D7.1 Data templates and training materials released

(Version 0.1, 17th June 2024)

THE JAMES HUTTON INSTITUTE (JHI)

DELIVERABLE:
D7.1 Data templates and training materials released
WORK PACKAGE:
WP7. Data management and biostatistics/bioinformatics support
AUTHOR(S):
Paul Shaw and Sebastian Raubach, The James Hutton Institute (JHI)
DUE DATE:
17/06/2024
ACTUAL SUBMISSION DATE:
17/06/2024
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	17/06/2024	JHI	Initial submission
0.2			
0.3			
0.4			
0.5			

Table of Contents

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.

GridScore   A software tool designed for experimental data collection in the field and which can be used by BEST-CROP partners to help them collect experimental data.

GitHub   A Microsoft owned online tool for the version control of computer code.

1. Introduction

The aim of this document is to provide guidelines on the use of Germinate data templates for the upload of data to the BEST-CROP Germinate database which will be used to ensure that data from BEST-CROP is made available publicly and to project partners in formats suitable for online exploration of data and downstream analysis.

2. Background

The use of standardised data upload templates is important for several reasons including:

Consistency: Templates ensure project partners uploading data uses the same format and includes the same fields/information as well as helping to improve standardisation of metadata. This consistency makes data easier to understand, compare, and perform analysis on the underlying data by downstream analysis applications.

Error Mitigation: Standardised templates can help in the elimination of errors caused by mistakes in formatting or missing data entries. By specifying required fields, data types and in some circumstances permitted values (such as lists of valid phenotypes for a trait) or acceptable value ranges (such as plant height being above a lower limit and below an upper limit), templates can help mitigate against nonsensical entries or irregularities.

Efficiency: Standardised templates can help save time and effort during the data upload process. For example, ensuring that our project partners are aware that data is required to be in a specific format gives them a template to work towards while the data upload and check features that Germinate offers will allow them to check their datasets for common issues before data is submitted to the databases. This helps us try to ensure data uploads become faster, more streamlined and as error free as possible.

Lastly, but most importantly the use of standardised data templates helps us to improve - **Data Quality:** By ensuring consistency and reducing errors, templates can help lead to overall better-quality data. This high-quality data is essential for accurate reporting, analysis, and decision-making downstream.

Data templates are all held under version control using GitHub so that changes can be efficiently tracked and accounted for.

In addition to the development of standard data templates we have also run online Germinate data template and GridScore training and made these videos available through YouTube to allow BEST-CROP (and the wider community) access to materials online, when it suits them.

3. Resources

- BEST-CROP data resources can be found on our BEST-CROP GitHub page https://github.com/best-crop
- Germinate Templates can be found on the main BEST-CROP GitHub page or directly on the main Germinate page https://github.com/germinateplatform then https://github.com/germinateplatform/germinate-data-templates.
- Germinate training that can be used by BEST-CROP partners (and will be updated throughout the project duration) can be found here https://germinateplatform.github.io/germinate-training.
- 'Introduction to Germinate Data Templates' which was undertaken in April 2024 for BEST-CROP partners is available from https://youtu.be/OwzDf4kxKH0

- 'Introduction to GridScore' training which was undertaken in April 2024 for BEST-CROP partners is available from https://youtu.be/Nq6rKtTWDBU

Most data will start coming soon and we will keep the training up to date and run additional sessions with groups once we start receiving data from them.

All resources are available through the main BEST-CROP GitHub page https://github.com/best-crop and links to select resources have also been made available through the BEST-CROP website in a dedicated 'GridScore and Germinate' page https://www.bestcrop.eu/GridScore&Germinate.html.