From Bits to Bites

Software development for genetics, genetic resources and pre-breeding

29th October 2026 Crop Trust GROW Webinar

Paul Shaw
Department of Information and Computational Sciences

The James Hutton Institute

INTERNATIONAL BARLEY HUB

paul.shaw@hutton.ac.uk





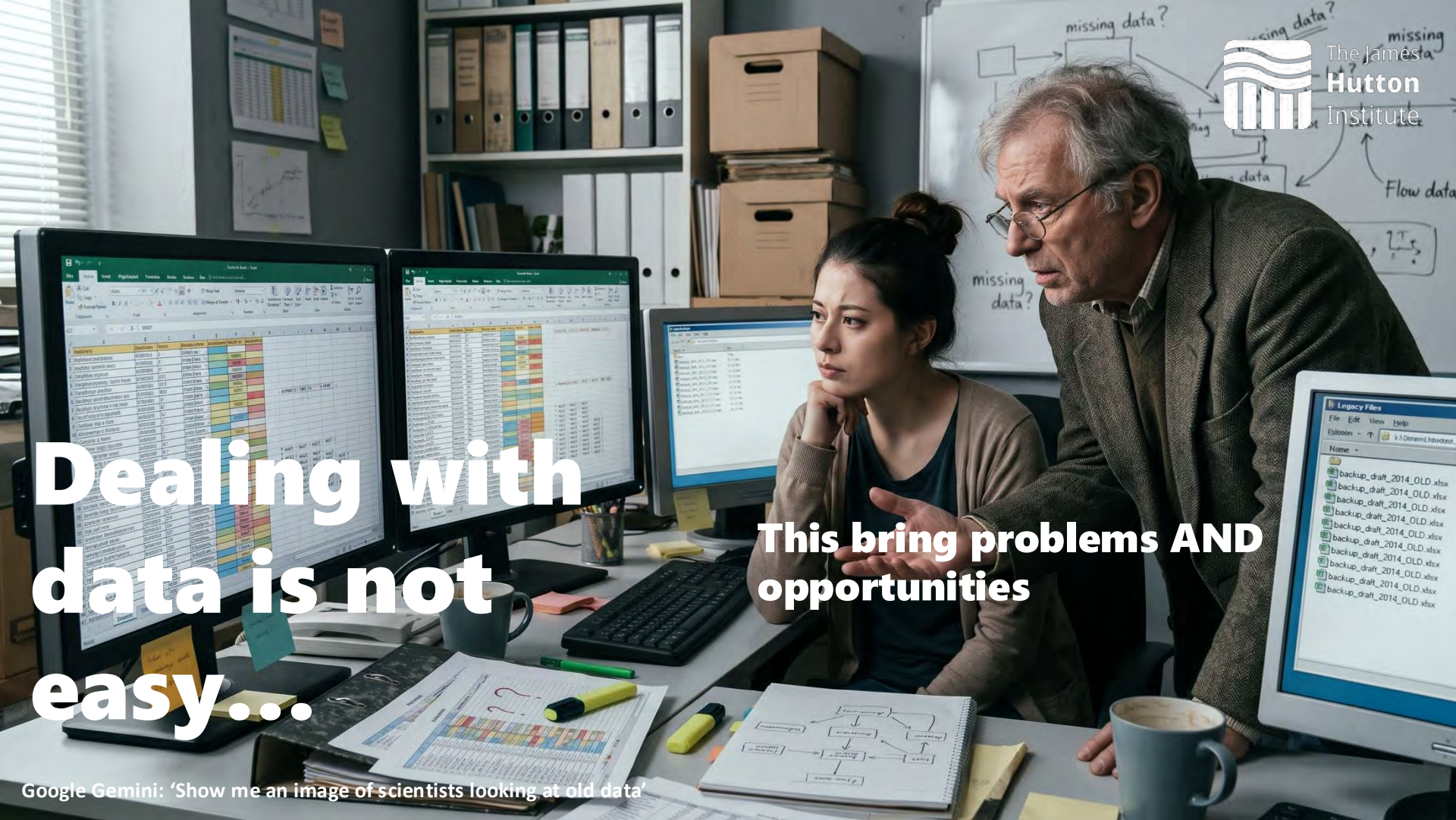




The James Hutton Institute



Google Gemini: 'Show me an image of scientists in a barley field working with data'



Dealing with data is not easy…

This bring problems AND opportunities

Google Gemini: ‘Show me an image of scientists looking at old data’

BIG DATA

The James Hutton Institute

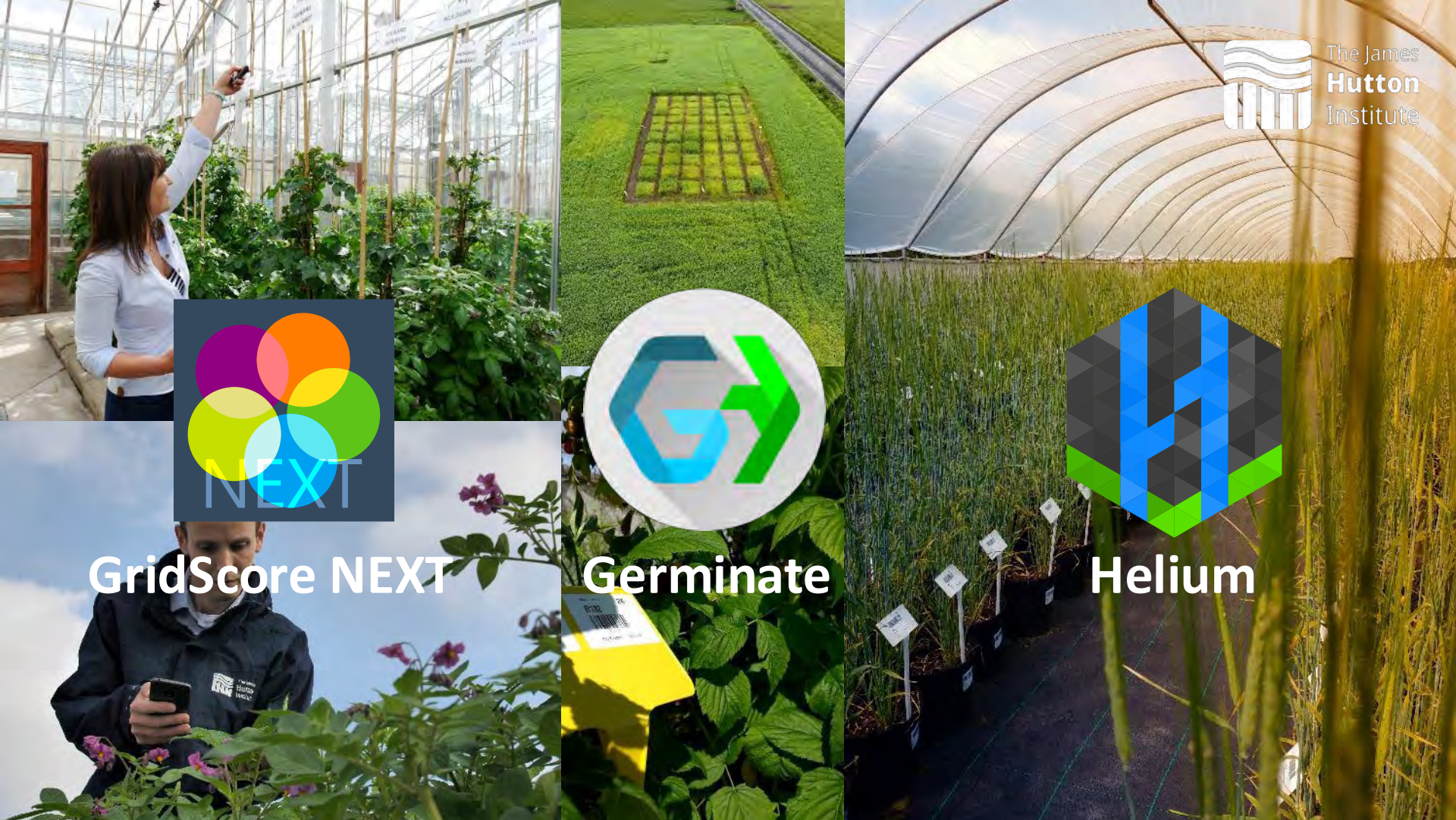
IS NORMAL DATA...

"THE NEXT AGRICULTURAL REVOLUTION WON'T BE SOWN IN THE FIELD - IT WILL BE COMPILED IN THE CLOUD."

The James Hutton Institute

BRINGING STRUCTURE TO CHAOS

IF YOUR DATA LIVES IN A FILE NAMED FINAL_VERSION_V3_ACTUALLY_FINAL.XLSX, THIS TALK IS FOR YOU ☺


The James
Hutton
Institute
NEXT
GridScore NEXT
Germinate
Helium

Why?

- Size and complexity of data is challenging.
- Allowing people to collect, store, explore & interact with data.
- Aid understanding.
- Show relationships and patterns.
- Provide easy access in digestible chunks.
- Availability of information (FAIR etc.).
- Reduce errors.

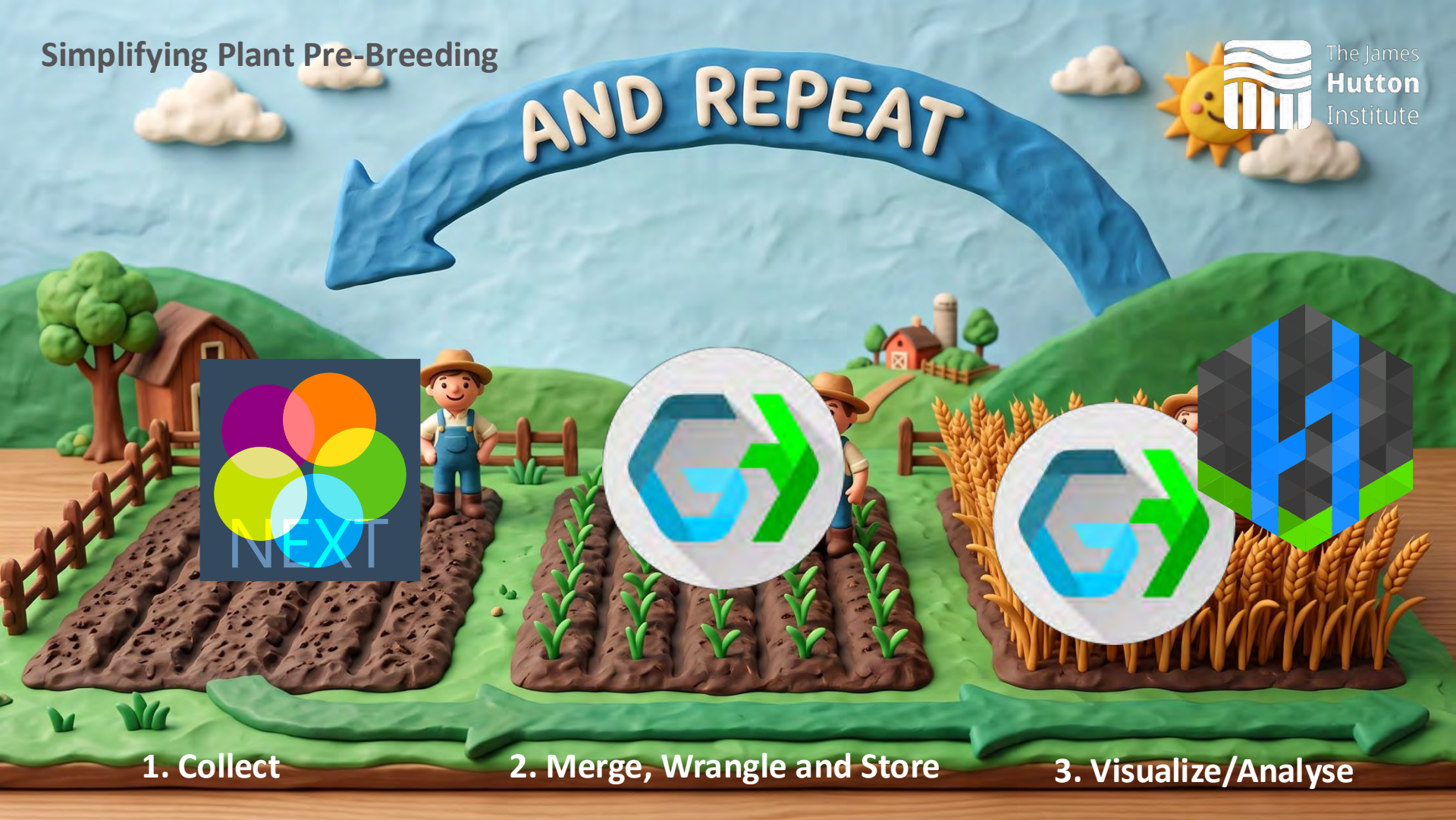


Why are we different?

- Long term software development -GridScore 2020, Germinate 2002, Helium in 2010.
- Everything you see here is free for you to use with no restrictions.
- We have had wide ranging funders for these tools over the years and that's important!
- All our code is open source and on GitHub.
- We are always as open as we can be.
- Tools developed in Scotland primarily for Hutton's needs but with worldwide impact.



Simplifying Plant Pre-Breeding



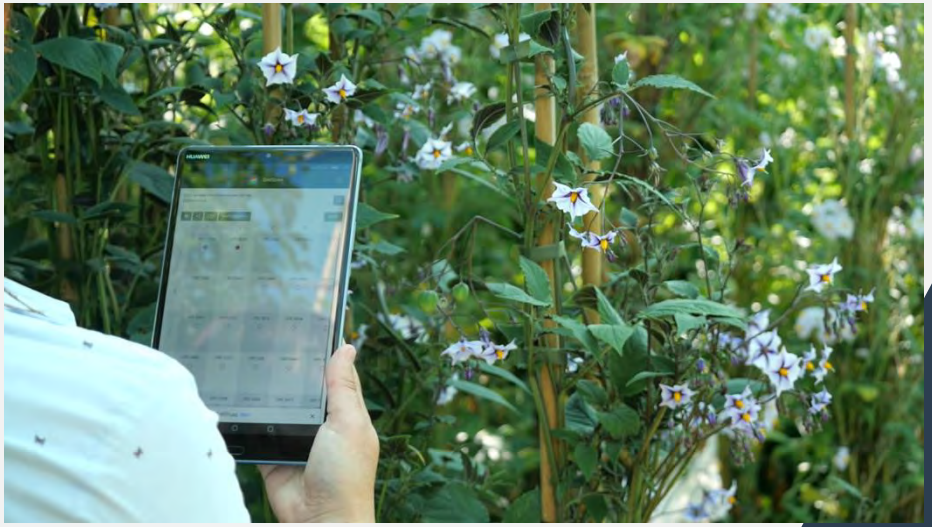
The James Hutton Institute

GridScore

Mobile phenotypic data collection

gridscore.hutton.ac.uk

The James Hutton Institute



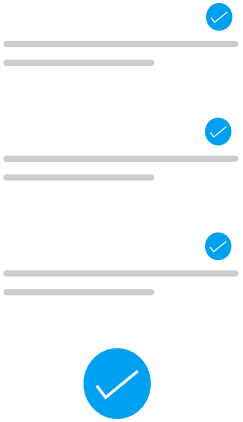


"FIELD WORK BUILDS CHARACTER. SOFTWARE KEEPS THAT CHARACTER FROM LOSING THREE YEARS OF PHENOTYPING DATA."

Raubach, S., Schreiber, M. & Shaw, P.D. GridScore: a tool for accurate, cross-platform phenotypic data collection and visualization. *BMC Bioinformatics* 23, 214 (2022). https://doi.org/10.1186/s12859-022-04755-2

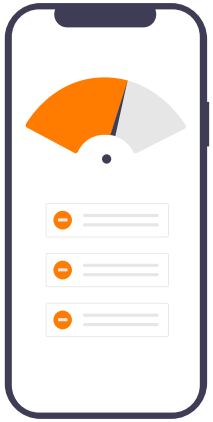
Raubach, S., Schreiber, M., Hamilton, R., McKenzie, G., McCallum, S., Kilian, B., Humphries, A., Nguyen, L.H., Quang, T.H., Singh, A., Sharma, S., Trinder, S., Feser, M., Shaw, P.D., 2026. Beyond the clipboard: data collection with GridScore NEXT. BMC Bioinformatics. https://doi.org/10.1186/s12859-025-06352-5

Benefits



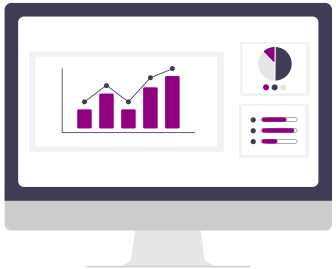
Data verification

Reduce errors by using automatic **data validation** coupled with interactive live visualizations of data.



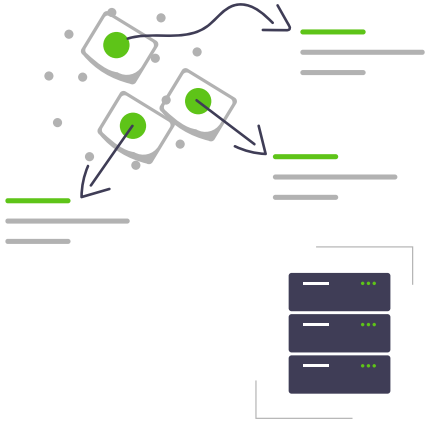
Better efficiency

Faster data collection, removal of hand-written notes, automatic **image tagging**, **GPS** tracking.



Data visualization

Spot outliers or interesting patterns in the data. These may **indicate errors** or **meaningful clustering** of data.



Data integration

Load germplasm and trait identifiers straight from central database. Everyone using the **same definitions**. Makes data **comparable/compatible**.



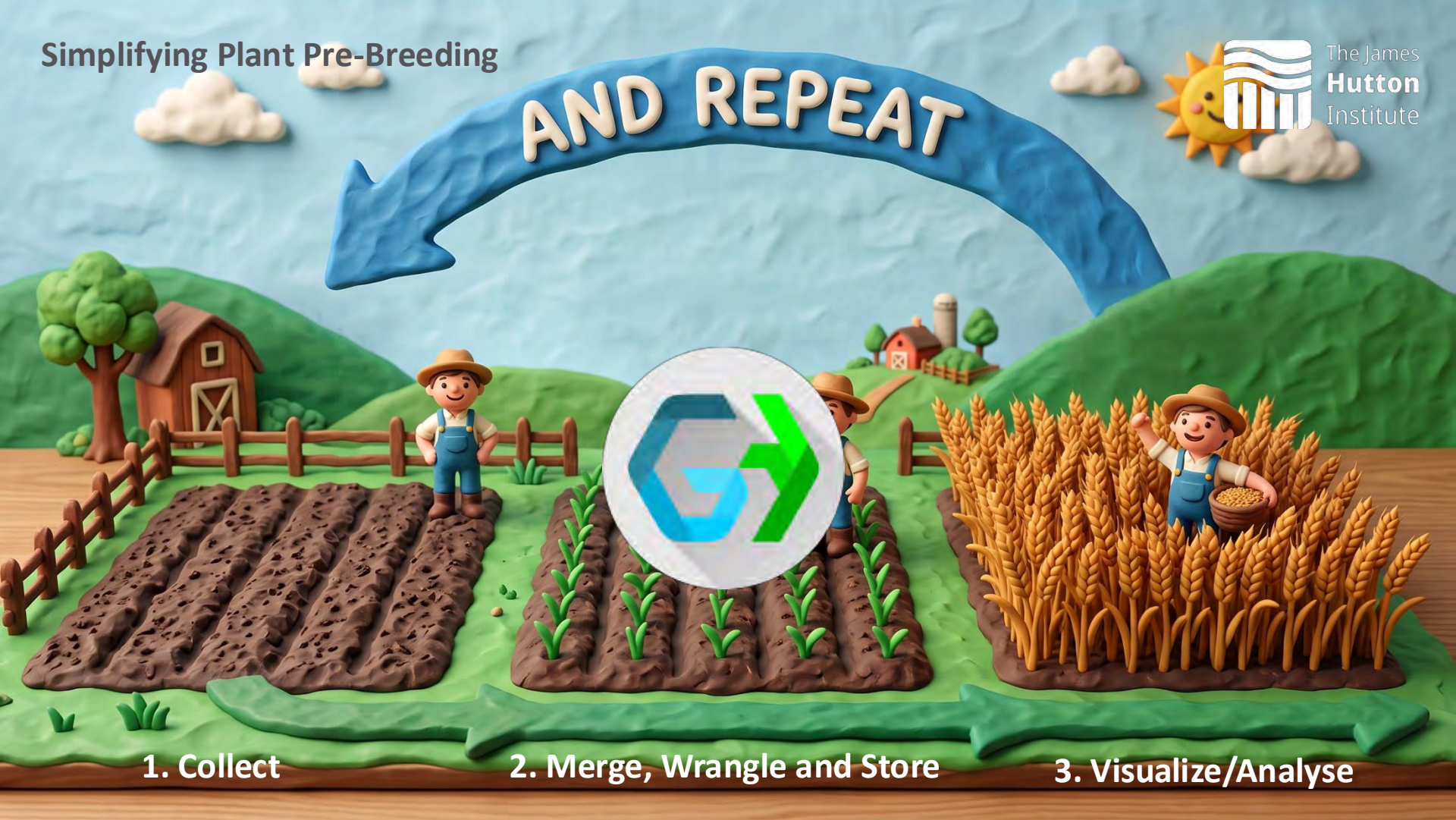




The James Hutton Institute


TRUST...

Simplifying Plant Pre-Breeding



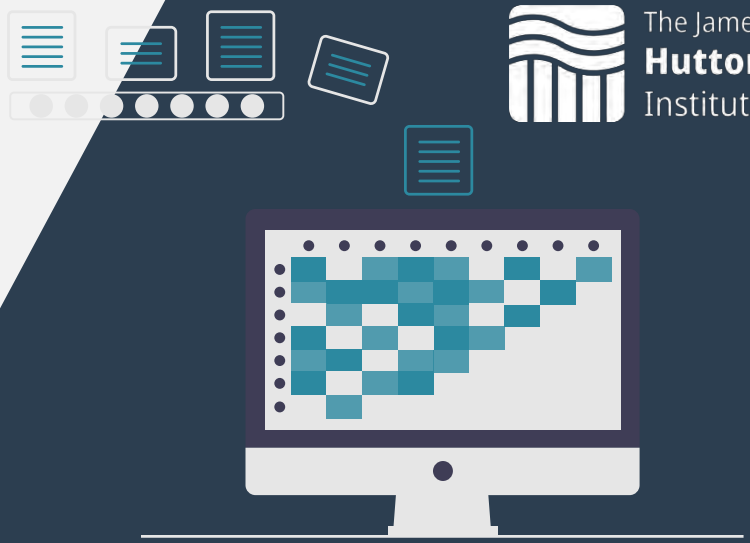
The James Hutton Institute

Germinate

Plant genetic resources database

germinate.hutton.ac.uk



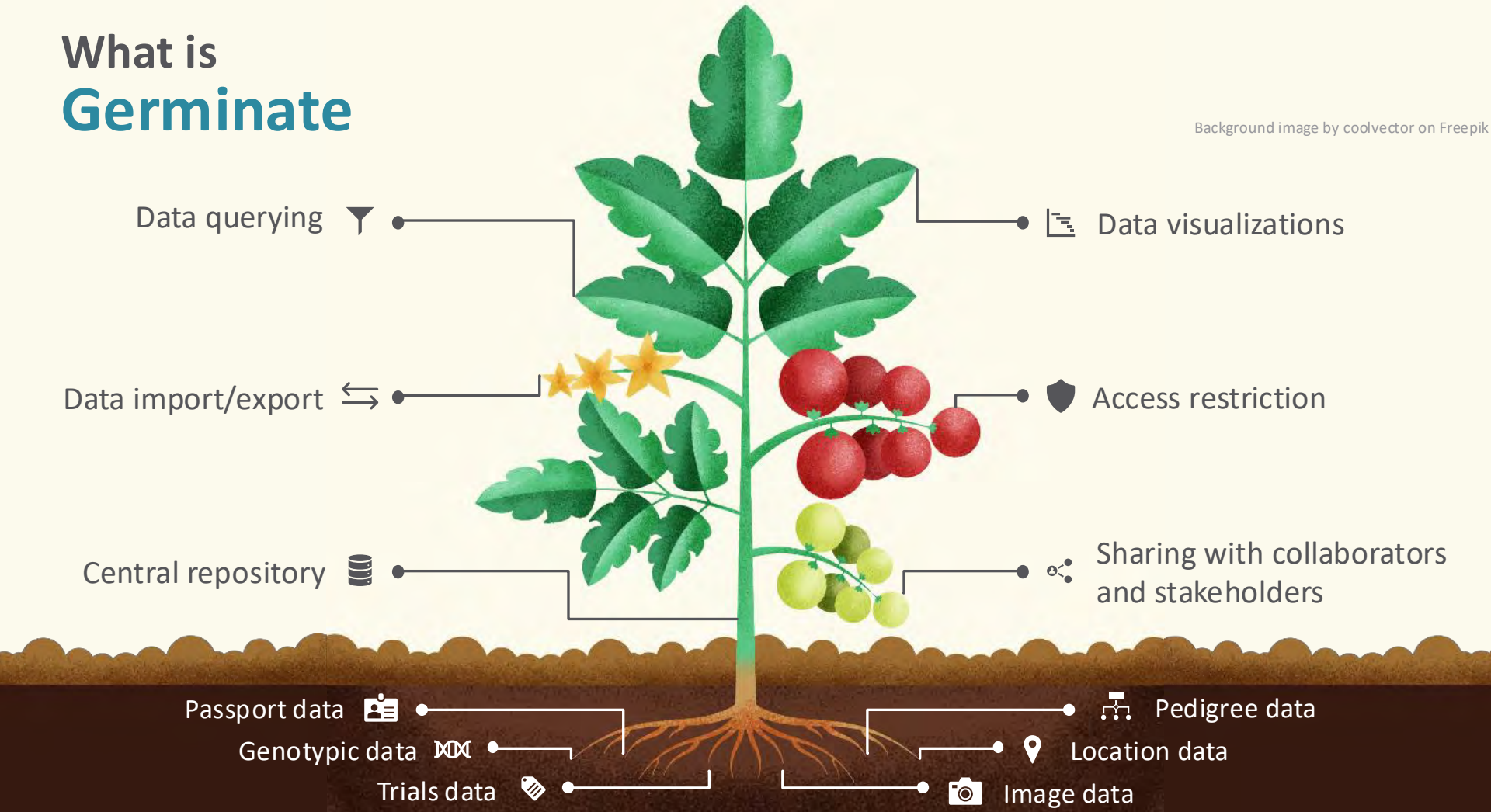


Raubach, S., Kilian, B., Dreher, K., Amri, A., Bassi, F. M., Boukar, O., Cook, D., Cruickshank, A., Fatokun, C., el Haddad, N., Humphries, A., Jordan, D., Kehel, Z., Kumar, S., Labarosa, S. J., Nguyen, L. H., Mace, E., McCouch, S., McNally, K., … Shaw, P. D. (2021). From bits to bites: Advancement of the Germinate platform to support prebreeding informatics for crop wild relatives. *Crop Science*, *61*(3), 1538–1566. https://doi.org/10.1002/csc2.20248

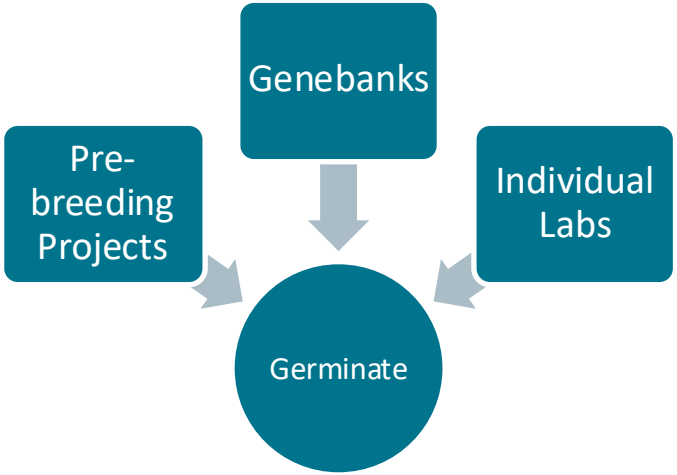
Shaw, Paul D., Sebastian Raubach, Sarah J. Hearne, Kate Dreher, Glenn Bryan, Gaynor McKenzie, Iain Milne, Gordon Stephen, and David F. Marshall. 2017. “Germinate 3: Development of a Common Platform to Support the Distribution of Experimental Data on Crop Wild Relatives.” Crop Science 57 (3): 1259–73.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

What is Germinate

Background image by coolvector on Freepik



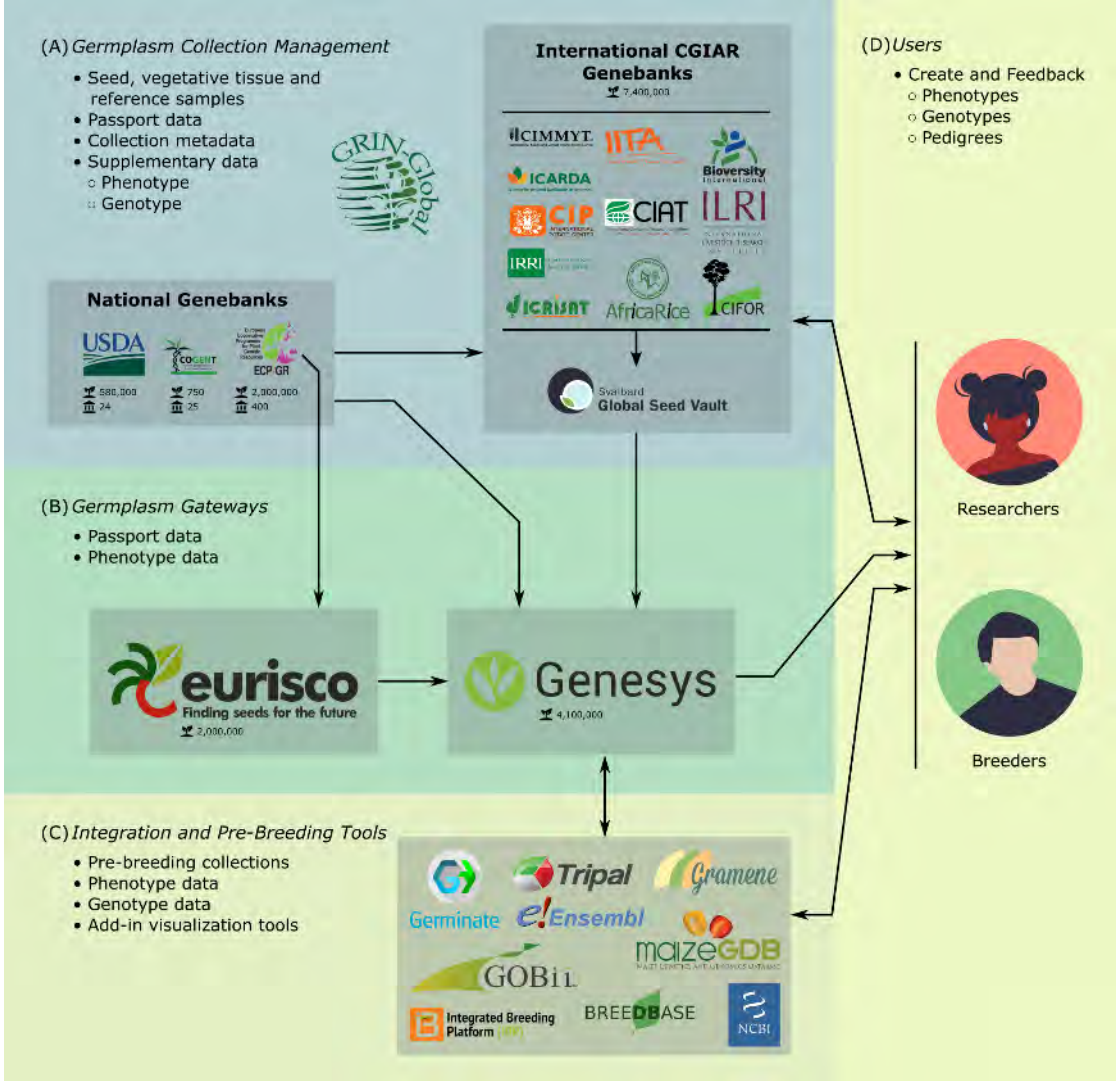
FAO GLIS Positioning

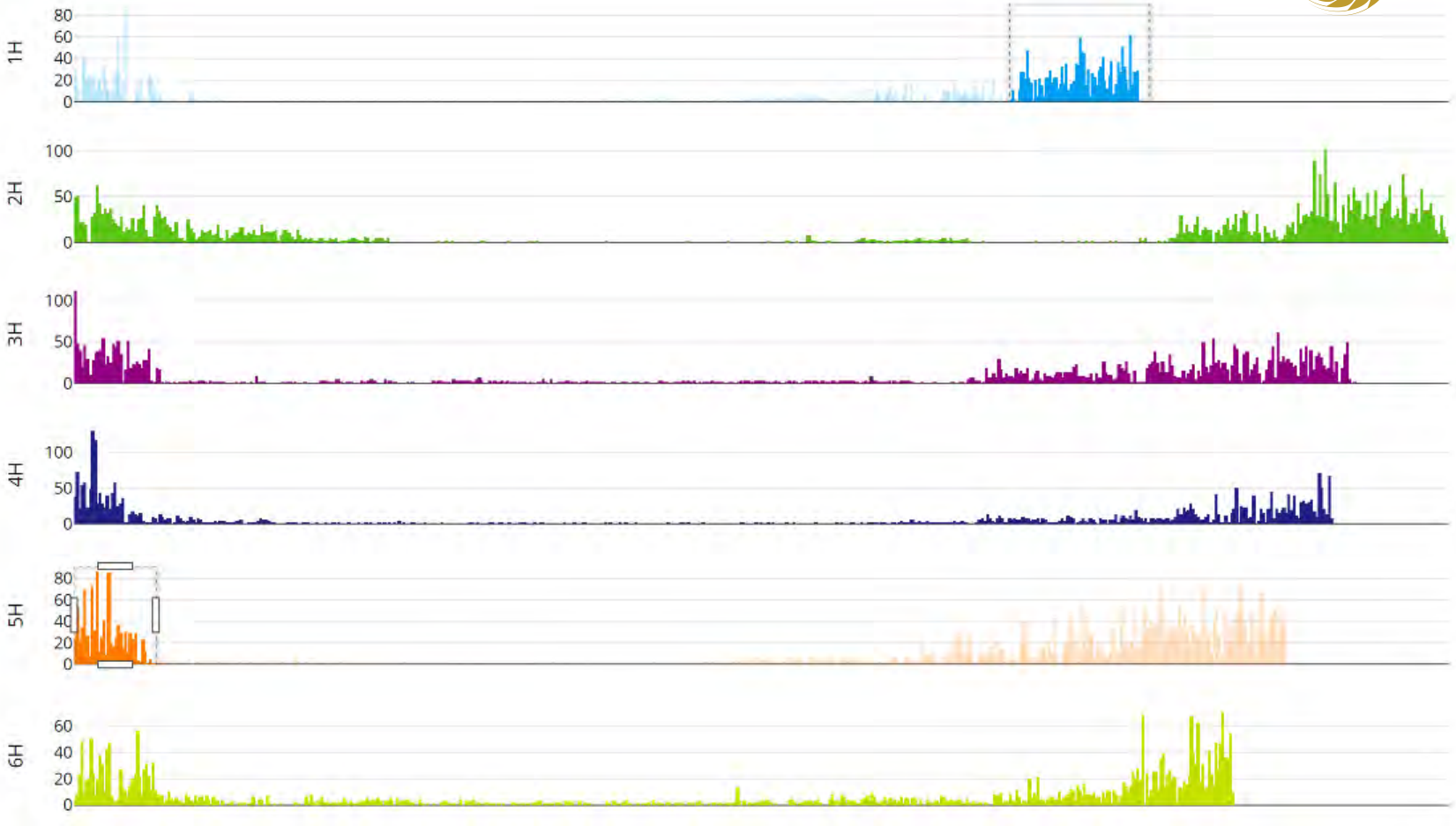


Shallow v Deep Data

Wide and shallow – Narrow and deep

Shaw, Paul D., Stephan Weise, Matija Obreza, Sebastian Raubach, Susan McCouch, Benjamin Kilian, and Peter Werner. 2022. "Database Solutions for Genebanks and Germplasm Collections." In *Plant Genetic Resources for the 21st Century*, 285–309. New York: Apple Academic Press. https://doi.org/10.1201/9781003302957-19.




1H
80
60
40
20
0
2H
100
50
0
3H
100
50
0
4H
100
50
0
5H
80
60
40
20
0
6H
60
40
20
0

MAIN NAVIGATION
Home
Data
Germplasm 2.4K
Trials data
Traits 51
Trials export 33
Geographic data
Climate data
Cross-data comparison
Datasets 35
Experiments 8
Data statistics
Publications 5
Groups 12
Search
About

Germplasm trait performance

This section shows the performance of the selected germplasm compared to all other germplasm in the database per trait. For this, the average value for this germplasm per trait is shown against the minimum and maximum across all the data.

Toggle

Blast score (1-9) | Dead plants | Digital biomass [mm²] | Height [mm] | Leaf angle [°] | Leaf area (projected) ...
Leaf area [mm²] | Leaf area index [mm²... | Leaf inclination [mm... | Light penetration de... | Resistance % | Resistant %
Resistant plants | Susceptible % | Total plants

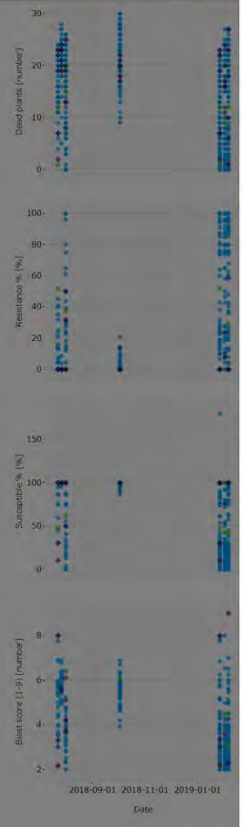
Datasets

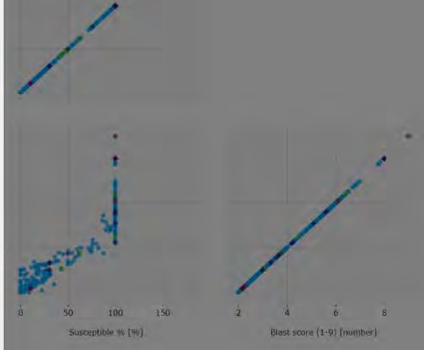
A germplasm can have many types of data associated with it. This table shows datasets of all types that contain this germplasm.

Use the table filtering to search for specific items. Showing 1 to 10 of 16 records 10

Id	Dataset name	Dataset description	Experiment name	Dataset type	Data type	License	Contact	Country	Locations	Start date	End date	ob
1	2018 Blast Screening Pg45 155 Lines 4 Populations	Using two wild Pennisetum violaceum accessions and four cultivated pearl...	Crop Trust CWR Blast Screening Data	Trials		CC-BY-NC	Dr. Shivali Sharma, Theme Leader, Pre-breeding, ICRISAT. Tel: +91-40 3071 3570. Email: shivalipbg@gmail.com	IN	1			
2	2018 Blast Screening Pg45 Drought and Heat Selected	Using two wild Pennisetum violaceum accessions and four cultivated pearl...	Crop Trust CWR Blast Screening Data	Trials		CC-BY-NC	Dr. Shivali Sharma, Theme Leader, Pre-breeding, ICRISAT. Tel: +91-40 3071 3570. Email: shivalipbg@gmail.com	IN	1			
3	2018 Blast Screening Pg138 155 Lines 4 Populations	Using two wild Pennisetum violaceum accessions	Crop Trust CWR Blast Screening Data	Trials		CC-BY-NC	Dr. Shivali Sharma, Theme Leader, Pre-breeding, ICRISAT. Tel: +91-40 3071 3570. Email:	IN	1			

Close

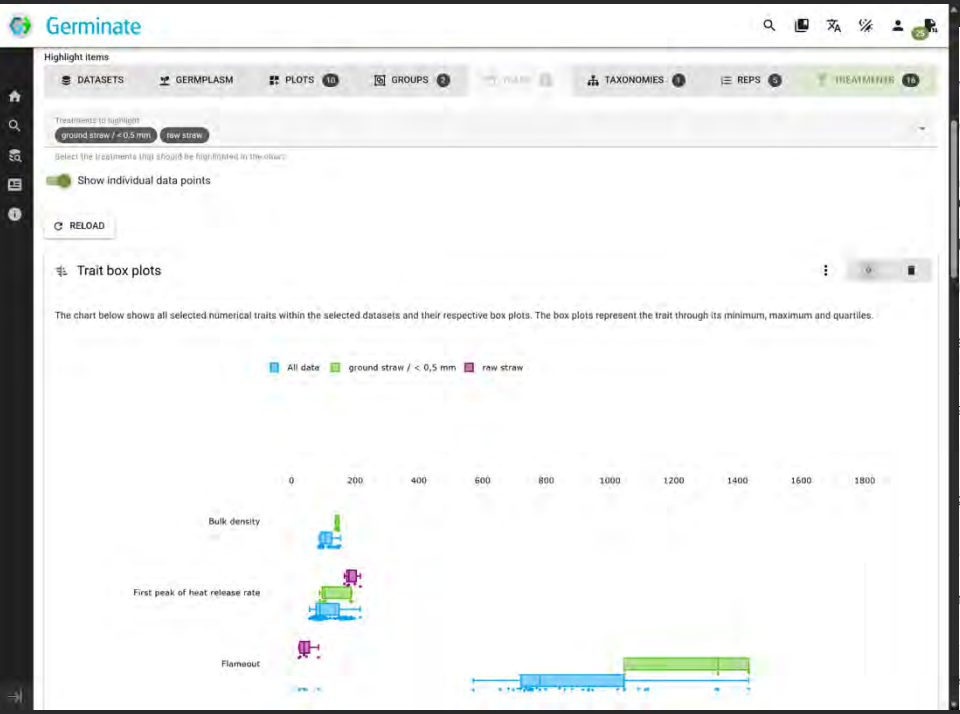


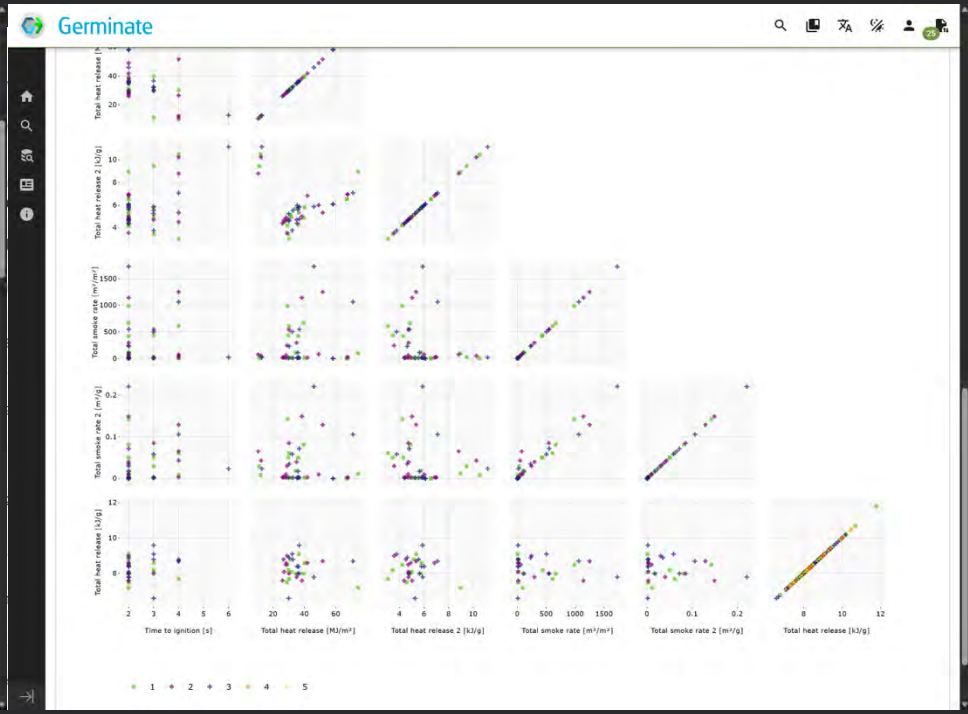


Introgression line | Wild pare

The New Germinate!

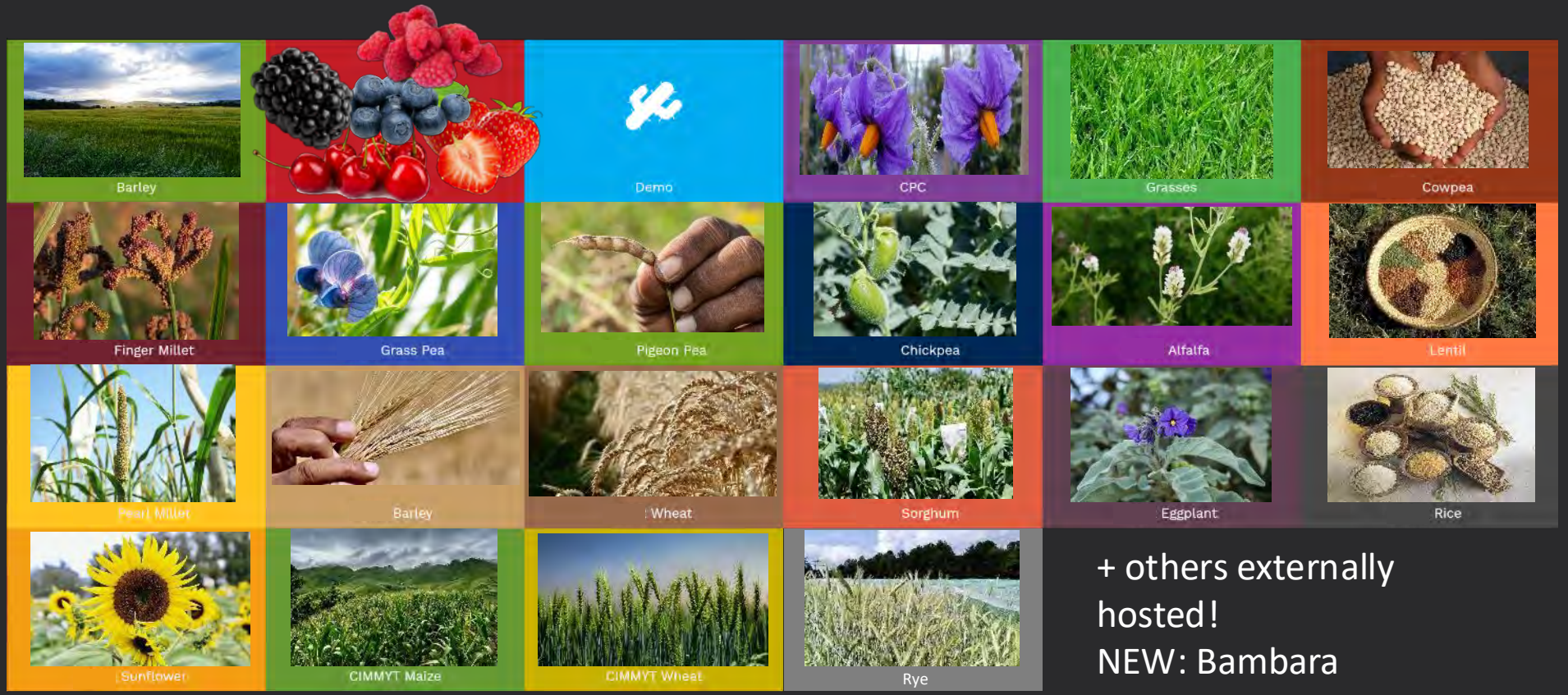
The James Hutton Institute





germinate.hutton.ac.uk

Germinate Species Diversity

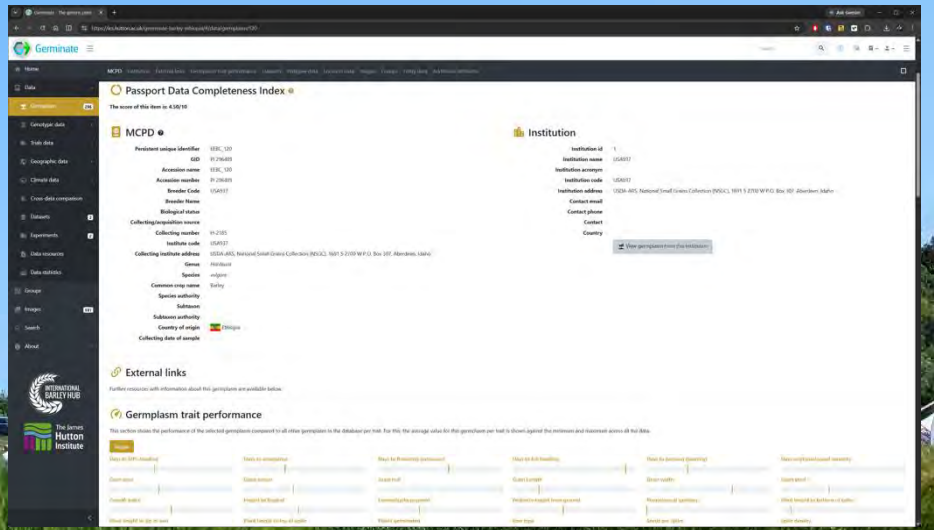


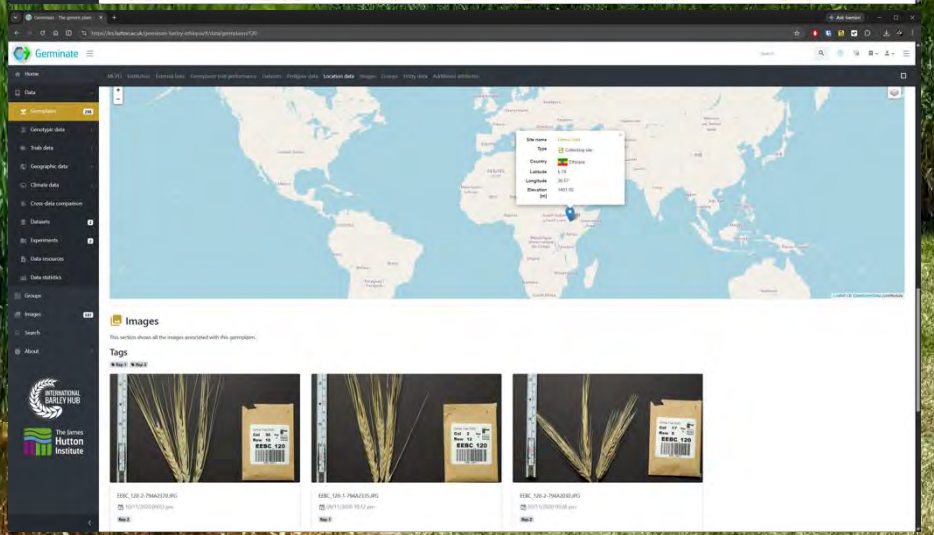
+ others externally hosted!

NEW: Bambara

Making Data Available







1030	E.ARZE 17KM ALONG ROAD FROM ANZALDO TO MOLLEPUJRO JUST ABOVE...	Collecting site	-17.92	-65.92	2600.00	BO
1031	E.ARZE 18KM ALONG ROAD FROM ANZALDO TO MOLLEPUJRO, ABOUT 1KM...	Collecting site	-17.92	-65.92	2550.00	BO
1032	VALLE GRANDE 28KM FROM VALLE GRANDE ON ROAD TO PUCQUIRA...	Collecting site	-18.63	-63.85	2850.00	BO
1033	VALLE GRANDE 28KM FROM VALLE GRANDE BY SIDE OF ROAD..	Collecting site	-18.63	-63.85	2850.00	BO

1 of 60 1 2 3 4 ...

Maps

Filtering the table will adjust the maps below. You can use this to, e.g., only look at trial sites or only locations from a specific country.

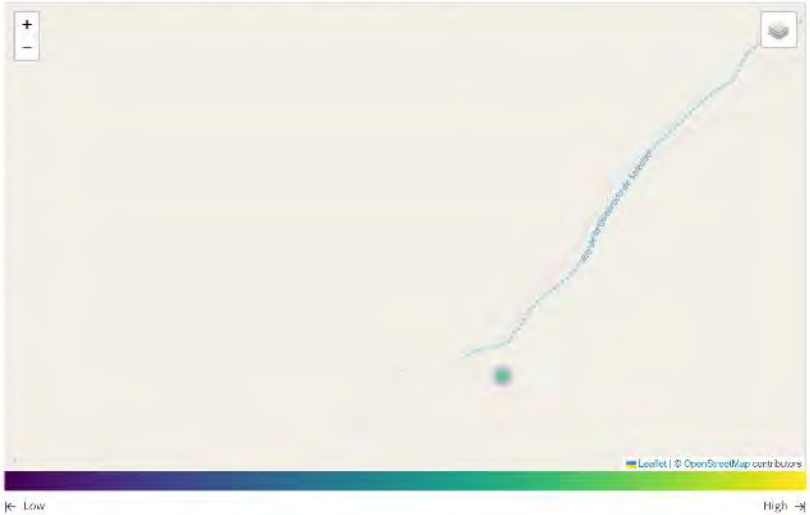
Clustered locations

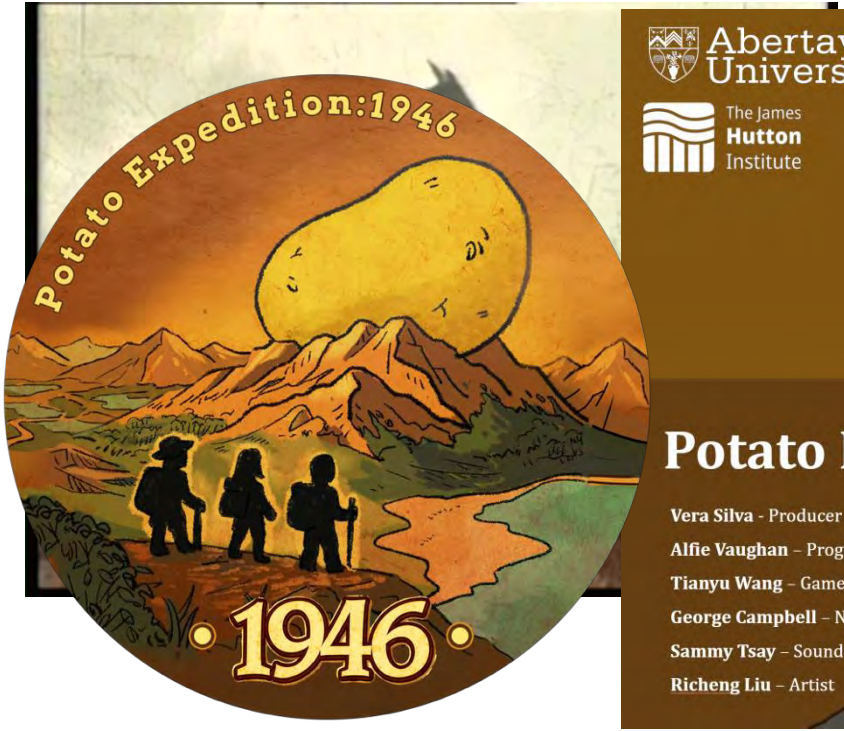
This map shows the locations clustered into groups. The numbers on the cluster nodes as well as their size and color represent the number of clustered locations. Clicking on a cluster node will expand this cluster and zoom in.

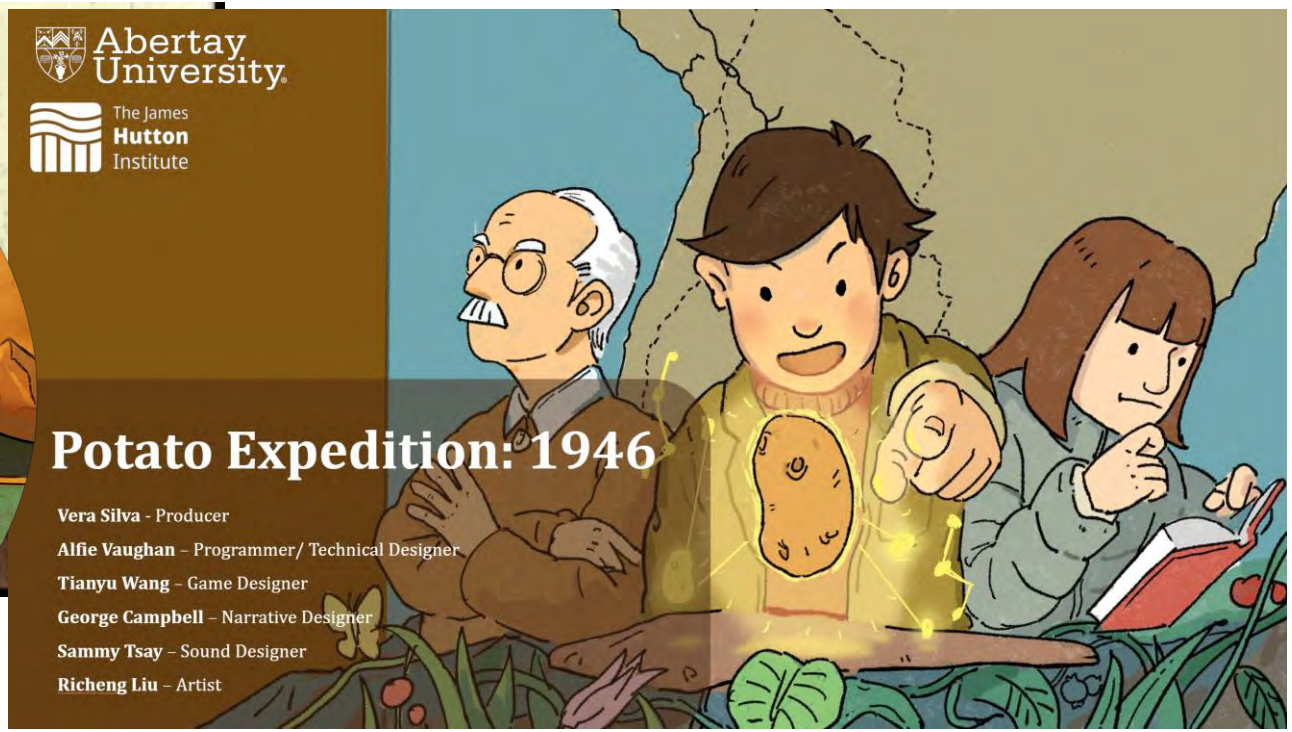


Heatmapped locations

The map below represents the density of locations with colors on a gradient. The gradient ranges from a low density (represented by the color white) to a high density (represented by the color blue). Zooming in will cause the density to dissipate.







National Potato Innovation Centre

CPC Commonwealth Potato Collection

Abertay University

Gaynor McKenzie
Dr Andrew Reid

https://team-root-games.itch.io/potato-expedition

Simplifying Plant Pre-Breeding



Helium and Neon

Pedigree tools

helium.hutton.ac.uk



A/B//C/D

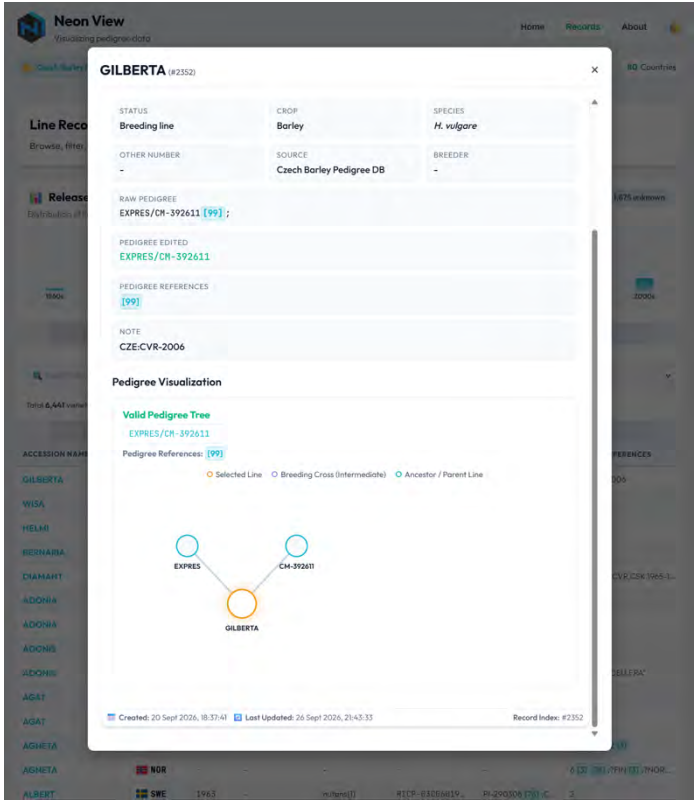
((A * B) * C) * D

((A x B) x C) x D

[A x {[(B x C) x D] x E}] x [F x A] x C

Shaw, P. D., Graham, M., Kennedy, J., Milne, I., & Marshall, D. F. (2014). *Helium: visualization of large scale plant pedigrees*. **BMC Bioinformatics**, 15(1),

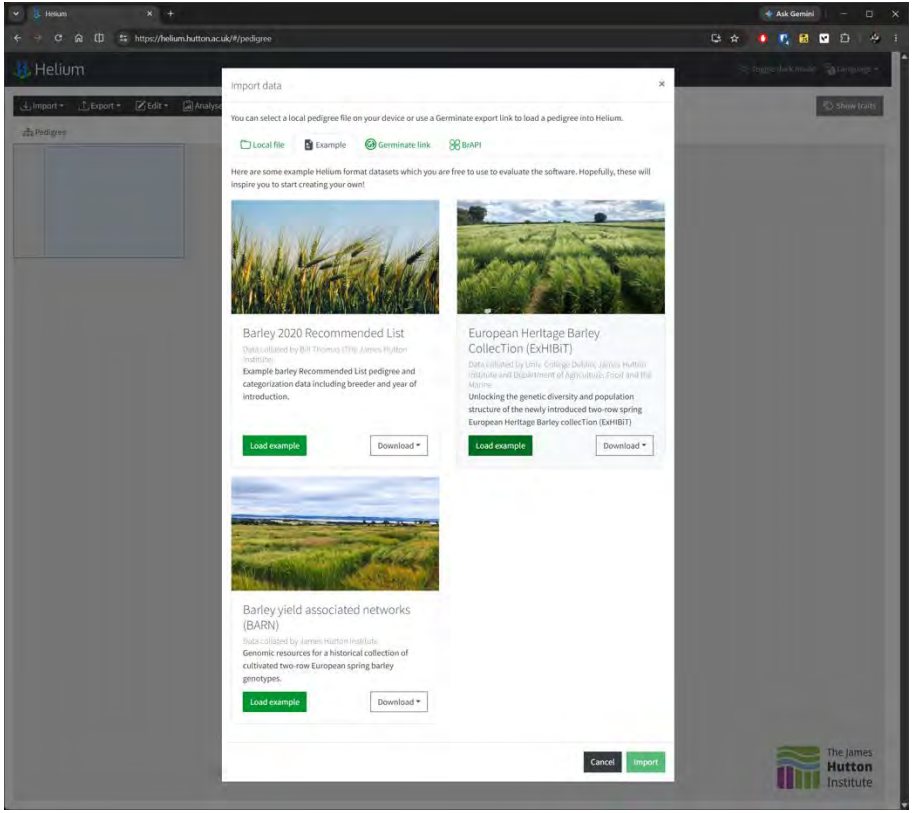
Helium and Neon*
Edit and Visualize Plant Pedigrees

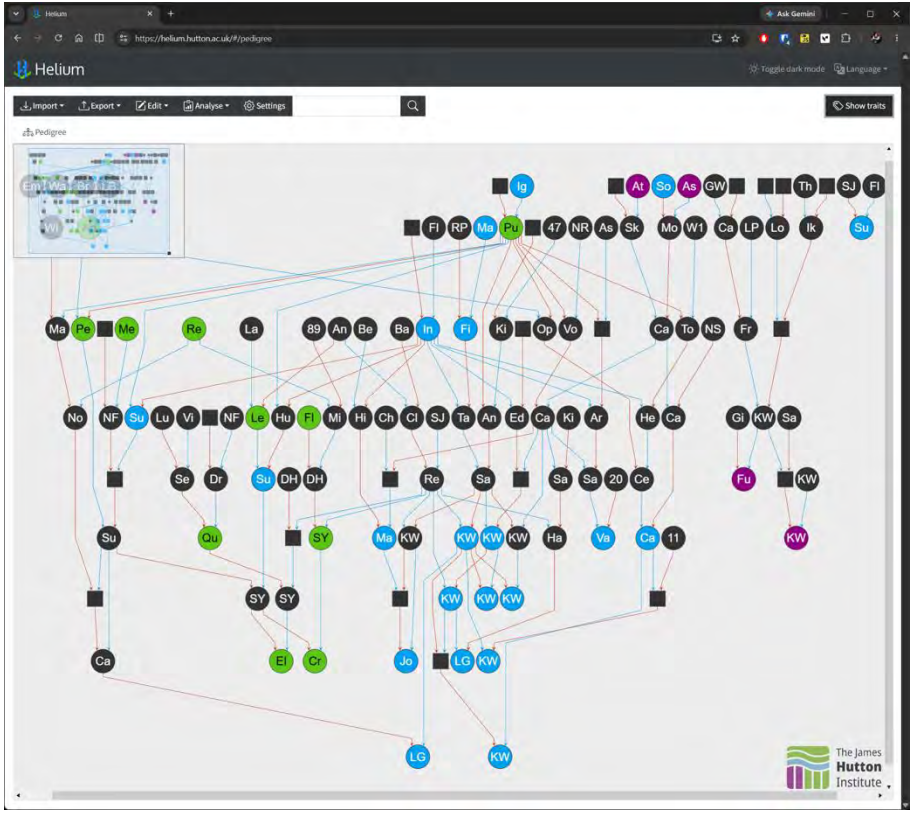


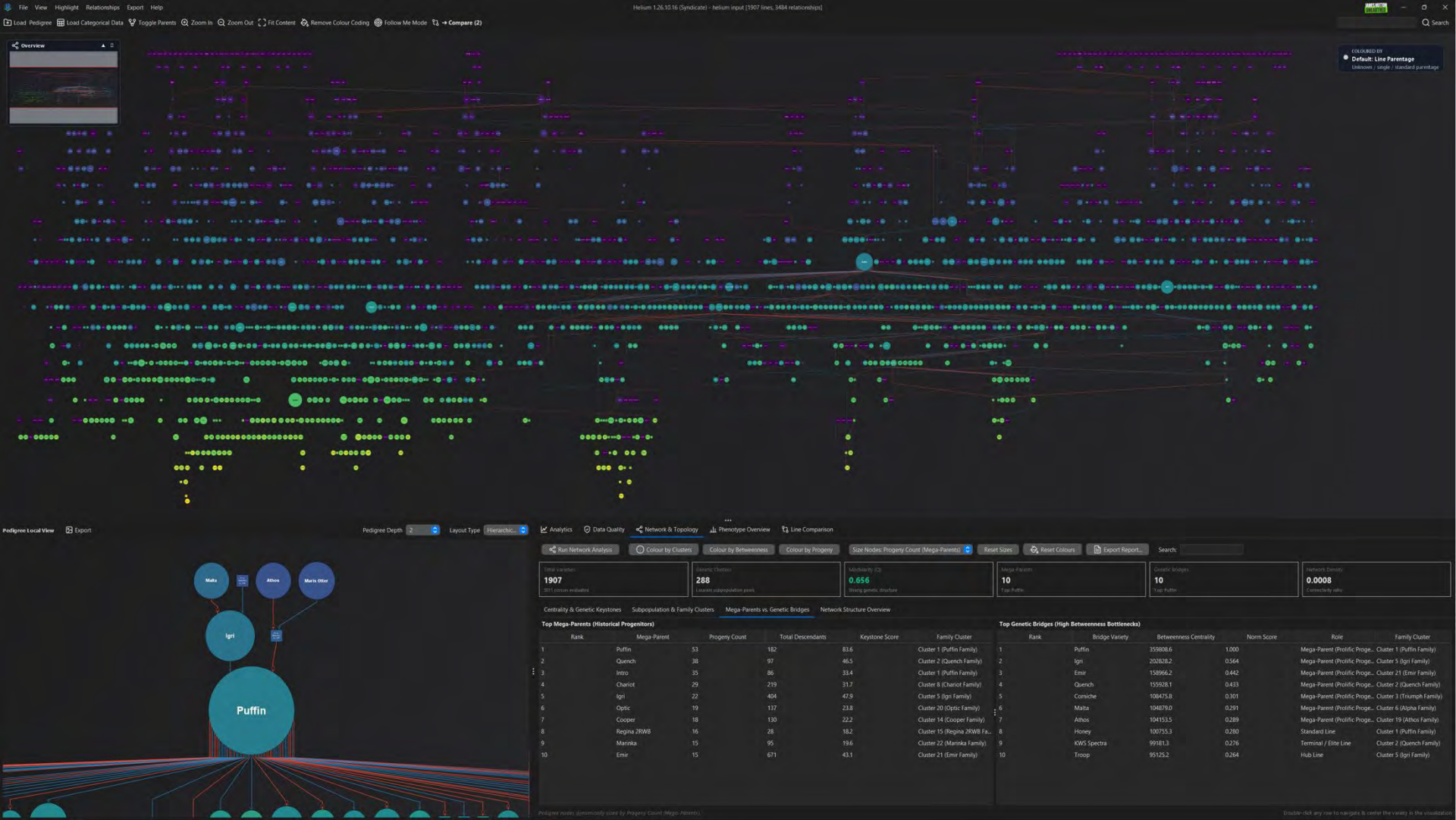
- Plant pedigrees can be complex.
- Purdy of Lamacraft format, a mix?
- Multiple versions of the same pedigree but which one is correct?
- Gives experts tools to maintain shared pedigrees.
- Add files and references to data.

https://helium.hutton.ac.uk

Helium Online (the preferred version)





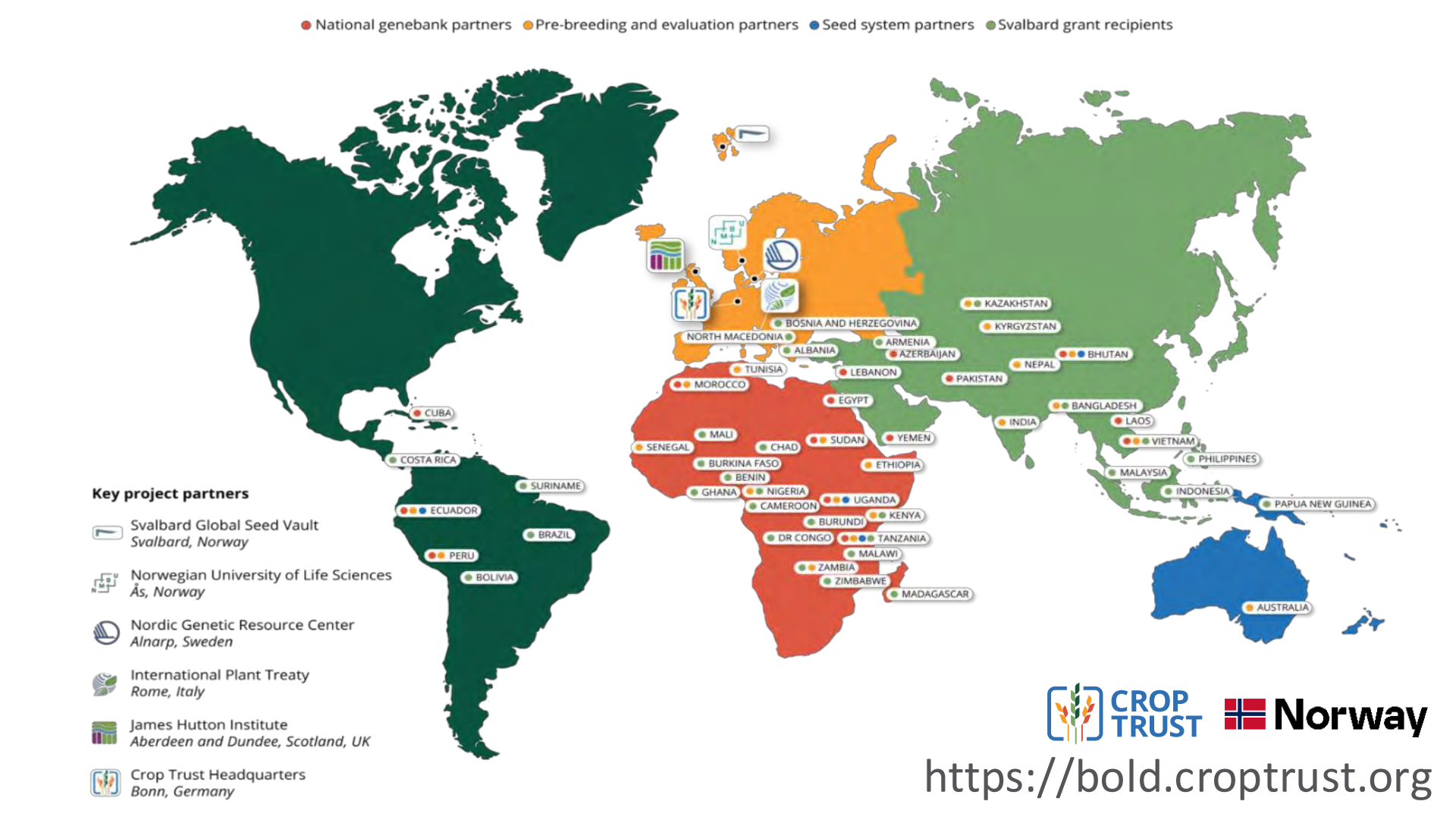

File View Highlight Relationships Export Help
Helium 1.26.10.16 (Syndicate) - helium input [1907 lines, 3484 relationships]
Load Pedigree Load Categorical Data Toggle Parents Zoom In Zoom Out Fit Content Remove Colour Coding Follow Me Mode → Compare (2)
Search
Overview
COLOURED BY
Default: Line Parentage
Unknown / single / standard parentage
Pedigree Local View
Export
Pedigree Depth 2
Layout Type Hierarchic...
Malta
Athos
Maris Otter
Igri
Puffin
Analytics
Data Quality
Network & Topology
Phenotype Overview
Line Comparison
Run Network Analysis
Colour by Clusters
Colour by Betweenness
Colour by Progeny
Size Nodes: Progeny Count (Mega-Parents)
Reset Sizes
Reset Colours
Export Report...
Search:
Total Varieties
1907
3011 crosses evaluated
Genetic Clusters
288
Discrete subpopulation pools
Modularity (Q)
0.656
Strong genetic structure
Mega-Parents
10
Top: Puffin
Genetic Bridges
10
Top: Puffin
Network Density
0.0008
Connectivity ratio
Centrality & Genetic Keystones
Subpopulation & Family Clusters
Mega-Parents vs. Genetic Bridges
Network Structure Overview
Top Mega-Parents (Historical Progenitors)
Rank Mega-Parent Progeny Count Total Descendants Keystone Score Family Cluster
1 Puffin 53 182 83.6 Cluster 1 (Puffin Family)
2 Quench 38 97 46.5 Cluster 2 (Quench Family)
3 Intro 35 86 33.4 Cluster 1 (Puffin Family)
4 Chariot 29 219 31.7 Cluster 8 (Chariot Family)
5 Igri 22 404 47.9 Cluster 5 (Igri Family)
6 Optic 19 137 23.8 Cluster 20 (Optic Family)
7 Cooper 18 130 22.2 Cluster 14 (Cooper Family)
8 Regina 2RWB 16 28 18.2 Cluster 15 (Regina 2RWB Fa...
9 Marinka 15 95 19.6 Cluster 22 (Marinka Family)
10 Emir 15 671 43.1 Cluster 21 (Emir Family)
Top Genetic Bridges (High Betweenness Bottlenecks)
Rank Bridge Variety Betweenness Centrality Norm Score Role Family Cluster
1 Puffin 359808.6 1.000 Mega-Parent (Prolific Proge... Cluster 1 (Puffin Family)
2 Igri 202828.2 0.564 Mega-Parent (Prolific Proge... Cluster 5 (Igri Family)
3 Emir 158966.2 0.442 Mega-Parent (Prolific Proge... Cluster 21 (Emir Family)
4 Quench 155928.1 0.433 Mega-Parent (Prolific Proge... Cluster 2 (Quench Family)
5 Corniche 108475.8 0.301 Mega-Parent (Prolific Proge... Cluster 3 (Triumph Family)
6 Malta 104879.0 0.291 Mega-Parent (Prolific Proge... Cluster 6 (Alpha Family)
7 Athos 104153.5 0.289 Mega-Parent (Prolific Proge... Cluster 19 (Athos Family)
8 Honey 100755.3 0.280 Standard Line Cluster 1 (Puffin Family)
9 KWS Spectra 99181.3 0.276 Terminal / Elite Line Cluster 2 (Quench Family)
10 Troop 95125.2 0.264 Hub Line Cluster 5 (Igri Family)
Pedigree nodes dynamically sized by Progeny Count (Mega-Parents).
Double-click any row to navigate & center the variety in the visualization

BOLD – **B**iodiversity for **O**pportunities, **L**ivelihoods and **D**evelopment

- BOLD is a 10 year 80 million USD Project that builds on the success of the previous 60 million USD Crop Wild Relatives Project.
- The main goal is to “Help ensure food and nutritional security in the face of climate change”
- Helps genebanks, breeders, local communities, scientists and the world!
- Hutton tools are contributing to this work.

CROP TRUST Norway





CROP TRUST Norway

https://bold.croptrust.org

Germinate BOLD WP2 Crops



The James Hutton Institute

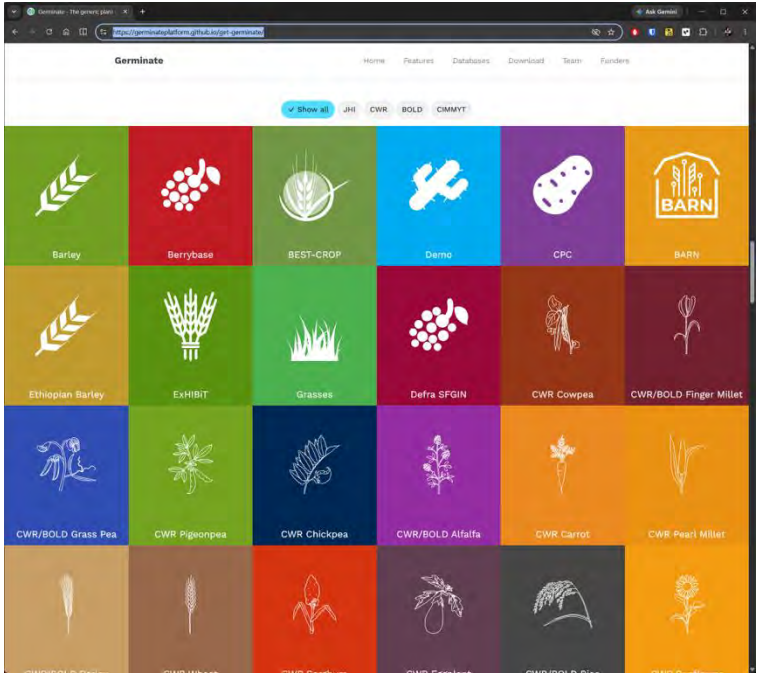
CROP TRUST

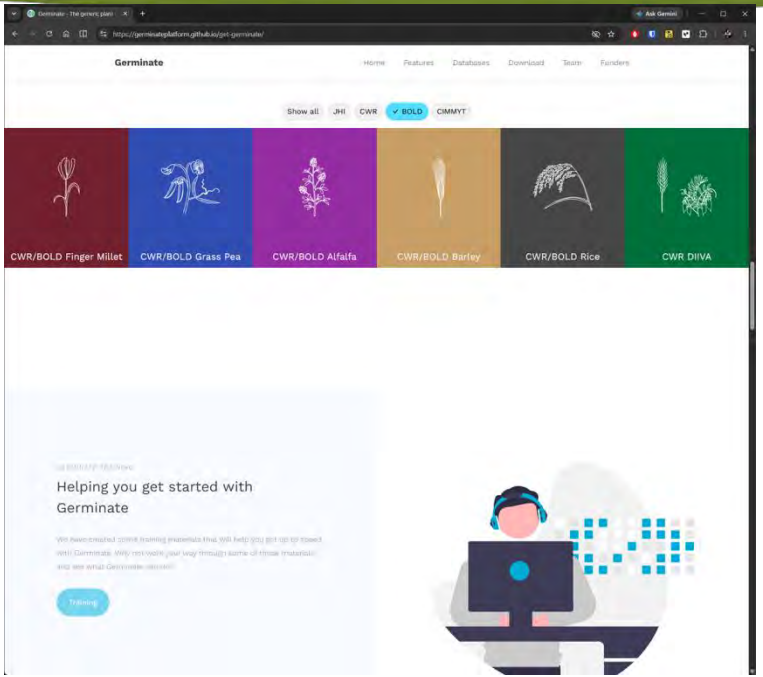
Norway

Accessing BOLD data

https://germinate.hutton.ac.uk

The James Hutton Institute






The James
Hutton
Institute
LOOKING TOWARDS AI...

25M Landscape Genomics and Climate Prediction using AI.

CROP TRUST

Norway

University of Dundee

University of Hawai'i Mānoa

Benjamin Kilian
Anna McCormick
Mikey Kantar



Hajk Drost



Sebastian Raubach



Kelly Houston

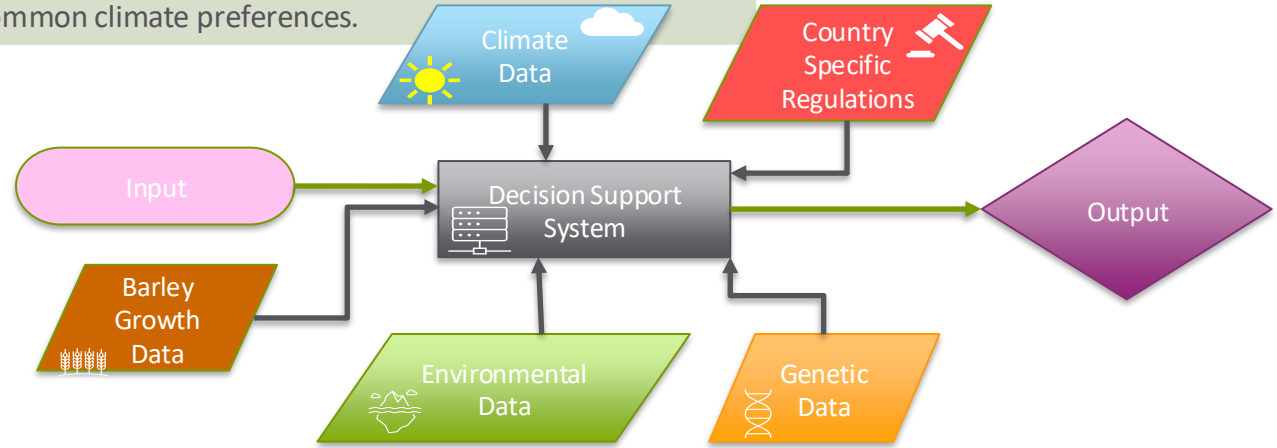


Projects Goals

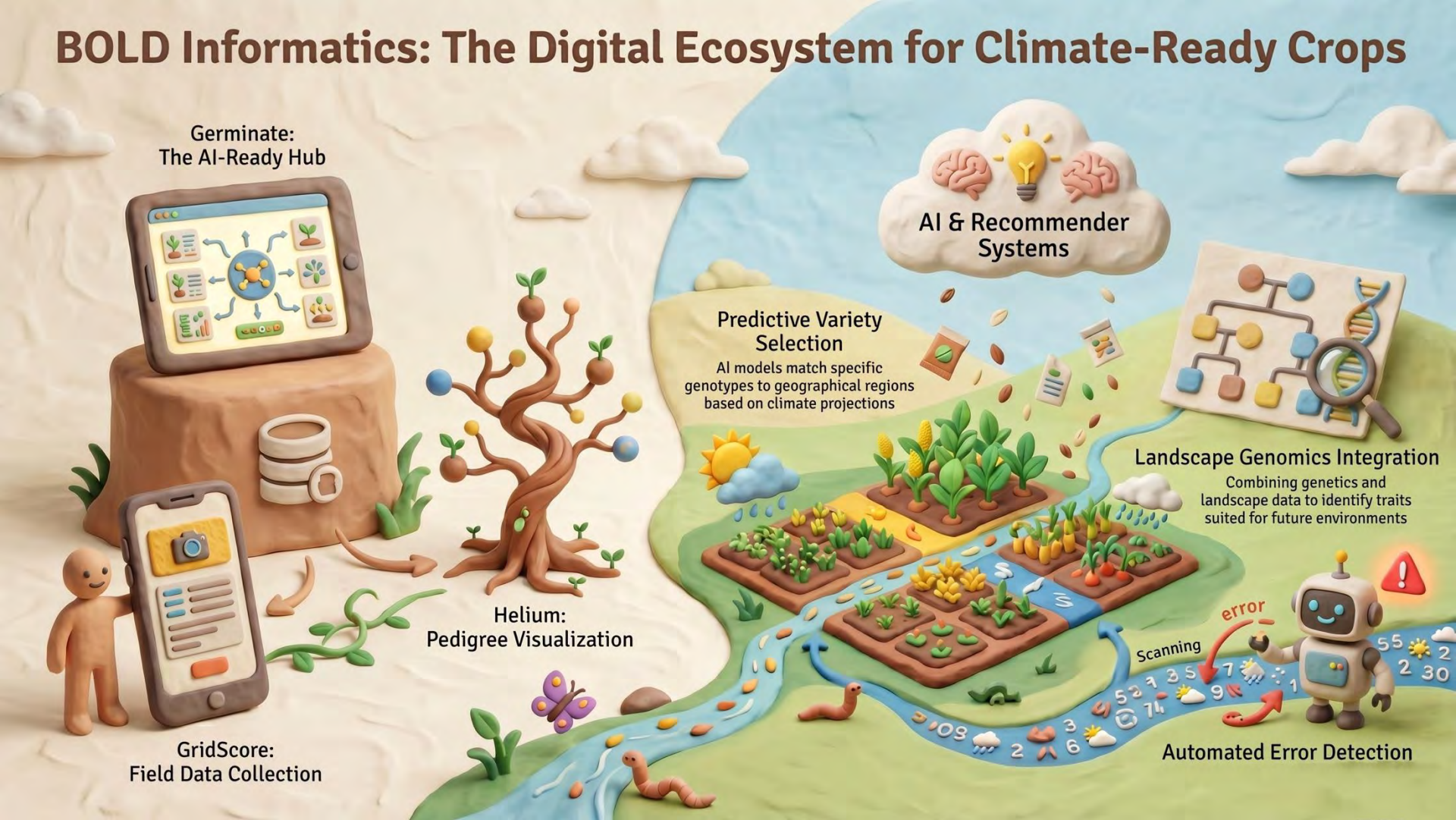
To understand the ways AI can be applied to the farming of barley

To address the key challenges associated with using AI to predict crop growth under the changing environmental landscape around the world.

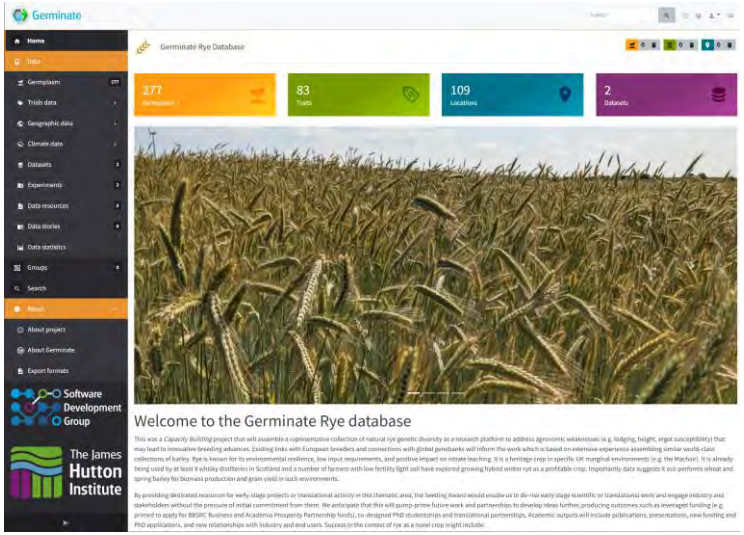
The objective is to develop an AI driven, farm specific variety recommender decision support system for barley which will effectively match genetic diversity with the local landscapes based on their common climate preferences.

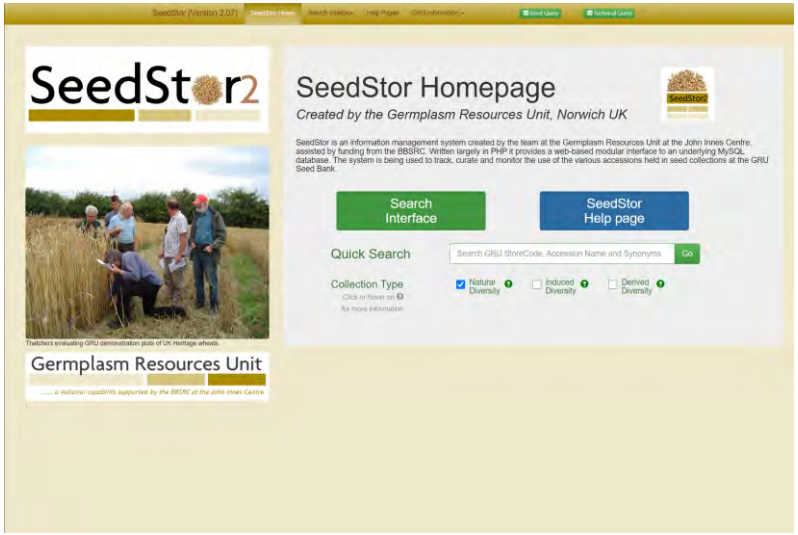


BOLD Informatics: The Digital Ecosystem for Climate-Ready Crops



BBSRC BBR AI-PGR-OMICs: An AI-literate IT infrastructure for genebank management and FAIR germplasm and data provision 2027-2030





Noam Chayut, Paul Shaw, Burkhard Stuernagel, Sebastian Raubach

The James Hutton Institute

John Innes Centre

Biotechnology and Biological Sciences Research Council

EU COST Action CA25121

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

HARVEST

Harnessing AI for Responsive, Valuable, and Efficient Sustainable Technologies (HARVEST)

The HARVEST project organises its efforts into four distinct Working Groups (WGs). These groups are designed to bridge the gap between agricultural data and AI application.

WG1: Omics Data Management
WG2: Phenotypic Data Standards
WG3: AI Tools for Research, Agriculture and Farm Recommendations
WG4: Dissemination, Communication & Training

This COST Action will directly benefit barley and other crops informatics across Europe.
Link to apply to our working group
https://www.cost.eu/actions/CA25121/



The James Hutton Institute

BIOLOGICAL BACKUPS WITH DIGITAL CATALOGUES OR DIGITAL CATALOGUES WITH BIOLOGICAL BACKUPS?

The James Hutton Institute

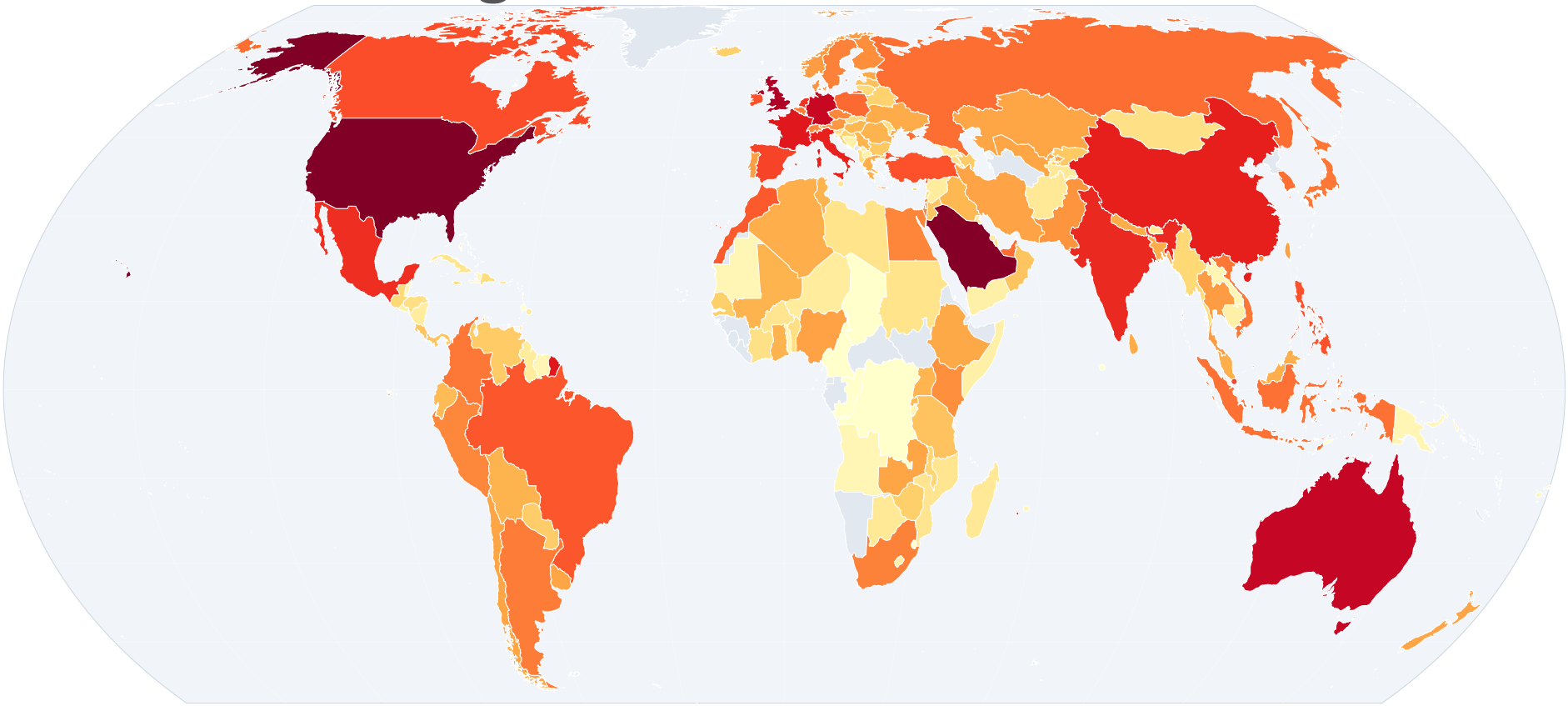
TRAINING?

https://germinateplatform.github.io/germinate-training/

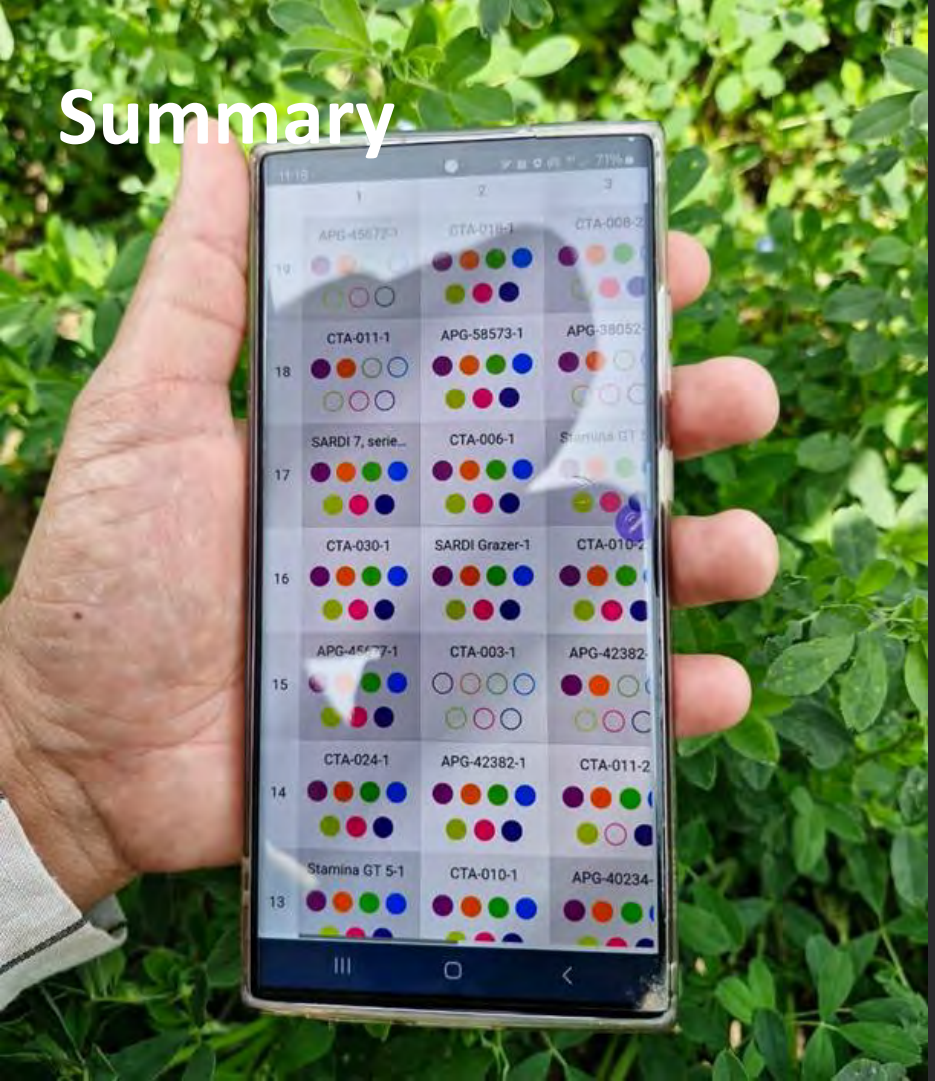
https://www.youtube.com/@cropgeeks

SPEAK TO US!

Worldwide usage



Summary



- Think about your data and how you will collect, store and analyse your data!
- All our software is free to use **for you**.* Other options are available so apologies for the sales pitch ☺
- Our software is being used to support large international projects as well as local needs.
- We welcome contributions, feedback and ideas on how we can improve.
- Please just give us a shout if you want to try any of our tools with your own data or want an extended demo.
- Whatever you do STOP WRITING STUFF ON PAPER! :-)

Thank you!

- **Sebastian Raubach**
- **Eloise Clarkson**
- **Amir Yousefi**
- Hajk-Georg Drost
- **Miriam Schreiber**
- Benjamin Kilian
- Ingo Hein
- Dan Sargent
- Guille Mendiondo
- Agnes Aboagye
- **Micha Bayer**
- Isabelle Colas
- Malcolm Macaulay
- Gaynor McKenzie
- Noam Chayut
- Andrew Reid

Scottish Government
Riaghaltas na h-Alba
gov.scot

RESAS
Rural & Environment Science and Analytical Services

CROP TRUST

Funded and Co-funded by the European Union

Norway
Norwegian Ministry of Foreign Affairs

INTERNATIONAL BARLEY HUB

The James Hutton Institute
Scientific Services

westfalia fruit

IPK LEIBNIZ INSTITUTE

National Potato Innovation Centre

CIP INTERNATIONAL POTATO CENTER
CGIAR

UKRI
Biotechnology and Biological Sciences Research Council

University of Dundee

UKRI Innovate UK
Knowledge Transfer Partnerships

The University of Nottingham

GOV·UK
Department for Environment Food & Rural Affairs

John Innes Centre

Abertay University

SEFARI

syngenta

ef Esmée Fairbairn FOUNDATION

Kirkhouse Trust

CFF
CROPS FOR THE FUTURE

paul.shaw@hutton.ac.uk