

INTEGRATED DATABASE “GERMINATE”

THE DATA REPOSITORY OF THE PROJECT BREEDINGVALUE

14. FEB. 2024 | ELISA SENGER

IBG-4, FORSCHUNGSZENTRUM JUELICH, GERMANY



OUTLOOK

- ▶ **Intro: Need for data repository**
- ▶ **The Germinate database**
- ▶ **Data preparation for upload**
- ▶ **BreedingValue berry data @ Germinate**



NEED FOR DATA REPOSITORY

Introduction

- FAIR¹ and open data
 - “desirable” or mandatory under public funding
 - increases value and utility of the data (repetitive and broader applications)
 - knowledge becomes more accessible for exploitation
- No common “plant database” present or limited for integrated², multi-omics data
- Databases host data (mid- to long-term)
- Databases provide tools to explore the data online

¹ Findable + Accessible + Interoperable + Re-useable

² data genres connected within database (e.g. genotypic, phenotypic, metabolomics, climate, passport)



GERMINATE (DEVELOPED BY JHI)

In a nutshell...

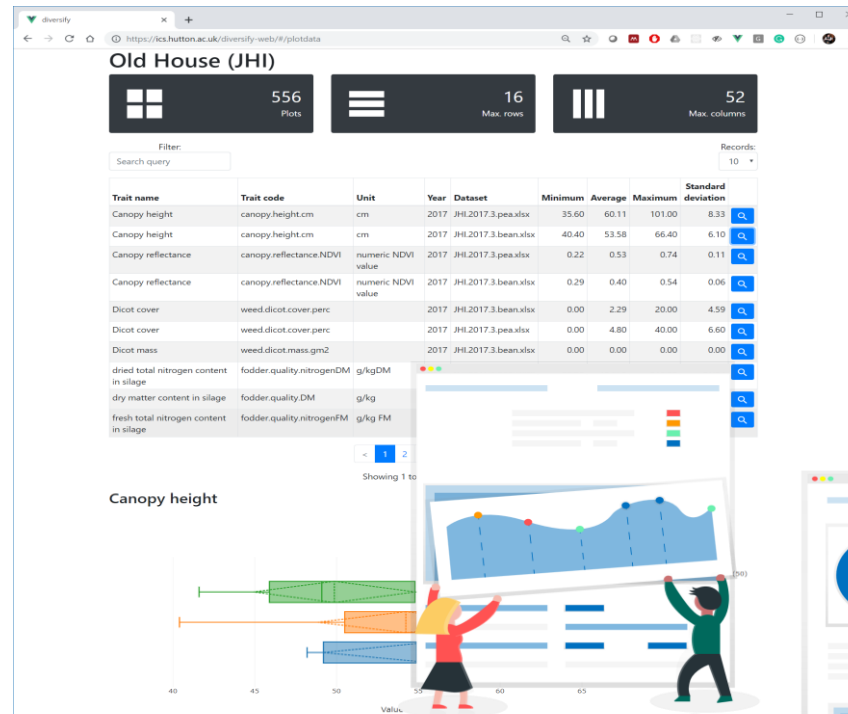
From this...

...to this



Source: Paul Shaw, JHI

Year	Plotcode	OwnCode	Site	Location	EUPartner	Block	Plot	Subplot	Row ID	Column (Bed)	PlantPartner	Rep	Subre
2017	OBJ11112	OBJ1111211	Old House	B52	JHI	1	1	1	1	1	2 Cereal	1	
2017	OBJ11112	OBJ1111221	Old House	B52	JHI	1	1	1	1	2	Legume	1	
2017	OBJ12123	OBJ1212311	Old House	B52	JHI	1	2	1	2	3	Cereal	1	
2017	OBJ12123	OBJ1212321	Old House	B52	JHI	1	2	1	2	3	Legume	1	
2017	OBJ13134	OBJ1313411	Old House	B52	JHI	1	3	1	3	4	Cereal	1	
2017	OBJ14145	OBJ1414521	Old House	B52	JHI	1	4	1	4	5	Legume	1	



Source: Paul Shaw, JHI

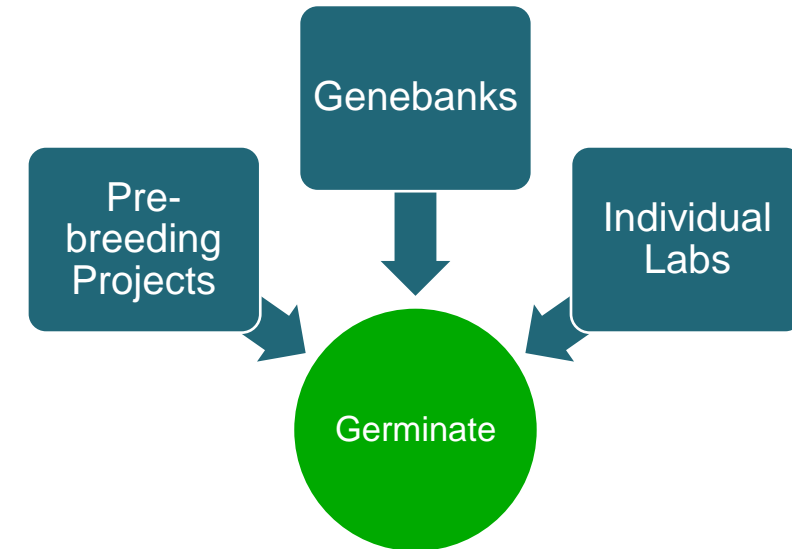
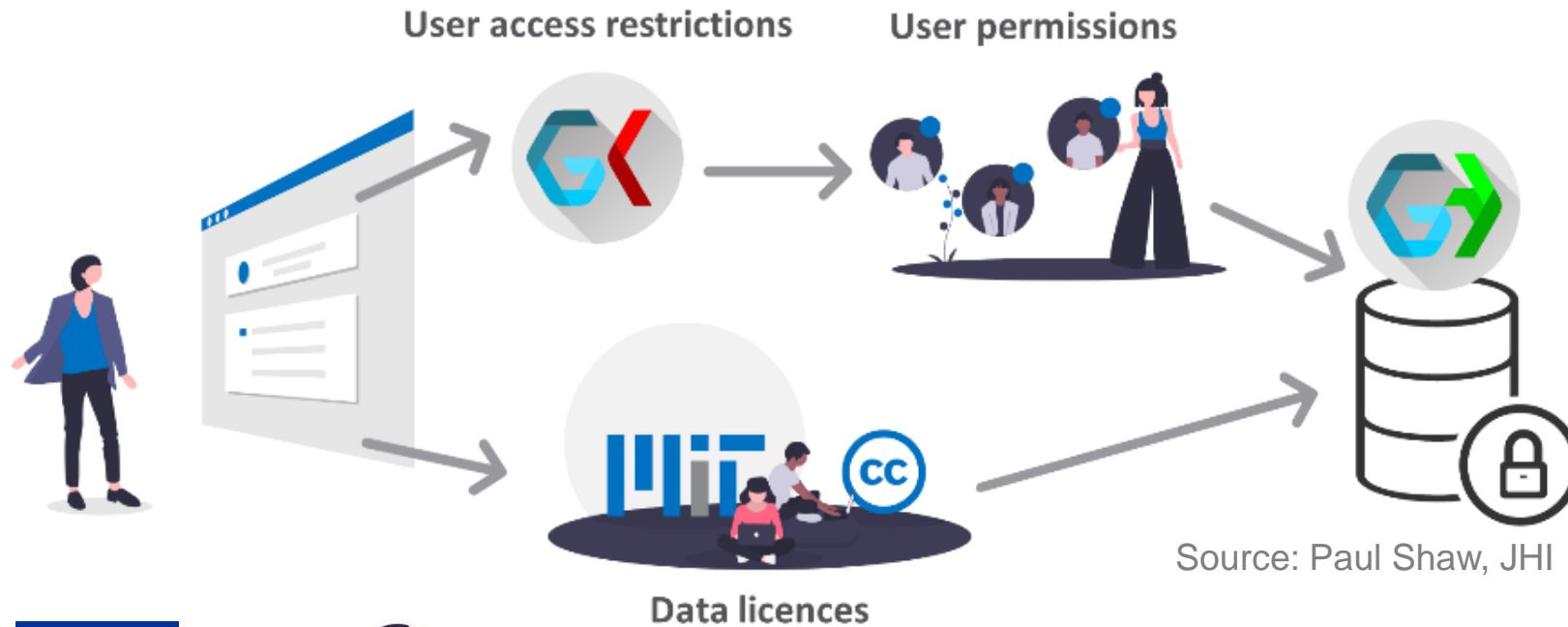


GERMINATE

Overall objective

“Making experimental data available to breeders, scientific researchers and farmers in approachable and tailored ways will increase the utility and value of the data and the knowledge that is generated by each project is more accessible for exploitation.”

→ But we can keep your data private as well!



GERMINATE

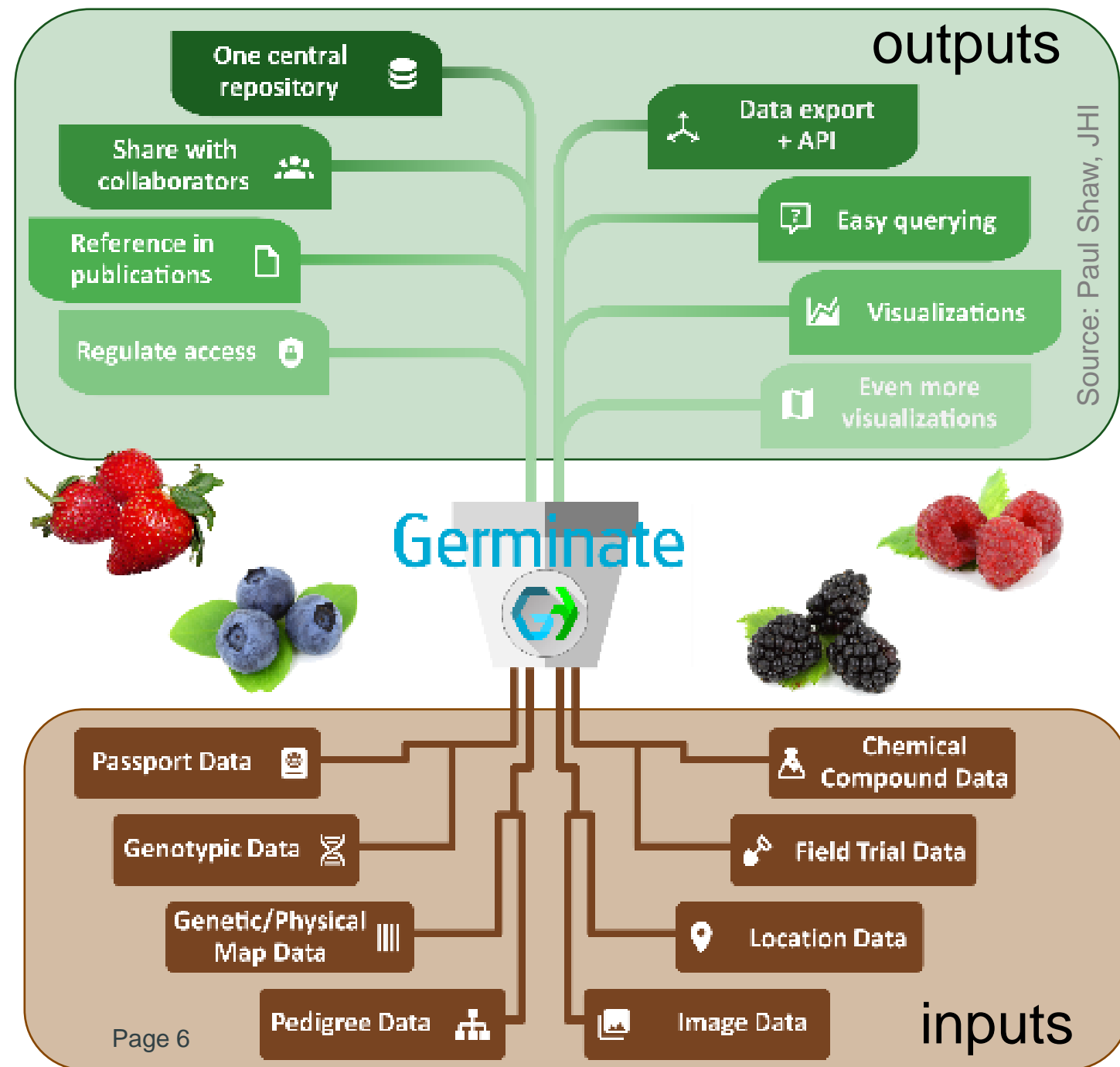
Data of plant genetic resources

- Connecting different data genre
- Visualizing + exploring
- Compatible with MCPD + Dublin Core
- BrAPI implementation currently being extended

Publications:

Shaw et al. (2017) Germinate 3
doi:10.2135/cropsci2016.09.0814

Raubach et al. (2020) From bits to bites
doi:10.1002/csc2.20248



GERMINATE

Data of plant genetic resources

- Download + citation

<https://germinateplatform.github.io/get-germinate/>

<https://ics.hutton.ac.uk/germinate-demo/#/home>

Germinate

Export formats

Data from Germinate can be exported into various formats. Below is a list of tools that can be used to open, explore and visualize these data types.

📌 All 📌 Genotype 📌 Phenotype 📌 Pedigree 📌 PCA

- Flapjack**
Flapjack is a free multi-platform tool for interactive visualization of high-throughput genotype data, allowing for rapid navigation and comparisons between lines, markers and chromosomes and visualization around QTL features and characterization data.
📌 Genotype 📌 Phenotype
Download
- Helium**
Helium is a free multi-platform tool for interactive visualization of plant pedigree data. It provides users with the ability to overlay varying data types onto a pedigree framework and provides data exploration tools to allow users to explore and interrogate their data.
📌 Pedigree 📌 Phenotype
Download
- CurlyWhirly**
CurlyWhirly is designed for the visualization of 3D data provided as Cartesian coordinate data and is particularly well suited for visualizing the outputs of analyses such as Multidimensional Scaling and Principal Coordinates Analysis.
📌 PCA
Download
- Strudel**
Strudel is a graphical tool for visualizing genetic and physical maps of genomes for comparative purposes. The application aims to let the user examine their data at a variety of different levels of resolution, from entire maps to individual markers, and explore syntenic relationships between genomes.
📌 Genotype
Download
- Excel**
Microsoft Excel is a spreadsheet developed by Microsoft for Windows, macOS, Android and iOS. It can be used to explore genotypic and phenotypic data.
📌 Genotype 📌 Phenotype
Download
- Tassel**
TASSEL is a software package used to evaluate traits associations, evolutionary patterns, and linkage disequilibrium.
📌 Genotype
Download

Version 4.2.0

Germinate © 2022 The James Hutton Institute.



GERMINATE

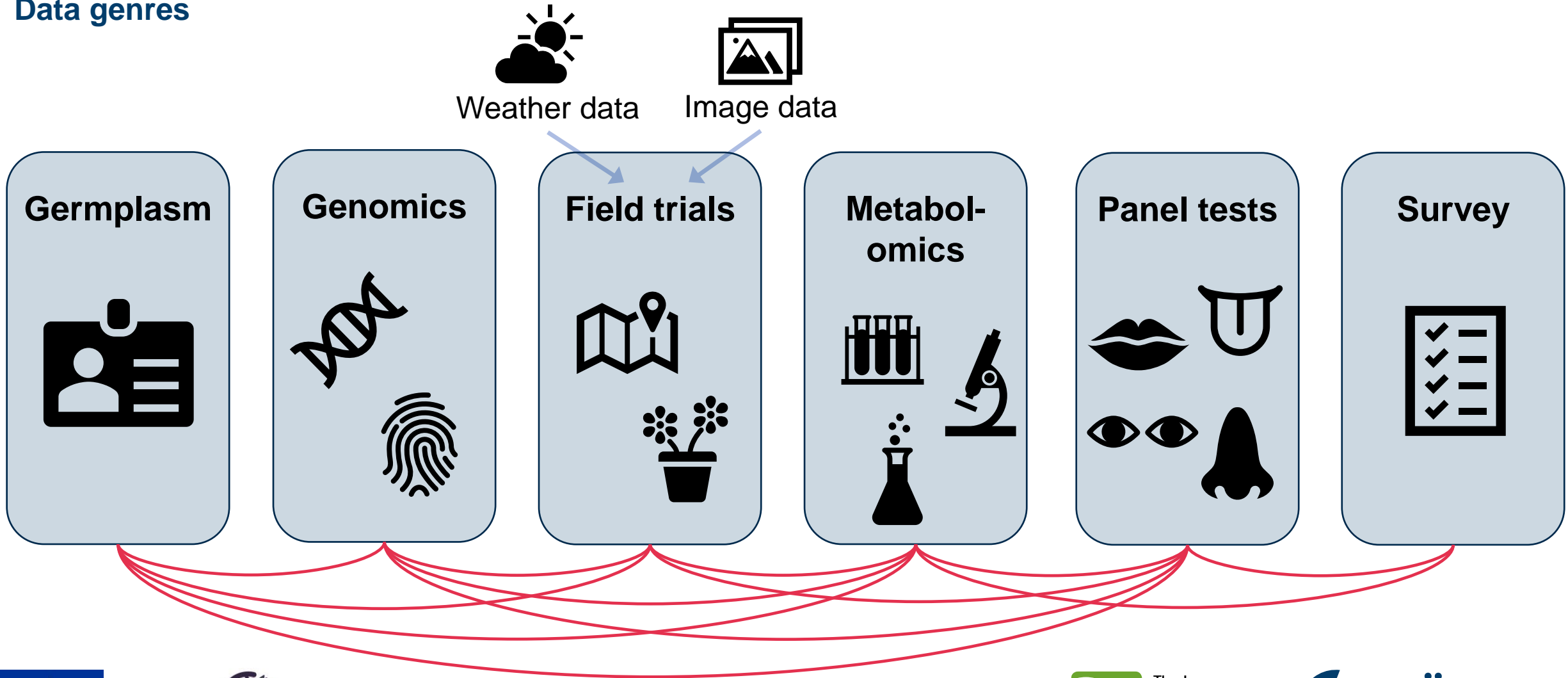
Data of plant genetic resources

- Growing number of users
- Growing number of plant species (21 public databases + few private)
- Growing number of tools and functions



BREEDINGVALUE DATA

Data genres



DATA PREPARATION FOR UPLOAD

Introduce strategy and template files to data providers

- Motivate, encourage, justify, clarify
- FAIR and open data (but requirements can be lowered and adapted)
- Share data with consortium and (ideally) with public

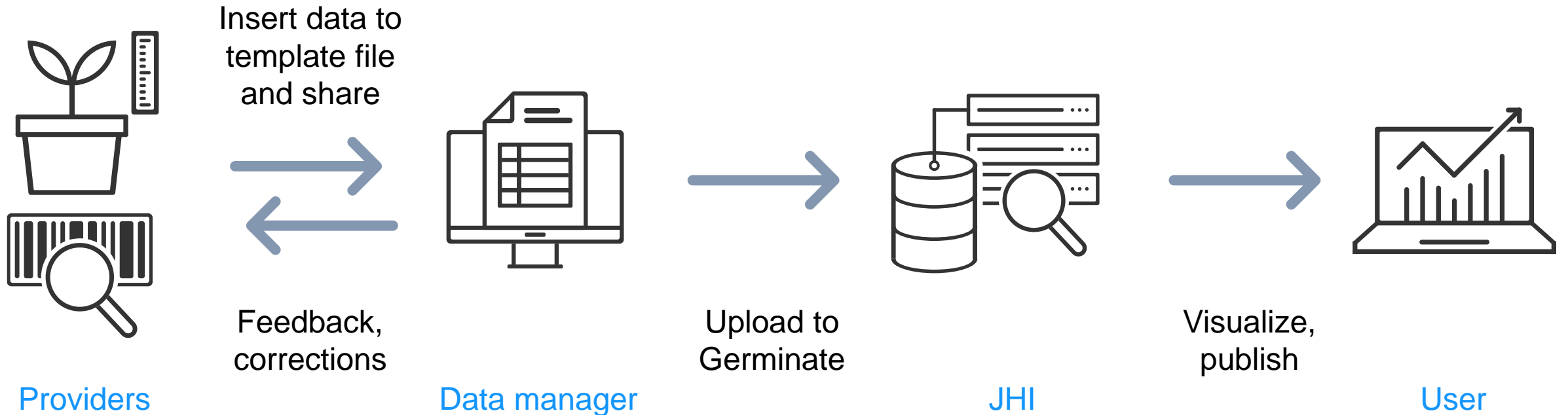
Name	Short Name	Description	Data Type	Unit Name	Unit Abbrev	Unit De	Trait categories (comma	Min (only fo	Max (only fo
Marketable Yield	MY	Marketable yie	numeric	grams per plar	g/plant			0	1000
Discarded Yield	DY	Discarded yiel	numeric	grams per plar	g/plant			0	1000
Average single fruit weight	AvSFW	Average fruit w	numeric	grams per frui	g/fruit			1	100
Balanced average single fruit weight	BalSFW	Calculated fron	numeric	grams per frui	g/fruit			1	100
Fruit size	FSI	Fruit size using	categorical				1: extremely small; 3: small; 5: medium; 7: large; 9: extrem		
Fruit size uniformity, Blueberry	FSIU_Vac	Fruit size unifo	categorical				1: extremely low; 3: low; 5: moderate; 7: high; 9: extremel		
Surface Color (CPVO 1-7), Strawberry	SCol_17	Surface color u	categorical				1: whitish yellow; 2: light orange; 3: medium orange; 4: ora		
Surface Color (CPVO 1-8), Raspberry	SCol_18	Surface color u	categorical				1: yellow; 2: orange; 3: light red; 4: medium red; 5: dark red		
Surface Color (Ctifl)	SCol_Ctifl	Surface color u	categorical				1; 2; 3; 4; 5; 6; 7; 8; 9		

INFO	METADATA	PHENOTYPES	DATA	RECORDING_DATES	ATTRIBUTES	COLLABORATORS	LOCATION
------	----------	------------	-------------	-----------------	------------	---------------	----------



DATA PREPARATION FOR UPLOAD

Data revision and upload



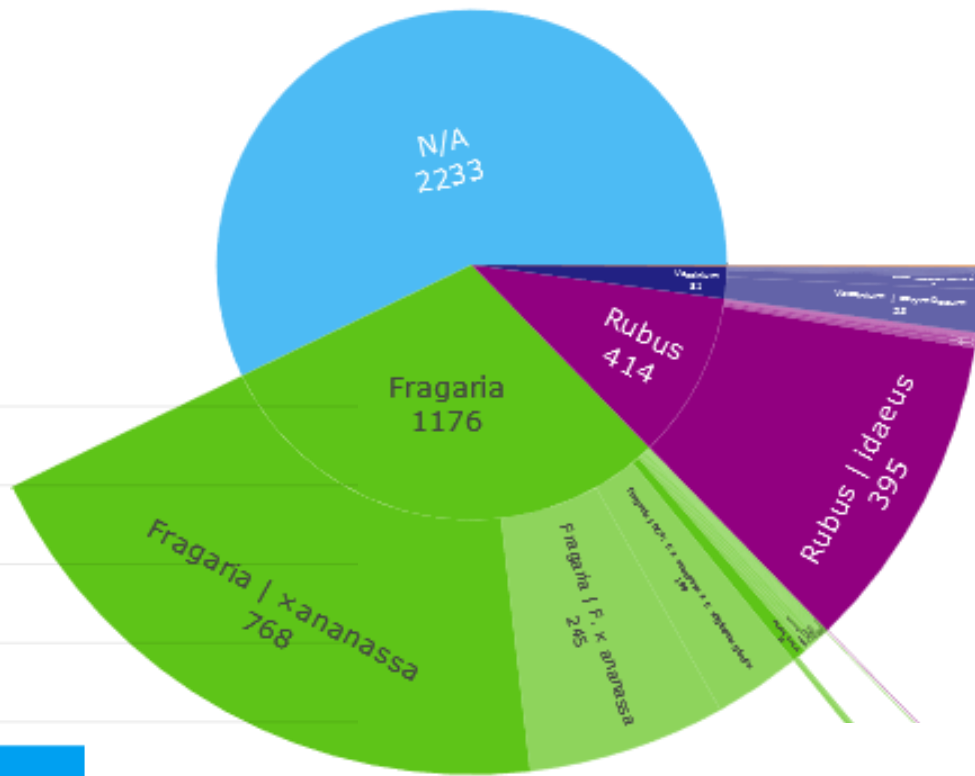
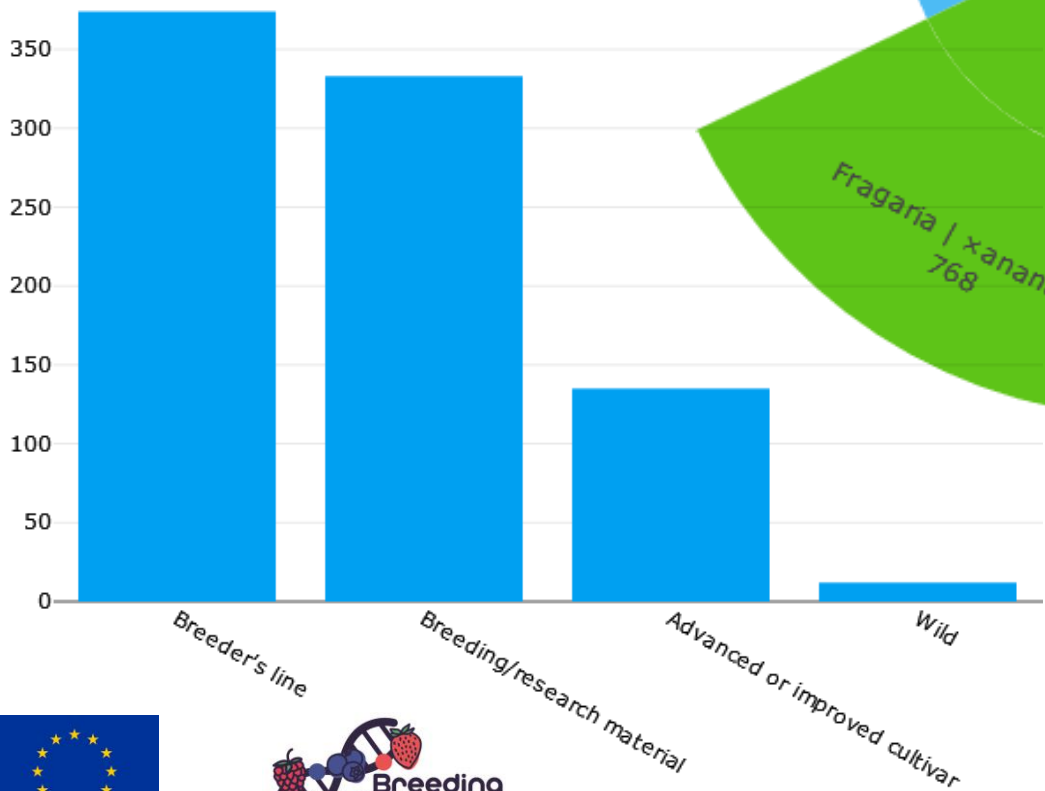
Source: Paul Shaw, JHI



BREEDINGVALUE BERRY DATA @ GERMINATE

Passport data

- Home
- Data**
 - Germplasm** 5K
 - Genotypic data
 - Markers 9K
 - Maps 2
 - Genotypic export 1
 - Trials data
 - Traits 69
 - Trials export 7
 - Climate data
 - Climate export 0
 - Datasets 9
 - Experiments 2
 - Data resources 19
 - Data stories 0
 - Data statistics
- Publications 2
- Search
- About







BREEDINGVALUE BERRY DATA @ GERMINATE





















Genotypic data

Maps

The following maps were found in this version of Germinate. To view markers contained on a map choose a map from the table below and click on its name.




[← Use the table filtering to search for specific items.](#)
Showing 1 to 2 of 2 records  100

Id	Map name	Map description	Marker count
1	2023 JHI Blueberry SNP Marker Data	2023 JHI Blueberry SNP Marker Data	4.011
2	2023 JHI Rapberry SNP Marker Data	2023 JHI Rapberry SNP Marker Data	4.989

-  Home
-  Data ^
-  Germplasm 5K
-  Genotypic data ^
-  Markers 9K
-  Maps 2
-  Genotypic export 1
-  Trials data ^
-  Traits 69
-  Trials export 7
-  Climate data ^
-  Climate export 0
-  Datasets 9
-  Experiments 2
-  Data resources 19
-  Data stories 0
-  Data statistics
-  Publications 2
-  Search
-  About <



BREEDINGVALUE BERRY DATA @ GERMINATE

Phenotypic data

Trials

Use the table filtering to search for specific items. Showing 1 to 7 of 7 records

<input type="checkbox"/>	Id	Dataset name	Dataset description	Experiment id	Experiment name	Dataset type	Data type	License	Data objects	Data points	Linked data resources	Ext.
<input type="checkbox"/>	2	2020 JHI Blueberry LAB field mapping population Tunnel J8A	A blueberry cross used for genetic mapping	2	Generic Experiment ⓘ	Trials		CC BY-SA	301	3.58K	Show	
<input type="checkbox"/>	3	2021 JHI Blueberry LAB field mapping population Tunnel J8A	A blueberry cross used for genetic mapping	2	Generic Experiment ⓘ	Trials		CC BY-SA	299	3.45K	Show	
<input type="checkbox"/>	4	2022 JHI Blueberry LAB field mapping population Tunnel J8A	A blueberry cross used for genetic mapping	2	Generic Experiment ⓘ	Trials		CC BY-SA	298	3.85K	Show	
<input type="checkbox"/>	17	2022 JHI Blueberry Bullionfield cultivars	Replicated commercial blueberry cultivars	2	Generic Experiment ⓘ	Trials		CC BY-SA	29	992	Show	
<input type="checkbox"/>	18	2022 JHI Blueberry replicated mapping population	Mapping progeny blueberry cultivars	2	Generic Experiment ⓘ	Trials		CC BY-SA	272	3.41K	Show	
<input type="checkbox"/>	19	2023 JHI Blueberry Bullionfield cultivars	Replicated commercial blueberry cultivars	2	Generic Experiment ⓘ	Trials		CC BY-SA	29	937	Show	
<input type="checkbox"/>	21	2023 JHI Blueberry LAB field mapping population Tunnel J8A	A blueberry cross used for genetic mapping	2	Generic Experiment ⓘ	Trials		CC BY-SA	293	3.78K	Show	

Use the checkboxes to select one or more rows.



Next

- Home
- Data**
- Germplasm 5K
- Genotypic data
- Markers 9K
- Maps 2
- Genotypic export 1
- Trials data**
- Traits 69
- Trials export 7
- Climate data
- Climate export 0
- Datasets 9
- Experiments 2
- Data resources 19
- Data stories 0
- Data statistics
- Publications 2
- Search
- About



BREEDINGVALUE BERRY DATA @ GERMINATE

Example for data export

Genotype data export

Selected datasets

Select germplasm groups

The exported files will contain the germplasm from the selected groups only.

Search

Marked Items (7)

Group selection All Items

Select map

Only markers on the map will be exported.

2023 JHI Raspberry SNP Marker Data (5K)

Select marker groups

The exported files will contain the markers from the selected groups only.

Search

Marked Items (0)

Group selection All Items

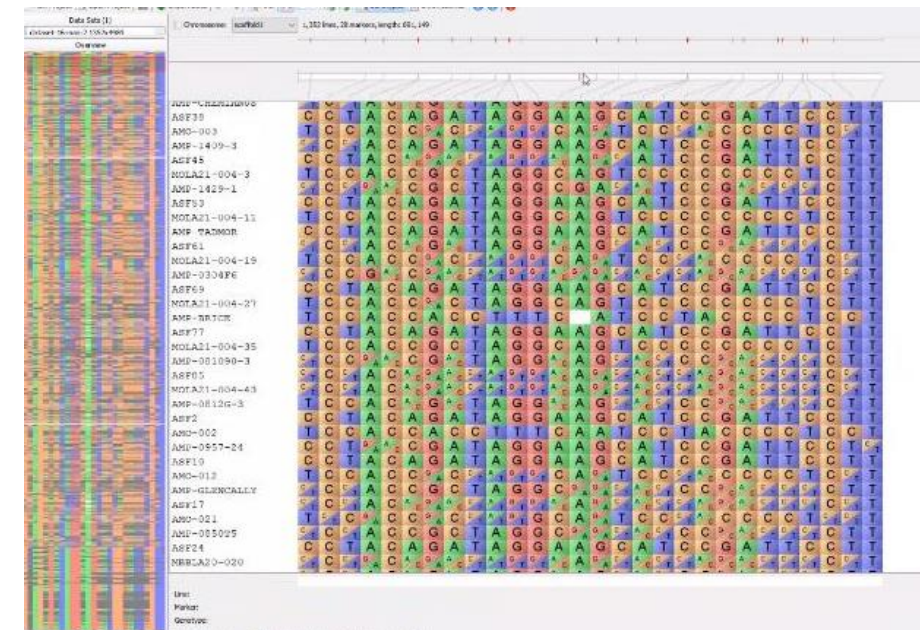
Generate flat text file?

By enabling this, Germinate will generate a flat tab-delimited file containing the genotypic data.

Yes

Generate Flapjack project?

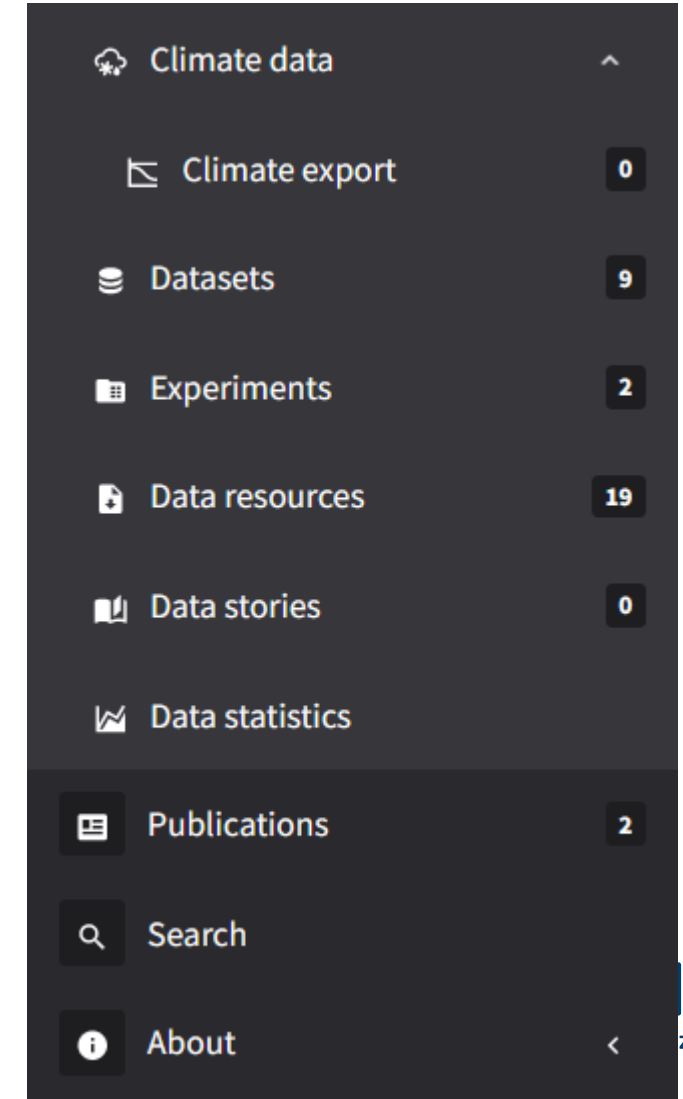
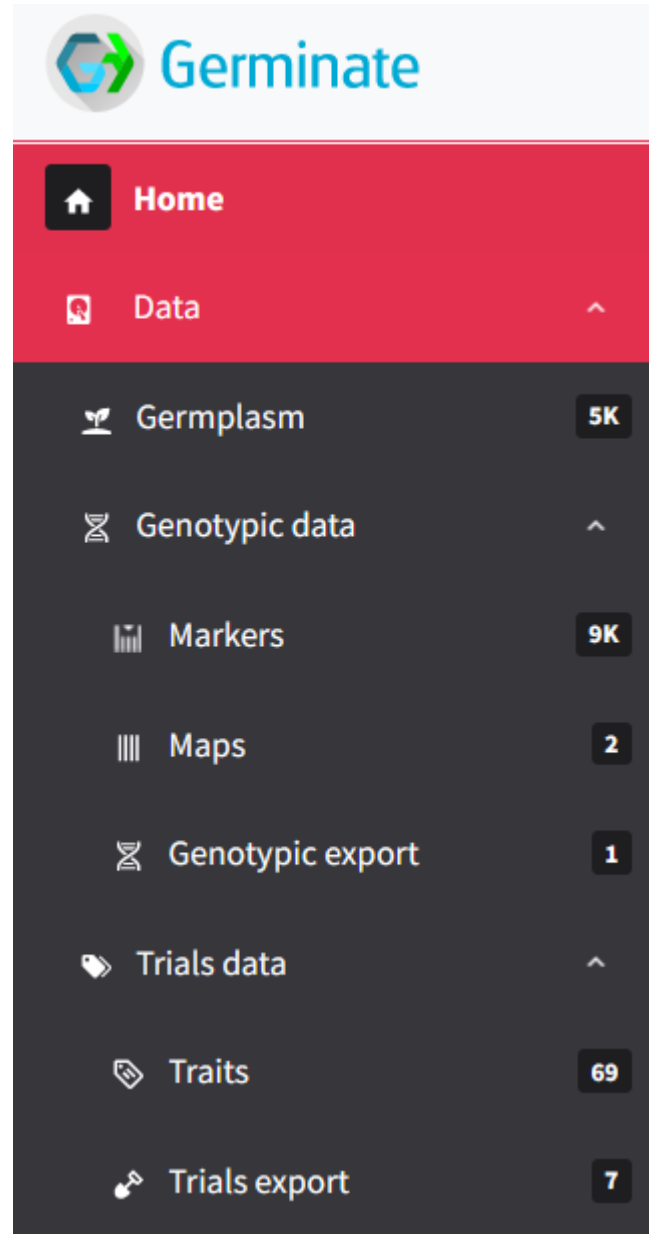
By enabling this, Germinate will generate a Flapjack project file in addition to the flat



BREEDINGVALUE BERRY DATA @ GERMINATE

Further data to come

- Data, data, data...
- Photographs of plants and fruits
- Metabolic profiles
- Location
- Climate



THANKS TO

- YOU ! ...FOR YOUR ATTENTION
- JAMES-HUTTON INSTITUTE ...FOR PROVIDING THE PLATFORM AND SOME SLIDES (PAUL SHAW, SEBASTIAN RAUBACH, DEPARTMENT OF INFORMATION AND COMPUTATIONAL SCIENCES, SUSAN MCCALLUM, DEPARTMENT OF CELL AND MOLECULAR SCIENCES, INVERGROWIE, SCTOLAND, UK
- BREEDINGVALUE PARTNERS ...FOR COLLABORATION AND PROVIDING DATA
- EU ...FOR FUNDING
- IBG-4 AT FZJ ...FOR TRAINING AND COMFORTABLE WORKING SPACE

