-CROP will be made available through a single interface suitable for online exploration of data and allow data to be exported in formats suitable for downstream analysis. All data that the BEST-CROP Germinate database holds will be available in uncompiled, standard text based formats as per our grant agreement and data management plan (D9.1). Work we undertake on Germinate under BEST-CROP will not only benefit the project, but also the Germinate ecosystem with any developments released back to the community under the unrestrictive open source Apache 2.0 licence that Germinate uses. As per our grant agreement data will be made available through Germinate under a permissive CC BY licence (although some may be embargoed for a period of time to allow publication or IP protection) using DOI's where appropriate and ensuring the long-term availability of data past the end of the project.

2. Background

There is a need to make sure that any experimental data generated under BEST-CROP is made available to project partners and the community more widely and in ways that promotes the use and reuse of data resources that BEST-CROP is generating. The use of a standardised data repository for BEST-CROP is therefore critically important for reasons including:

Transparency and Accountability: Public access allows for scrutiny and verification of the project research findings, ensuring transparency and holding project partners accountable for their work. It also ensures that any interested parties can interact with the data and contact the original data providers if required.

Archival of Research Data: BEST-CROP data made available through Germinate will be maintained for a period of at least 3 years past the end date of the project ensuring access well into the future.

Reproducibility: Other researchers can replicate experiments and analyse data if required along with having access to contact details of the original data provider and any code that was used to analyse data.

Community Collaboration: Sharing data enables community collaboration and the building upon existing research, potentially leading to the synthesis of new knowledge.

Public Trust: Openness nurtures trust in the scientific process and the institutions that conduct research. This is particularly important with BEST-CROP where there are gene editing experiments being undertaken. Being open and up front about the gene editing data that is being generated is an opportunity for us to engage positively with the public and other researchers.

Ensuring BEST-CROP data is as FAIR as possible: Ensuring that data from the BEST-CROP project is made available in ways that are as FAIR as possible will promote the future use of data that the project has generated.

Why Germinate?

To meet project requirements a centralised database which will allow us to store, distribute and curate experimental data from the BEST-CROP project is required Germinate is and has been used in a number of other PGR type projects and provides an 'almost' off the shelf solution to meet BEST-CROPs requirements. Any additional functionality that this project needs will be implemented using Germinate then made freely available back to the plant sciences community. This is one of the core benefits of using Germinate and allows BEST-CROP to contribute to, and benefit from the development of Germinate across all the projects that use it.

The BEST-CROP Germinate database will provide both project partners and the wider scientific community with tools to allow them to interact with the experimental data generated in BEST-CROP through a single web-based interface.

Why not develop from scratch?

Leveraging an existing data platform like Germinate offers significant advantages over developing one from scratch. Firstly, it significantly reduces development time and costs. Building a data platform from the ground up requires substantial investment in infrastructure, personnel, and software development and is often outwith the scope of many projects. In contrast, drawing on an established platform provides pre-built components, tools, and functionalities, which cover most, albeit potentially not all, of the required functionality a project needs. Secondly, Germinate benefits from continuous updates and enhancements funded by the different projects that use it. This allows it to keep pace with evolving technological standards and industry best practices while adding new features which benefit the community as a whole.

Finally, existing data platforms often have a large and active user community, providing access to valuable knowledge, support, and best practices. This community-driven approach fosters collaboration, problem-solving, and knowledge sharing, accelerating learning and innovation.

Building a platform from scratch limits access to such a supportive ecosystem, potentially hindering progress and increasing the risk of errors.

What types of data will we store?

Germinate will store, plant passport, C&E, genotypic, image, phenotypic and field trials datasets generated under BEST-CROP. We will hold sequence data in appropriate public repositories and link to them where appropriate from Germinate. We will work to include new database features as these become required by the project.

Where will the data be stored?

The data held in Germinate will be hosted at the Hutton Institute in Invergowrie, Scotland. The data will be backed up by the backup plan detailed in D9.1 DMP which includes daily backups both onsite and at Huttons Aberdeen site. Backups and archiving are as per UKRI guidelines (10 year archives of data).

Other benefits

Germinate will also allow us to link the germplasm used in BEST-CROP to the European (EURISCO) and worldwide (Genesys) germplasm catalogues where appropriate. These systems (including Germinate) all contribute towards the FAO GLIS system.

Germinate is also BrAPI (Selby et al. 2019) compliant allowing other BrAPI tools to interact with the data that our database contains, opening up the possibility of linking other tools with the BEST-CROP database and allowing us to offer this as a core functionality of the resource.

3. What we have implemented

Germinate has been deployed on Hutton servers and can be accessed from <u>https://germinate.hutton.ac.uk/best-crop</u>. Initial background data in plant lines that will be used in BEST-CROP is being gathered and will be uploaded to this resource by the end of January 2025. This will then be added to as new data sets become available. Logos and links to all BEST-CROP project partners have been added along with recognition of EU and UKRI funding for this project. We have run training courses for Germinate data templates but we intend to offer these again now that the resource is up and running along with introduction to using the Germinate interface training for BEST-CROP partners. We will provide recordings of this training as well as access to any printed materials from the BEST-CROP website (<u>https://www.bestcrop.eu</u>).

Part of the work we will do over the remaining BEST-CROP funding is to look at how our partners are interacting with the data and directly liaise with them to create a feature priority list that will be used to direct the development of BEST-CROP specific Germinate tools. We will do this with our previously created 'Data Contact' group which includes a representative from each BEST-CROP partner and will run this after we have provided introductory training on using the Germinate interface to give partners time to look at the system with some BEST-CROP specific data. We will use the user group to prioritise development based on needs and publish this priority list on the BEST-CROP GitHub page (<u>https://github.com/best-crop</u>) to ensure openness and encourage discussion. We will also highlight this via our own Germinate social media channels to raise awareness of the development that BEST-CROP is undertaking.

4. What's next

We have started uploading basic background information on the barley varieties and lines that are being used in BEST-CROP. This should be completed by the end of January 2025 and form the basis for inclusion of all BEST-CROP experimental data going forward. Once this is done we will start promoting the resource to the wider community through social media channels, highlighting that data is sparse, but will increase over the duration of the project. Feedback from our data group will come after we run introduction to Germinate training (February/March 2025) and form the basis for future BEST-CROP Germinate development. Details will be published on our GitHub page (March 2025) for comment from the BEST-CROP consortium.

5. Resources

- The main BEST-CROP data repository has been set up and is available from https://germinate.hutton.ac.uk/best-crop. This will be added to over time as new data becomes available.
- Links to this resource are included on the main BEST-CROP webpages <u>https://www.bestcrop.eu/GridScore&Germinate.html</u>.

- Germinate itself is open source and all code and training materials available from https://germinateplatform.github.io/get-germinate.
- All BEST-CROP code will be available from <u>https://github.com/best-crop</u> which will be added to as new code is developed (including data analysis scripts).

Raubach, Sebastian, Benjamin Kilian, Kate Dreher, Ahmed Amri, Filippo M. Bassi, Ousmane Boukar, Douglas Cook, et al. 2021. "From Bits to Bites: Advancement of the Germinate Platform to Support Prebreeding Informatics for Crop Wild Relatives." *Crop Science* 61 (3): 1538–66. https://doi.org/10.1002/csc2.20248.

Selby, Peter, Rafael Abbeloos, Jan Erik Backlund, Martin Basterrechea Salido, Guillaume Bauchet, Omar E. Benites-Alfaro, Clay Birkett, et al. 2019. "BrAPI—an Application Programming Interface for Plant Breeding Applications." *Bioinformatics* 35 (20): 4147–55.

https://scholar.google.co.uk/citations?view_op=view_citation&hl=en&user=kwK7rKsA AAAJ&citation_for_view=kwK7rKsAAAAJ:RHpTSmoSYBkC.

 Shaw, Paul D., Sebastian Raubach, Sarah J. Hearne, Kate Dreher, Glenn Bryan, Gaynor McKenzie, Iain Milne, Gordon Stephen, and David F. Marshall. 2017. "Germinate 3: Development of a Common Platform to Support the Distribution of Experimental Data on Crop Wild Relatives." Crop Science 57 (3): 1259–73. https://doi.org/10.2135/cropsci2016.09.0814.