From Bits to Bites: Software Development supporting Plant Genetics, Breeding and Genetic Resources

2nd April 2025 Marche Polytechnic University, Ancona, Italy

Dr Paul Shaw and Sebastian Raubach

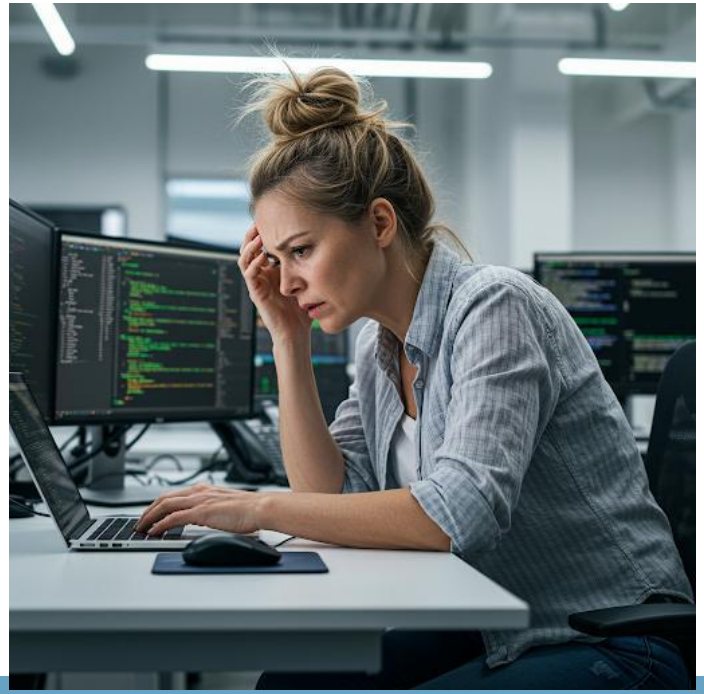
Department of Information and Computational Sciences

The James Hutton Institute

Breeding Value

Plant science produces a lot of data…

Google Gemini: ‘Show me an image of a scientist working with data’



3* V’s
Volume,
Variety,
Velocity,
Veracity and Value

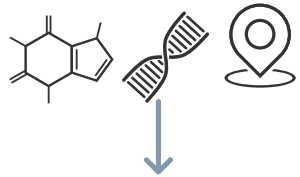
But we really don’t have any option…

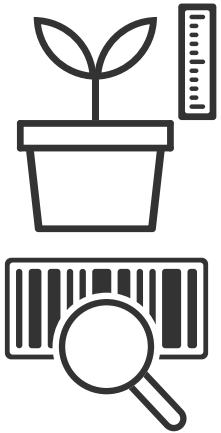
Simplifying plant pre-breeding…

The James Hutton Institute

2. Merge and Wrangle

(Historical and background data)





4. Visualize

(and export)

1. Collect

3. Store

Potential Solutions?

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.

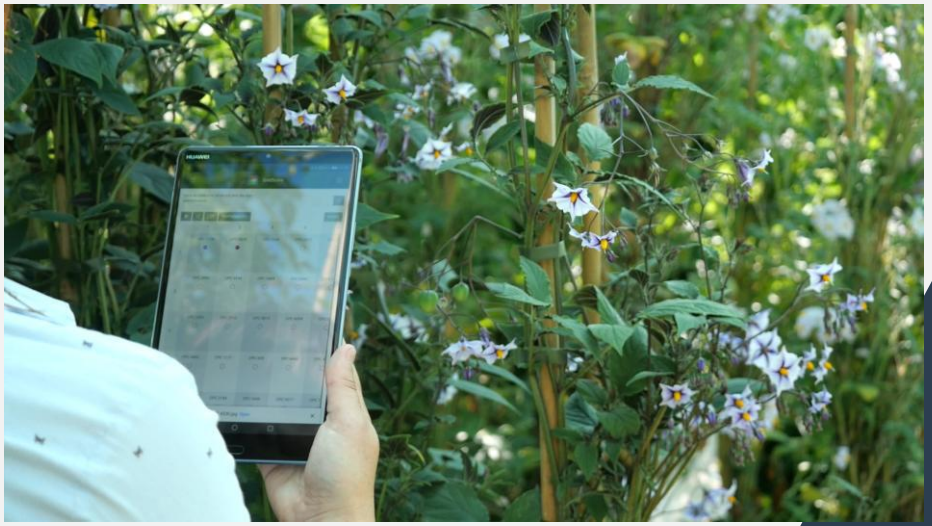
Making Diversity Available

The James Hutton Institute

GridScore

Mobile phenotypic data collection

gridscore.hutton.ac.uk



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

Common Errors and Mistakes

Number swaps

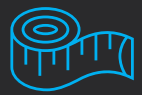
An easy mistake to make, e.g. 24 vs 42.
Trait limits and visualizations can help prevent and discover these errors.

Decimal point issues

Decimal point in the wrong place.
Trait limits and visualizations can help prevent and discover these errors.

User differences

Different people may score the same trait differently. Flower colour for example, "dark purple" vs "medium purple". Choose distinct categories and provide reference sheets.



Different units

Using different units to measure the same thing.
E.g. centimetres vs inches. Seems easy to avoid, but still happens. Provide trait descriptions including units.

Different categories between sites

The same trait might be scored differently between sites. "1, 3, 5, 7, 9" vs "very low, low, medium, high, very high". Share trait definitions with collaborators.

Plot mix-up

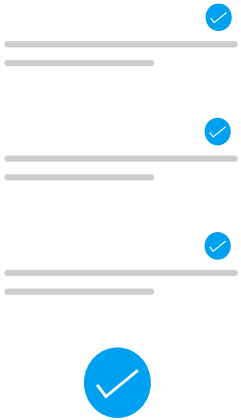
Scoring data for the wrong plot.
Use barcodes/QR-codes or guided walks to prevent this.

Stop Writing Stuff on Paper!

10%

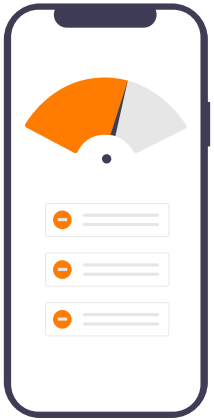
Benefits

The James Hutton Institute



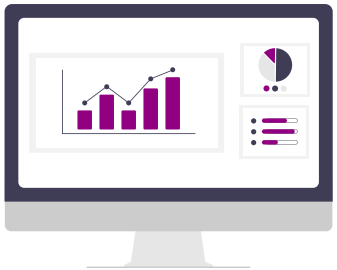
Data verification

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



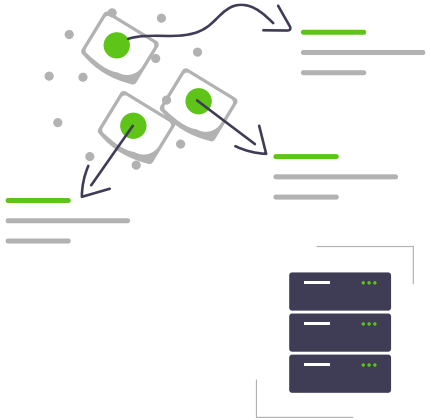
Better efficiency

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**.

What happens if you don’t collect high quality data?

1. Inaccurate conclusions, false discoveries and skewed results
2. Wasted resources
3. Damage to scientific integrity and loss of trust
4. Delayed advancement
5. Difficulties in reproducibility

Science fictions: exposing fraud, bias, negligence and hype in science by Stuart J. Ritchie

Germinate

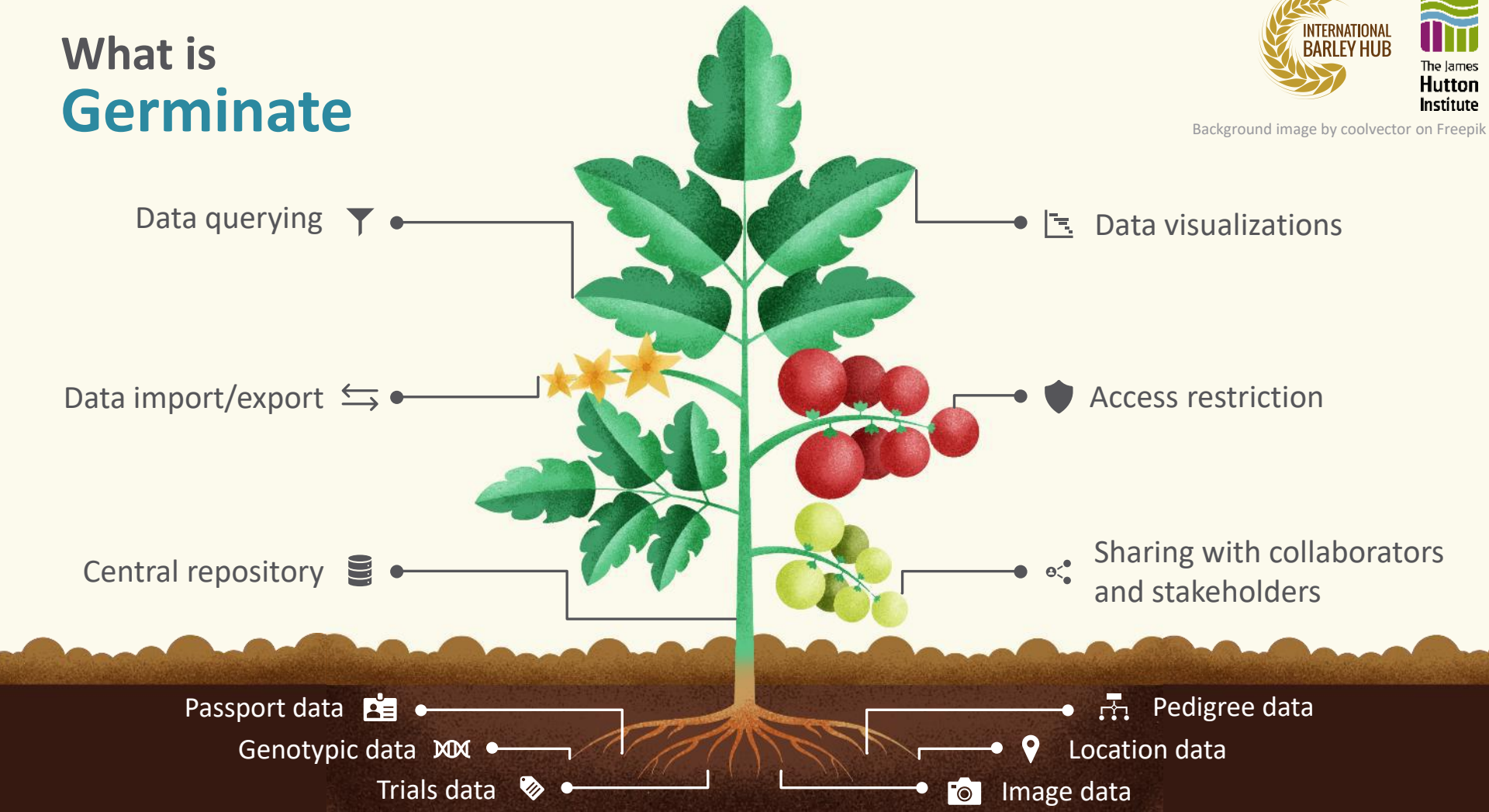
Plant genetic resources database

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

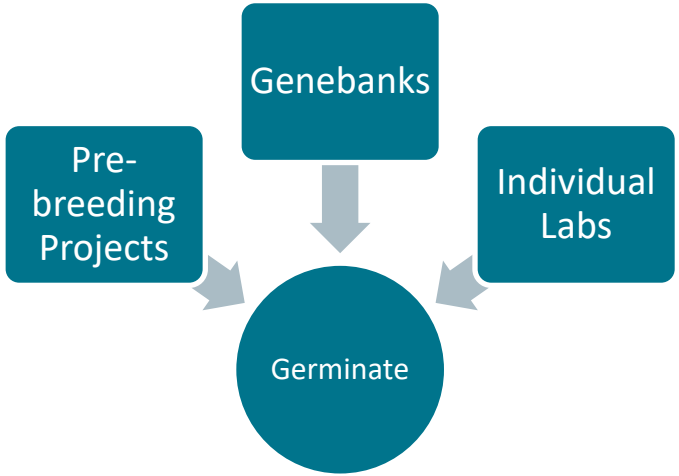


INTERNATIONAL BARLEY HUB

The James Hutton Institute

Background image by coolvector on Freepik

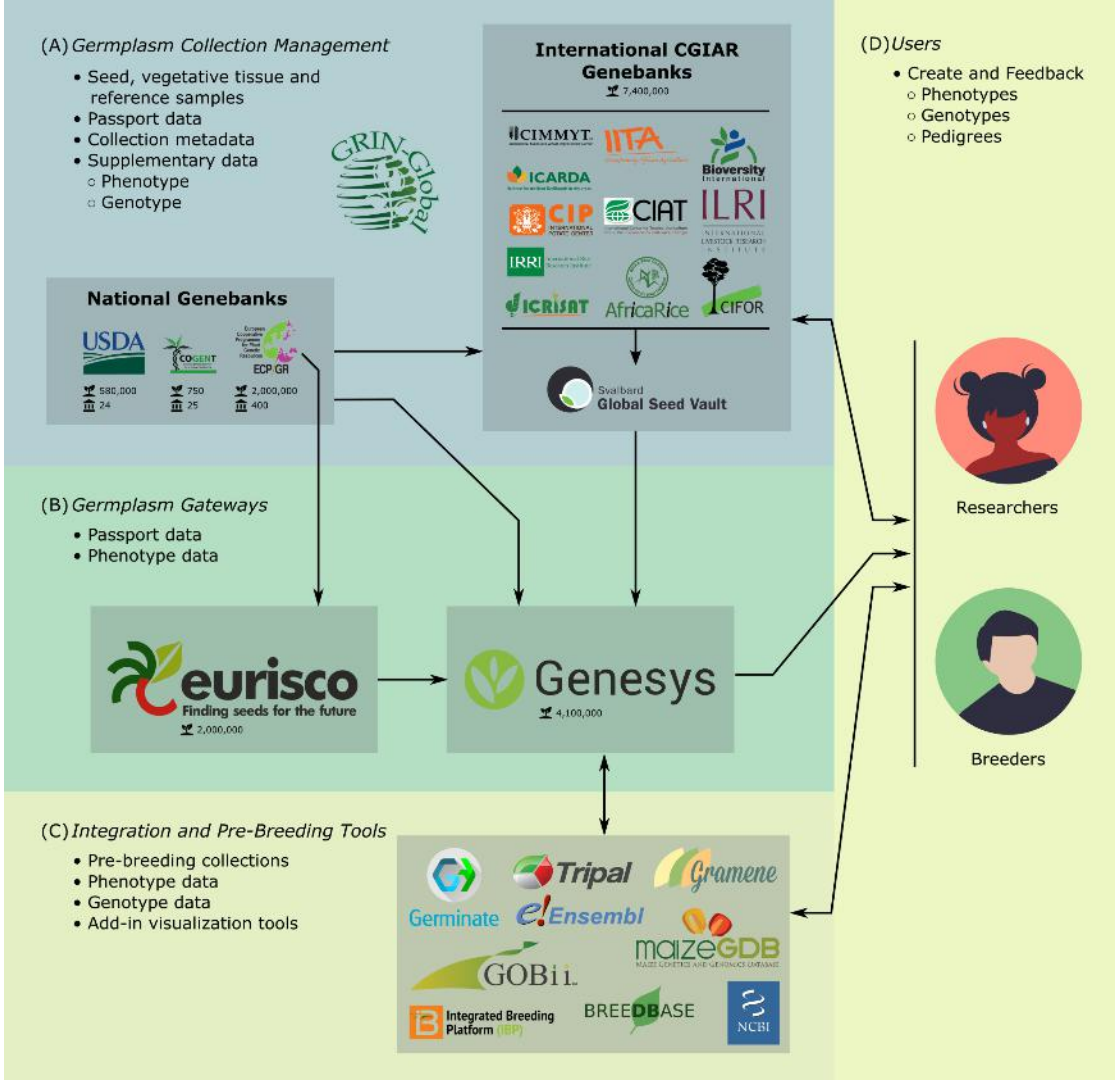
FAO GLIS Positioning



Shallow v Deep Data

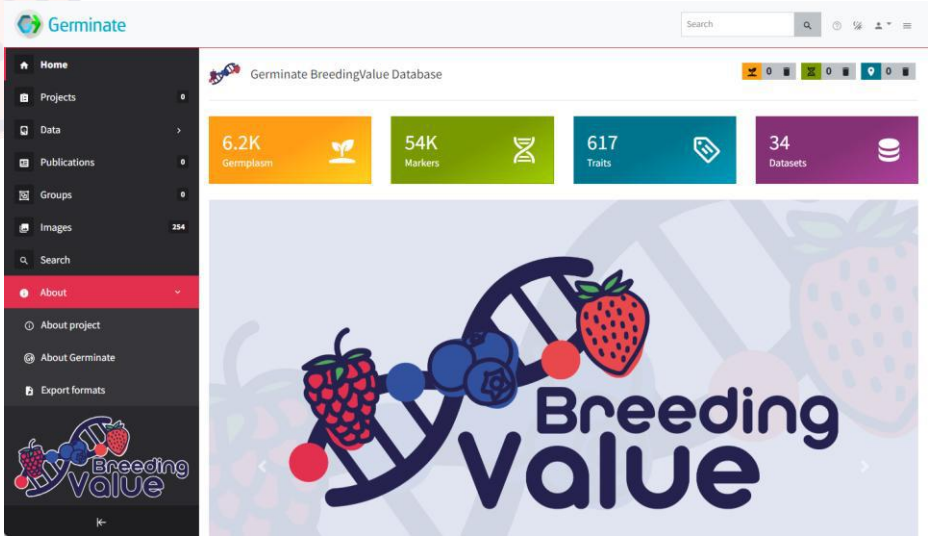
Wide and shallow – Narrow and deep

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.

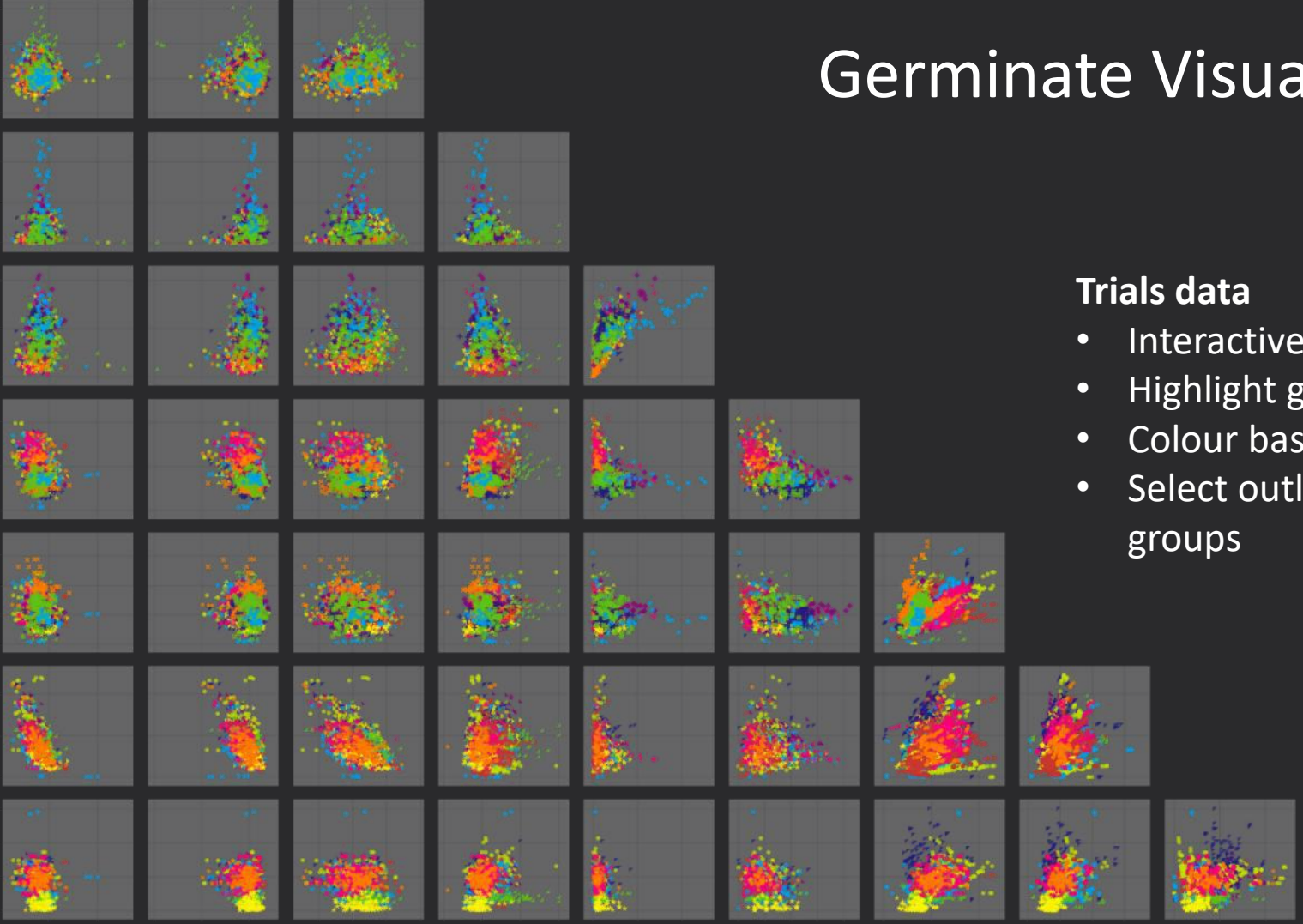


Where to find BreedingValue data?

- BreedingValue data is available at: https://germinate.hutton.ac.uk/breedingvalue
- Not all data public yet, but lots to explore
- No user account/registration required

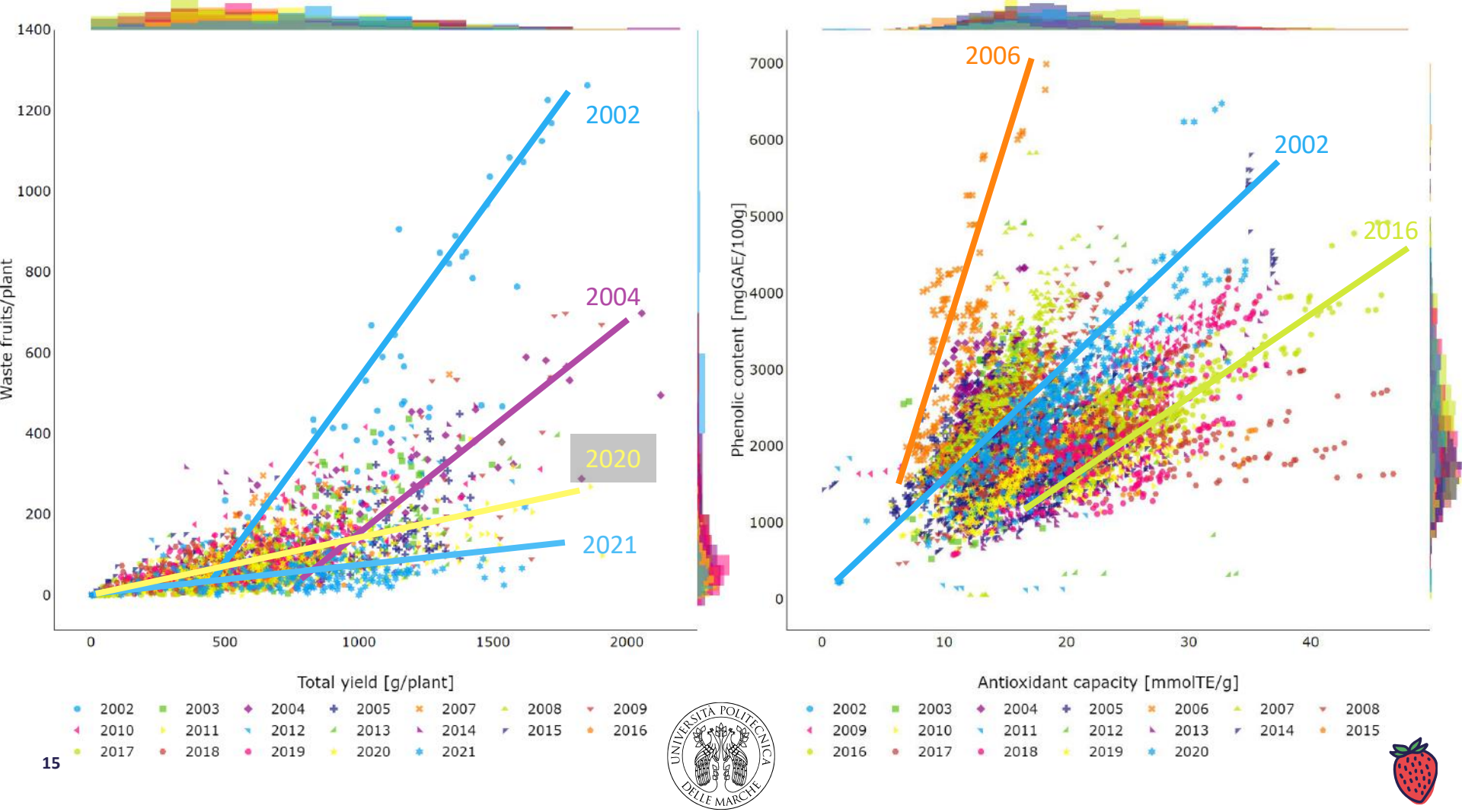


Germinate Visual Analytics



Trials data

- Interactive plots
- Highlight germplasm groups
- Colour based on *things*
- Select outliers and create new groups

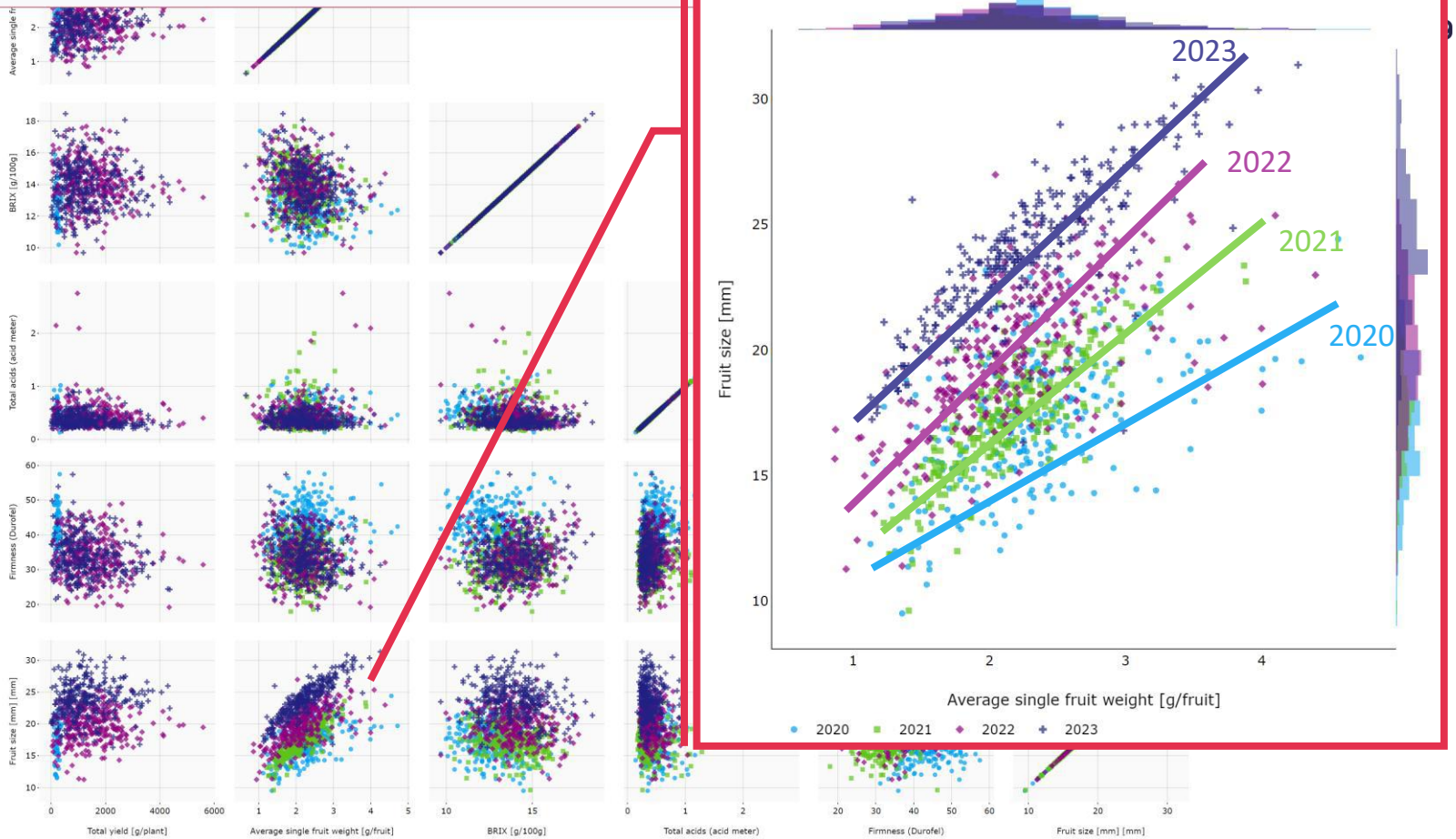

Waste fruits/plant
Total yield [g/plant]
2002
2004
2020
2021
Phenolic content [mgGAE/100g]
Antioxidant capacity [mmolTE/g]
2006
2002
2016
2002 2003 2004 2005 2007 2008 2009
2010 2011 2012 2013 2014 2015 2016
2017 2018 2019 2020 2021
2002 2003 2004 2005 2006 2007 2008
2009 2010 2011 2012 2013 2014 2015
2016 2017 2018 2019 2020

Germinate

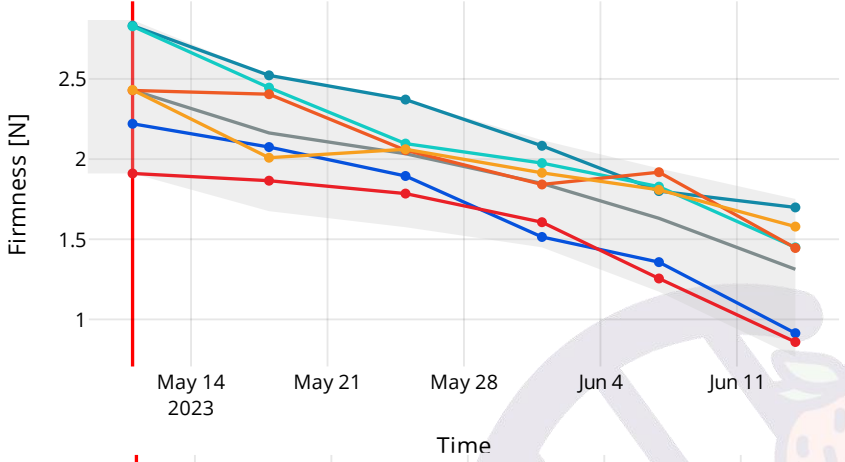
Search

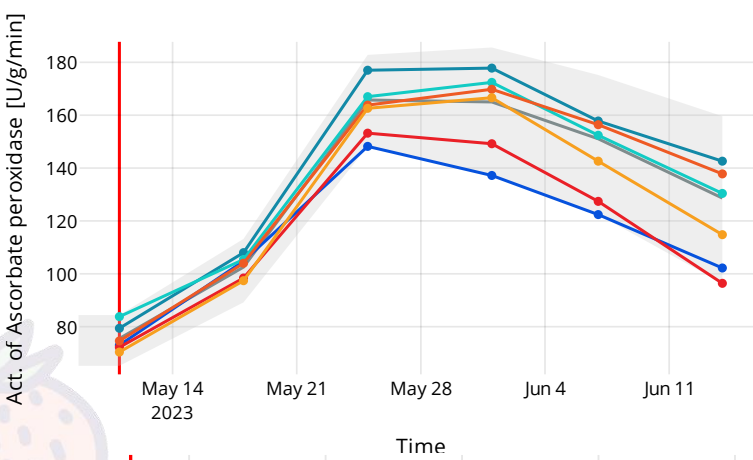
- Home
- Projects 0
- Data
- Publications 0
- Groups 0
- Images 254
- Search
- About

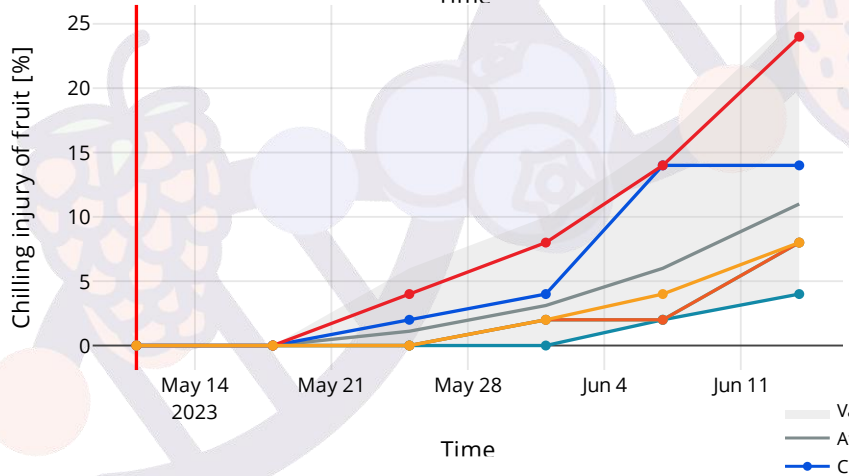
Breeding Value

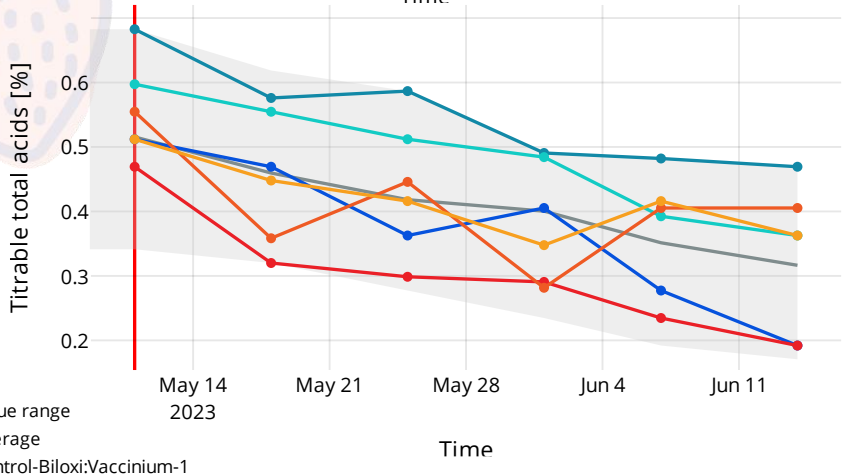


The James Hutton Institute









Value range
Average
Control-Biloxi:Vaccinium-1
GABA-Biloxi:Vaccinium-1
Hexanal-Biloxi:Vaccinium-1
Control-Star:Vaccinium-1
GABA-Star:Vaccinium-1
Hexanal-Star:Vaccinium-1

Summary

- Think about your data and how you will collect, store and analyse it!
- **Go and use the data!**
- **Watch the seminar, have an explore, reach out to us!**
- 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.
- **Don’t write stuff down with pen and paper!**
- **Think about naming of samples!**

PGR Work @ Hutton

- Sebastian Raubach
- Elisa Senger
- Miriam Schreiber
- Micha Bayer
- Kelly Houston
- Susan McCallum
- Julie Graham
- Malcolm Macaulay
- Niki McCallum
- Mohamed Salama
- Gaynor McKenzie
- Joanne Russell
- Pauline Smith
- Luke Ramsay
- Robbie Waugh
- + barley, potato and soft fruit groups @Hutton

Acknowledgements

Scottish Government
Riaghaltas na h-Alba

CROP TRUST

European Commission

Department for Environment Food & Rural Affairs

Norway

Norwegian Ministry of Foreign Affairs

International Barley Hub

The James Hutton Institute
Scientific Services

Norad

RESAS

DIAGEO

syngenta

DIVSEEK International Network Inc.

SEFARI

SusCrop – ERA-NET
Cofund on Sustainable Crop Production
FACCEJPI

BBSRC

Templeton World Charity Foundation

UKRI Innovate UK

Esmée Fairbairn Foundation

University of Dundee

SGS

SGS

Athena SWAN Bronze Award
Gender Charter

Stonewall Diversity Champion Scotland

The King's Awards for Enterprise 2023 Sustainable Development

Scottish Government
Riaghaltas na h-Alba
gov.scot