

Grassroots – an online data repository – what, why and how?



Andrew Riche
WGIN Stakeholders meeting
February 2023

Introduction



What:

- DFW is the national co-ordinated wheat research program
- Developed an online repository for field trials data, funded by BBSRC, but not limited to BBSRC funded projects

Why:

- To provide secure long-term easy access to data
- FAIR data available to all

How:

- Web based

Where:

- Development ongoing but ready to use

A screenshot of the Grassroots Infrastructure website. The header includes the title 'Grassroots Infrastructure' and navigation links for 'SERVICES', 'DATA', 'DOCS', 'BLOG', 'DFW', and 'SEARCH'. The Earham Institute logo is in the top right. The main content area has a green background with a wheat field image and text describing the project's goals and technical details. A teal section at the bottom is titled 'SERVICES' with a star icon and lists existing services.

Grassroots Infrastructure SERVICES DATA DOCS BLOG DFW SEARCH Earham Institute

The Grassroots Infrastructure project aims to create an easily-deployable suite of computing middleware tools to help users and developers gain access to scientific data infrastructure that can easily be interconnected.

With the data-generative approaches that are increasingly common in modern life science research, it is vital that the data and metadata produced by these efforts can be shared and reused. The Grassroots Infrastructure project wraps up industry-standard software tools with a consistent API that can be federated on a number of levels. This means institutions and groups can deploy a simple lightweight virtual machine, expose local data, connect up any existing data services, and federate their instance of the Grassroots with others out-of-the-box.

The Grassroots Infrastructure uses a controlled vocabulary of JSON messages to communicate, so any server or client that can understand JSON can be used to access and connect to the platform. We provide infrastructure to ensure that the scientific data remains the important factor, and not the worry about how to build a system to expose your data.

SERVICES

The Grassroots Infrastructure has been developed to expose a number of important and powerful technologies that can benefit researchers and developers to realise their data management needs.

Existing services:

- [Search Grassroots](#)
- [Field Trials](#)
- [Submitting Field Trials](#)
- [Blast searches](#)
- [Grassroots Data Portal](#)
- [Yellow Rust Field Pathogenomics](#)
- [List of all public services](#)



What? (1)

- Funded as part of DFW WP4
- Development on-going – your involvement means ideas are taken on board
- Web based database for storing data, images, meta data
- Open source code
- Could be run as a local instance or use the Earlham system

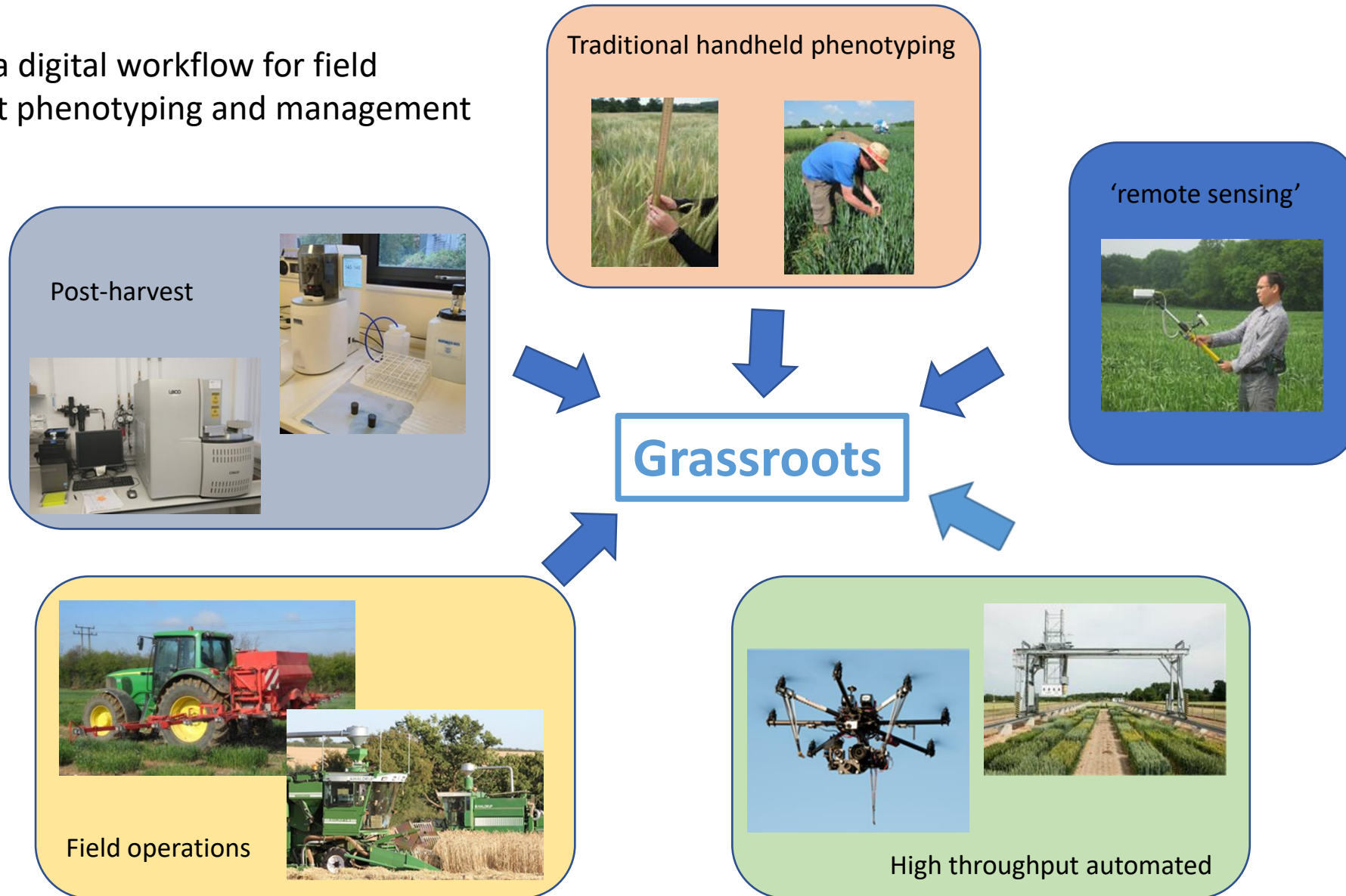


Within DFW many field trials across the JIC, NIAB, Rothamsted and University of Nottingham



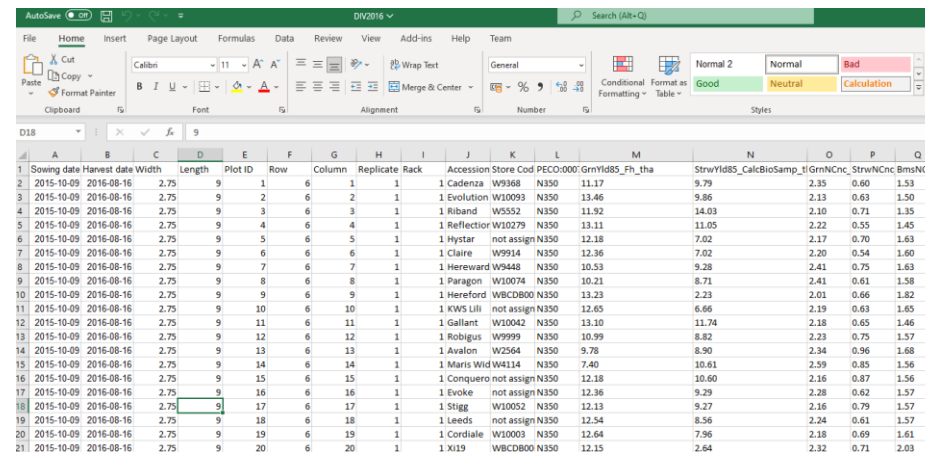
What? (2)

- Central to a digital workflow for field experiment phenotyping and management

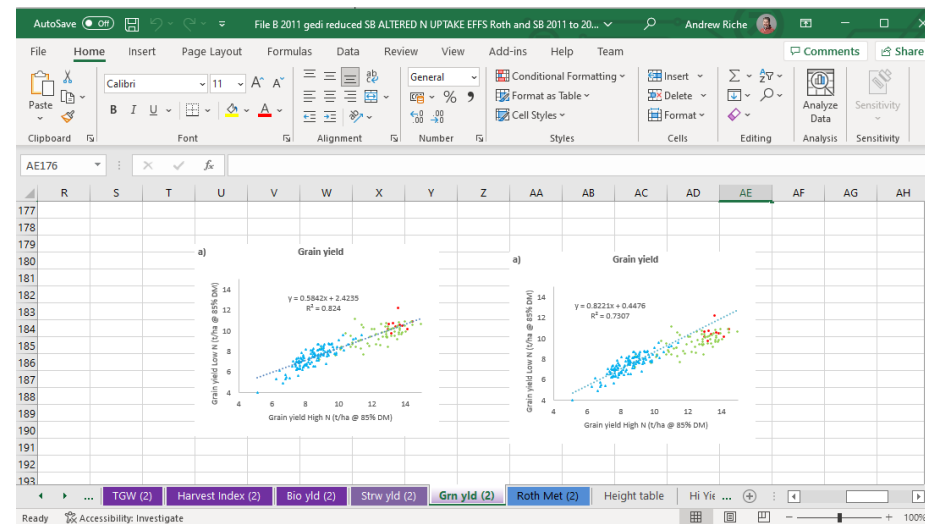


Why? (1)

- Traditional file systems often far from ideal!
- Excel can't cope!
- Dispersed data locations
- A lot of data collected is not easy to access
- Diverse sources of data
- Enforces use of ontologies
- High throughput phenotyping leads to huge datasets
- UAV and other image based systems produce data an order of magnitude greater



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Sowing date	Harvest date	Width	Length	Plot ID	Row	Column	Replicate	Rack	Accession	Store Cod	PECCO000	GrnYld85_Fh_tha	StrwYld85_CalcBioSamp_1	GrnNCnc_StrwNCnc	BmsNCnc	
2	2015-10-09	2016-08-16	2.75	9	1	6	1	1	1	Cadenza	W9368	N350	11.17	9.79	2.35	0.60	1.53
3	2015-10-09	2016-08-16	2.75	9	2	6	2	1	1	Evolution	W10093	N350	13.46	9.86	2.13	0.63	1.50
4	2015-10-09	2016-08-16	2.75	9	3	6	3	1	1	Riband	W5552	N350	11.92	14.03	2.10	0.71	1.35
5	2015-10-09	2016-08-16	2.75	9	4	6	4	1	1	Reflector	W10279	N350	13.11	11.05	2.22	0.55	1.45
6	2015-10-09	2016-08-16	2.75	9	5	6	5	1	1	Hystar	not assign	N350	12.18	12.18	2.17	0.70	1.63
7	2015-10-09	2016-08-16	2.75	9	6	6	6	1	1	Clare	W9914	N350	12.36	7.02	2.20	0.54	1.60
8	2015-10-09	2016-08-16	2.75	9	7	6	7	1	1	Hereford	W9440	N350	10.53	9.28	2.41	0.75	1.63
9	2015-10-09	2016-08-16	2.75	9	8	6	8	1	1	Paragon	W10074	N350	10.21	8.71	2.41	0.61	1.58
10	2015-10-09	2016-08-16	2.75	9	9	6	9	1	1	Hereford	WBCDB00	N350	13.23	2.23	2.01	0.66	1.82
11	2015-10-09	2016-08-16	2.75	9	10	6	10	1	1	KWS Lili	not assign	N350	12.65	6.66	2.19	0.63	1.65
12	2015-10-09	2016-08-16	2.75	9	11	6	11	1	1	Gallant	W10042	N350	13.10	11.74	2.18	0.65	1.46
13	2015-10-09	2016-08-16	2.75	9	12	6	12	1	1	Robigus	W9999	N350	10.99	8.82	2.23	0.75	1.57
14	2015-10-09	2016-08-16	2.75	9	13	6	13	1	1	Avalon	W2564	N350	9.78	8.90	2.34	0.96	1.68
15	2015-10-09	2016-08-16	2.75	9	14	6	14	1	1	Marris Wild	W4114	N350	7.40	10.61	2.59	0.85	1.56
16	2015-10-09	2016-08-16	2.75	9	15	6	15	1	1	Conquero	not assign	N350	12.18	10.60	2.16	0.87	1.56
17	2015-10-09	2016-08-16	2.75	9	16	6	16	1	1	Evolve	not assign	N350	12.36	9.29	2.28	0.62	1.57
18	2015-10-09	2016-08-16	2.75	9	17	6	17	1	1	Stigg	W10052	N350	12.13	9.27	2.16	0.79	1.57
19	2015-10-09	2016-08-16	2.75	9	18	6	18	1	1	Leeds	not assign	N350	12.54	8.56	2.24	0.61	1.57
20	2015-10-09	2016-08-16	2.75	9	19	6	19	1	1	Cordiale	W10003	N350	12.64	7.96	2.18	0.69	1.61
21	2015-10-09	2016-08-16	2.75	9	20	6	20	1	1	Xi19	WBCDB00	N350	12.15	2.64	2.32	0.71	2.03



200Mb, 128 sheets



Why? (2) – UAV case study

Conventional RGB+NIR

Hyperspectral

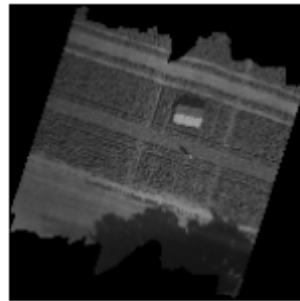


9Gb/flight

200 Gb/flight

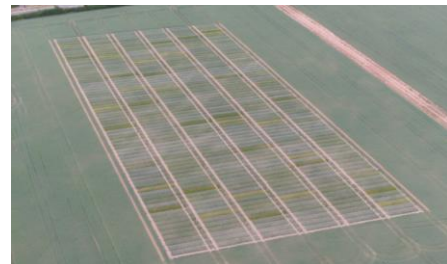


3 bands 3D



540 bands

70*170m



- High throughput phenotyping leads to large data sets
- RGB 0.25Tb/yr
- Hyperspectral 5Tb data/expt/yr
- Efficient data pipelines and storage essential
- Need to store both the images and data
- Why store locally when central would be most efficient?

How? Data entry



A location has:
Name
Street
Town
County
Country
Post code
Coordinates
Altitude
Soil type

SUBMITTING FIELD TRIAL PLOTS

The form for submitting or editing existing Plots is available at https://grassroots.tools/private/service/field_trial-submit_plots

Plots can be added to the study division table.

SUBMITTING A LOCATION

The form for submitting or editing an existing Location is available at https://grassroots.tools/private/service/field_trial-submit_location.

A Location defines the geographical location of the field where a Study takes place. The location can be defined by a normal postal address or GPS coordinates. The Grassroots system has libraries for determining GPS coordinates from any address data that is submitted, although the accuracy will depend upon how much data is given. You can alter the GPS coordinates and altitude of the location yourself should the information that is discovered automatically is not correct or not accurate.

The pieces of information that you can enter are:

- **Name ***: The name of this Location. This is a required field.
- **Street**: The street for this Location.
- **Town**: The town for this Location.
- **County**: The county for this Location.
- **Country**: The country for this Location.
- **Postal code**: The town for this Location.
- **Supply your own GPS coordinates**: Tick this box if you wish to override the GPS coordinates
- **Latitude**: If *Supply your own GPS coordinates* is ticked, then this will specify the latitude of this Location in decimal degrees e.g. 41.40338
- **Longitude**: If *Supply your own GPS coordinates* is ticked, then this will specify the longitude of this Location in decimal degrees e.g. -12.87665
- **Altitude**: If *Supply your own GPS coordinates* is ticked, then this will specify the altitude of this Location in metres.

SUBMITTING PROGRAMME
The form for submitting or editing an existing Programme is available at https://grassroots.tools/private/service/field_trial-submit_programme

Lead Programme: If you wish to edit an existing Programme, choose it from this list and its values will be loaded.

A Programme consists of the following pieces of information:

- **Name ***: The name of the Programme.
- **Abbreviation**: An abbreviation for the Programme.
- **Code**: This is the code for the Programme. It is identifying.
- **Objective**: This is a description of the objective of the Programme.
- **What address**: If there is a web page associated with the Programme, please enter it here.
- **Principal Investigator Name ***: This is the name of the person for this Programme.
- **Principal Investigator Email ***: This is the email of the person for this Programme.
- **Image**: The web address of the image for the Programme top. If the one you wish to use doesn't have a web address, get it from within a web browser and it will be uploaded to the grassroots tool website and the data will be added to your data base.

SUBMIT FIELD TRIAL PROGRAMME
Please use the form to submit a new Programme or to edit an existing Programme.

Programme Name:

Abbreviation:

Code:

Objective:

What address:

Principal Investigator Name:

Principal Investigator Email:

Image:

Submit Cancel

SUBMITTING TRIALS
This form for submitting or editing an existing Field Trial is available at https://grassroots.tools/private/service/field_trial-submit_trials

Lead Field Trial: If you wish to edit an existing Field Trial, choose it from this list and its values will be loaded.

A Trial consists of the following pieces of information:

- **Name ***: The name of the Field Trial.
- **Name**: The name or group that you are conducting the Field Trial.
- **Programme ***: This is the Programme that this Field Trial is part of.

SUBMIT FIELD TRIALS
Add a new trial to the system. Existing trials can be edited in an existing Field Trial contains multiple Studies. This is available in investigation records.

Show studies Add new studies

Name:

Programme:

Town:

SUBMITTING STUDIES
The form for submitting or editing an existing Study is available at https://grassroots.tools/private/service/field_trial-submit_studies

A Study is an experiment that takes place at a location to measure a variety of phenotypes. It is one of the main forms that you will need to complete and has a number of fields:

- **Lead Study**: If you wish to edit an existing Study, choose it from this list and its values will be loaded.
- **Name ***: The name of the Study.
- **Field Trial ***: This is the Field Trial that this Study belongs to. If the Field Trial that you require is not in this list, then you'll need to create it by following the instructions for [creating a Field Trial](#).
- **Location ***: This is the location where the Study is taking place. If the Location that you require is not in this list, then you'll need to create it by following the instructions for [creating a Location](#).
- **Sowing date**: This is the date when the seeds are sown in the field.
- **Harvest date**: This is the date when the crop is harvested. The date can be entered on the left hand page which some this field can be left blank.
- **Growing Conditions**: This is a free text field where you can describe the growing conditions for this Study can be placed.
- **Design**: This is a free text field where you can describe the experimental design of this Study can be placed.
- **Phenotype gathering notes**: This is a free text field where you can describe the collection of phenotypes for this Study can be placed.
- **Weather**: If there is a system with the weather details for this Study, enter the webpage for this.
- **Photo taken**: This is a free text field where you can enter the photos for this Study.
- **Data not included**: This is a free text field where you can describe data collected but not currently stored within Grassroots.
- **Photos not included**: This is a free text field where you can describe photos taken but not currently stored within Grassroots.
- **Crop**: This is a free text field where you can describe the crop that is being grown in this field.
- **Fieldwork stop**: This is the crop that was previously seen in the field that this Study is in. If the required crop is missing, please contact us for help in establishing the system.
- **pH alkalinity**: If the pH levels of the soil are known, enter the alkalinity value here.
- **pH acidity**: If the pH levels of the soil are known, enter the acidity value here.

SUBMITTING FIELD TRIAL PLOTS
The form for submitting or editing existing Field Trial Plots is available at https://grassroots.tools/private/service/field_trial-submit_plots

Plots can be added to the study division table.

SUBMIT FIELD TRIAL PLOTS
Please use the form to submit a new Field Trial Plot or to edit an existing Field Trial Plot.

Field Trial Plot Name:

Field Trial:

Location:

Sowing Date:

Harvest Date:

Growing Conditions:

Design:

Phenotype Gathering Notes:

Weather:

Photo Taken:

Data Not Included:

Photos Not Included:

Crop:

Fieldwork Stop:

pH Alkalinity:

pH Acidity:

Submit Cancel

SUBMITTING A LOCATION
The form for submitting or editing an existing Location is available at https://grassroots.tools/private/service/field_trial-submit_location.

A Location defines the geographical location of the field where a Study takes place. The location can be defined by a normal postal address or GPS coordinates. The Grassroots system has libraries for determining GPS coordinates from any address data that is submitted, although the accuracy will depend upon how much data is given. You can alter the GPS coordinates and altitude of the location yourself should the information that is discovered automatically is not correct or not accurate.

The pieces of information that you can enter are:

- **Name ***: The name of this Location. This is a required field.
- **Street**: The street for this Location.
- **Town**: The town for this Location.
- **County**: The county for this Location.
- **Country**: The country for this Location.
- **Postal code**: The town for this Location.
- **Supply your own GPS coordinates**: Tick this box if you wish to override the GPS coordinates
- **Latitude**: If *Supply your own GPS coordinates* is ticked, then this will specify the latitude of this Location in decimal degrees e.g. 41.40338
- **Longitude**: If *Supply your own GPS coordinates* is ticked, then this will specify the longitude of this Location in decimal degrees e.g. -12.87665
- **Altitude**: If *Supply your own GPS coordinates* is ticked, then this will specify the altitude of this Location in metres.

How? Interrogation



- Web based
- Access to all on FAIR (Findable, Accessible, Interoperable, Reusable) principles
- Open to whole community
- Standard ontologies
- Tools for data extraction
- Currently 142 studies entered
- Download complete datasets
- Can be interrogated via a mobile phone in the field

Treatment Factors

Pesticide Exposure

Grassroots Infrastructure

SERVICES DOCS DATA ABOUT BL

BROWSE FIELD TRIALS STUDY

My Location

Plot ID	Sowing date	Harvest date	Width	Length	Row	Column	Replicate	Rack	Accession	PECO:000	PECO:000	StrwSOFF	BmsSCnc	BmsASOFF	StrwMgOf	MnH_Cal	GmrcCnc	GmNacCnc	BmsCoFF	StrwAscN	Spk
1	09/10/2018	01/09/2019	1.8	9	6	3	1	1	KWS Zyatt RFP	N350	7.521694	0.115074	5.054926	39.92212	0.000483	0.002106	23.52791			45	
2	09/10/2018	01/09/2019	1.8	9	6	4	1	1	Robigus RFP	N350	4.674834	0.142965	1.627543	36.04821	0.000972	0.001958	9.443784			27	
3	09/10/2018	01/09/2019	1.8	9	6	5	1	1	Mercia RFP	N350	9.672766	0.130172	0.009718	3.335313	51.83088	0.00014	0.002146	28.05858	6.58868	54	
4	09/10/2018	01/09/2019	1.8	9	6	6	1	1	Solstice RFP	N350	9.898397	0.174195	3.880157	27.15484	0.000173	0.001842	24.12794			35	
5	09/10/2018	01/09/2019	1.8	9	6	7	1	1	Crusoe RFP	N350	9.835071	0.137475	4.235676	41.62705	0.000241	0.002021	23.29902			51	
6	09/10/2018	01/09/2019	1.8	9	6	8	1	1	Hereward RFP	N350	10.448608	0.141907	4.371226	41.50631	0.000118	0.001671	27.09774			49	
7	09/10/2018	01/09/2019	1.8	9	6	9	1	1	Malacca RFP	N350	8.414942	0.129224	4.681025	43.85665	7.69687	0.002015	24.35707			55	
8	09/10/2018	01/09/2019	1.8	9	6	10	1	1	Soissons RFP	N350	8.722628	0.127078	5.709529	50.08997	8.21826	0.002019	25.23055			52	
9	09/10/2018	01/09/2019	1.8	9	6	11	1	1	Riband RFP	N350	10.92145	0.156675	4.43775	36.98736	6.05613	0.001689	30.04453			41	
10	09/10/2018	01/09/2019	1.8	9	6	12	1	1	Cadenza RFP	N350	8.650728	0.129903	4.783775	45.43221	4.2892	0.002256	20.82185			50	
11	09/10/2018	01/09/2019	1.8	9	6	13	1	1	Graham RFP	N350	9.855605	0.126906	4.253906	39.17892	0.003005	0.001704	28.02961			60	
12	09/10/2018	01/09/2019	1.8	9	6	14	1	1	KWS Barre RFP	N350	9.318384	0.12941	4.85244	43.42935	0.000174	0.001709	21.50593			49	
13	09/10/2018	01/09/2019	1.8	9	6	15	1	1	Claire RFP	N350	3.018527	0.158896	0.00135	1.326356	52.61387	0.001447	0.001522	6.487578	1.72724	1	
14	09/10/2018	01/09/2019	1.8	9	6	16	1	1	Xi19 RFP	N350	9.282837	0.142674	0.00063	4.841973	41.92527	4.6621	0.001871	21.23884	6.06582	40	
15	09/10/2018	01/09/2019	1.8	9	6	17	1	1	Avalon RFP	N350	8.560368	0.160977	0.004809	3.092209	37.08024	5.07946	0.00163	18.70857	5.7743	33	
16	09/10/2018	01/09/2019	1.8	9	6	18	1	1	Siskin RFP	N350	9.18638	0.130632	4.488379	47.56575	6.52662	0.001528	27.90789			47	
17	09/10/2018	01/09/2019	1.8	9	6	19	1	1	Istabraq RFP	N350	8.975572	0.122514	3.901512	44.83847	3.22614	0.001587	24.14825			44	
18	09/10/2018	01/09/2019	1.8	9	6	20	1	1	Hylux RFP	N350	8.281668	0.139454	3.792903	42.11396	5.56136	0.001591	21.16428			44	
19	09/10/2018	01/09/2019	1.8	9	6	21	1	1	Marris Wild RFP	N350	11.17985	0.137785	3.585202	41.34956	3.0987	0.001857	26.28926			38	
20	09/10/2018	01/09/2019	1.8	9	6	22	1	1	Paragon RFP	N350	8.188383	0.127377	3.056872	38.34941	4.86589	0.001958	19.72448			38	
21	09/10/2018	01/09/2019	1.8	9	6	27	2	1	Siskin SFP	N200	6.618574	0.105971	5.198453	44.53225	0.000525	0.001831	26.85327			42	
22	09/10/2018	01/09/2019	1.8	9	6	28	2	1	Riband SFP	N200	6.948967	0.111448	4.07341	45.59929	7.36712	0.001476	23.9837			45	
23	09/10/2018	01/09/2019	1.8	9	6	29	2	1	Paragon SFP	N200	8.364276	0.11396	3.689766	44.65429	0.00011	0.00196	25.95615			44	
24	09/10/2018	01/09/2019	1.8	9	6	30	2	1	Xi19 SFP	N200	7.871594	0.115197	5.651221	46.90144	8.7272	0.001384	25.87111			44	
25	09/10/2018	01/09/2019	1.8	9	6	31	2	1	Avalon SFP	N200	8.258784	0.120077	4.661216	44.97459	0.000159	0.001451	27.83643			4	
26	09/10/2018	01/09/2019	1.8	9	6	32	2	1	Hylux SFP	N200	7.307399	0.088728	5.305583	39.24676	7.76812	0.002116	27.87715			50	
27	09/10/2018	01/09/2019	1.8	9	6	33	2	1	Graham SFP	N200	8.197159	0.106345	4.737363	42.70668	4.21815	0.001295	31.26126			65	

Phenotype Gathering Notes:

Physical Samples Collected:

Data not included:

Trait Description: Weight of grains per spike

Trait Abbreviation: WtGrSp

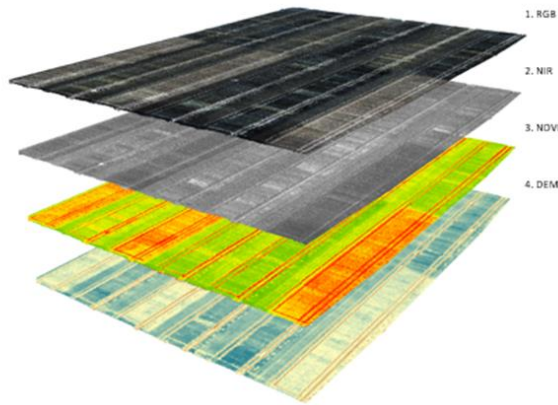
Measurement Name: Calculation from grab sampling data

Measurement Description: Weight of grains per spike, calculated from the pre-harvest grab-sample data (dry weight of grain divided by number of spikes)

Unit Name: g



Where?



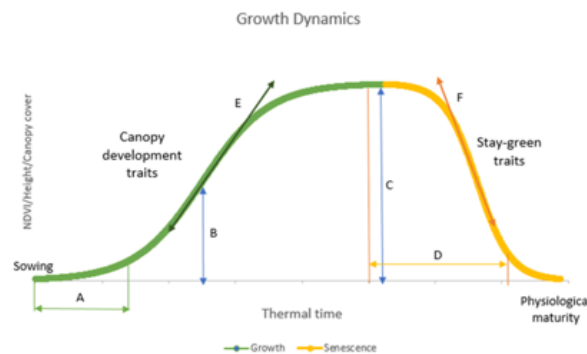
Benefits

- Using Grassroots is a great discipline in data organisation and standardisation
- Data centrally stored
- Making data FAIR

Developments

- Develop tools for data extraction
- Build in the ability to store and share large image files
- Develop to calculate other traits from 1st order data

2nd order traits



- Rates of increase and decrease
- Maxima & minima
- Time to reach key points

Who?



Rothamsted:

Malcolm Hawkesford
March Castle
Parul Sehrawat
Fenner Holman
Chris Rawlings
Richard Ostler



JIC:

Luzie Wingen

Earlham Institute:

David Swarbreck
Simon Tyrell
Daniel Olvera
Rob Davey
Xingdong Bian



Department
for Environment
Food & Rural Affairs

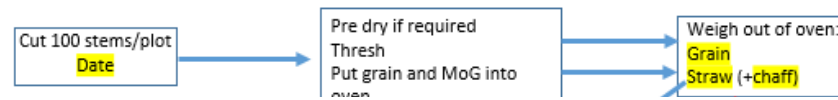


Biotechnology and
Biological Sciences
Research Council



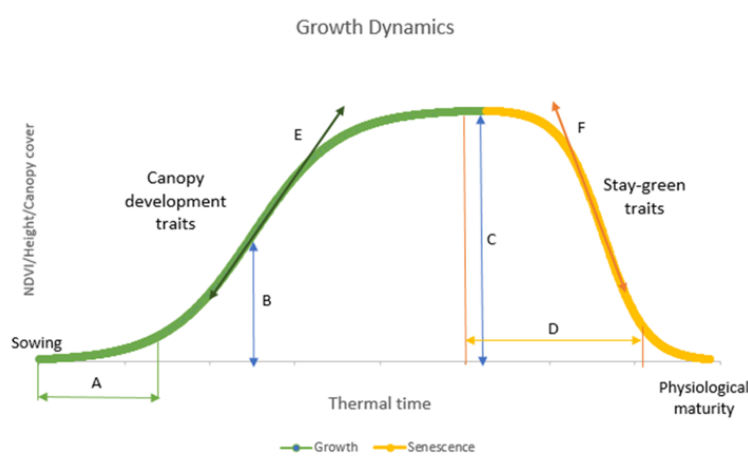
ROTHAMSTED
RESEARCH

Data & routine calculations

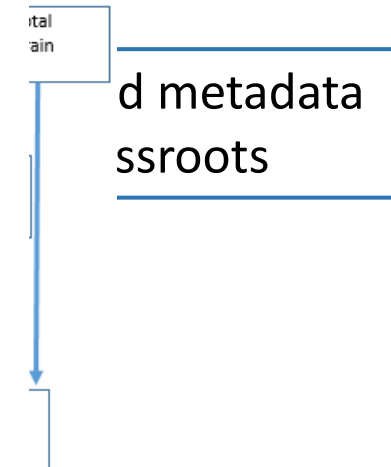


Measure plot length & width

Second order traits



- Rates of increase and decrease
- Maxima & minima
- Time to reach key points



Data from Grassroots

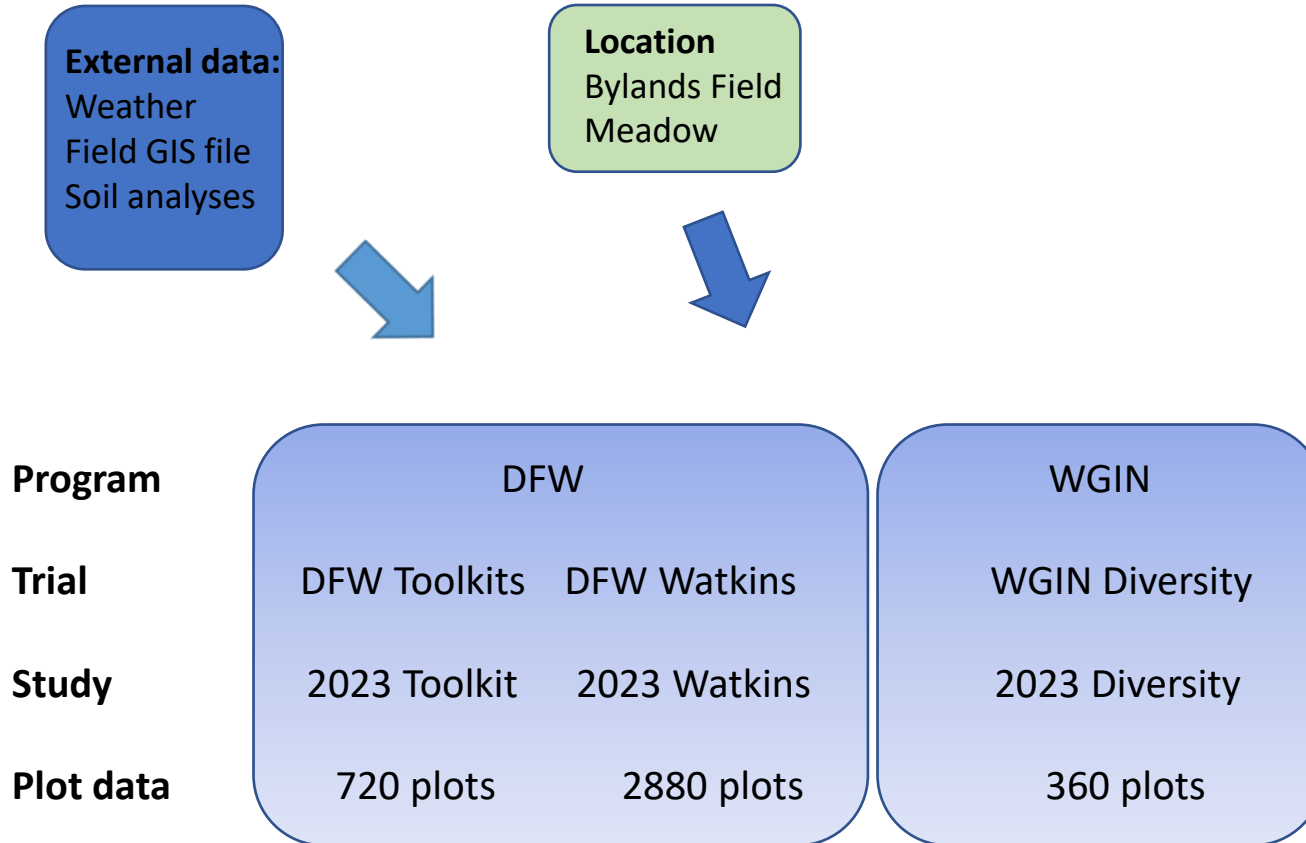


Calculate 2nd order trait data



Upload 2nd order plot data to Grassroots

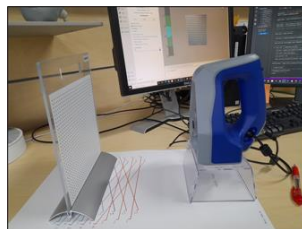
How? (1)



How? (1)



- Many systems producing large data sets



Introduction

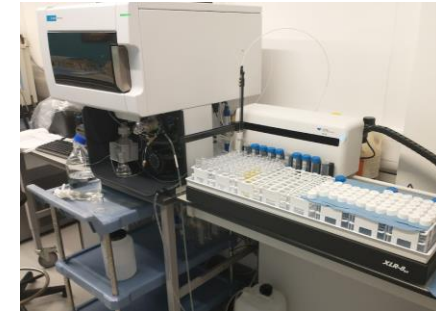
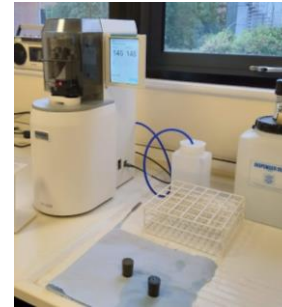


Efficient ways of arranging, storing and making data available are current & common issues across disciplines. Within Designing Future Wheat, the Grassroots database has been developed so that data is efficiently organised, using standardised terms, has adequate background information, and is available online. This resource will become increasingly useful, e.g. for meta-analyses as the database grows, and in a future project the capability of the system should be developed to handle new data formats, such as images, with bespoke tools for data extraction.

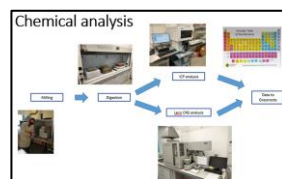
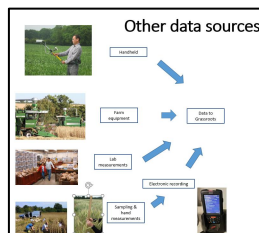
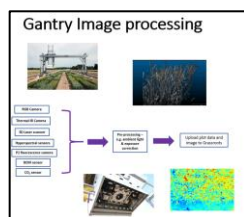
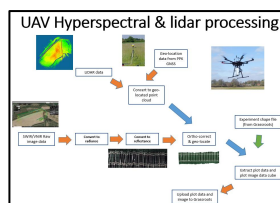
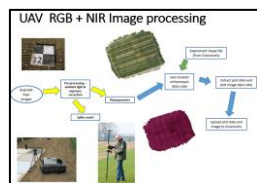
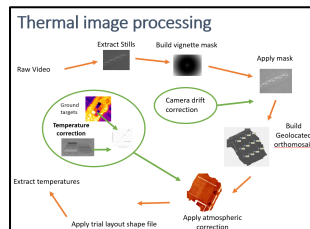
Grassroots: More than a database



1. Experiment proposal stage
2. Study planning
3. In-season phenotyping
4. Harvest & post-harvest data
5. Data integration

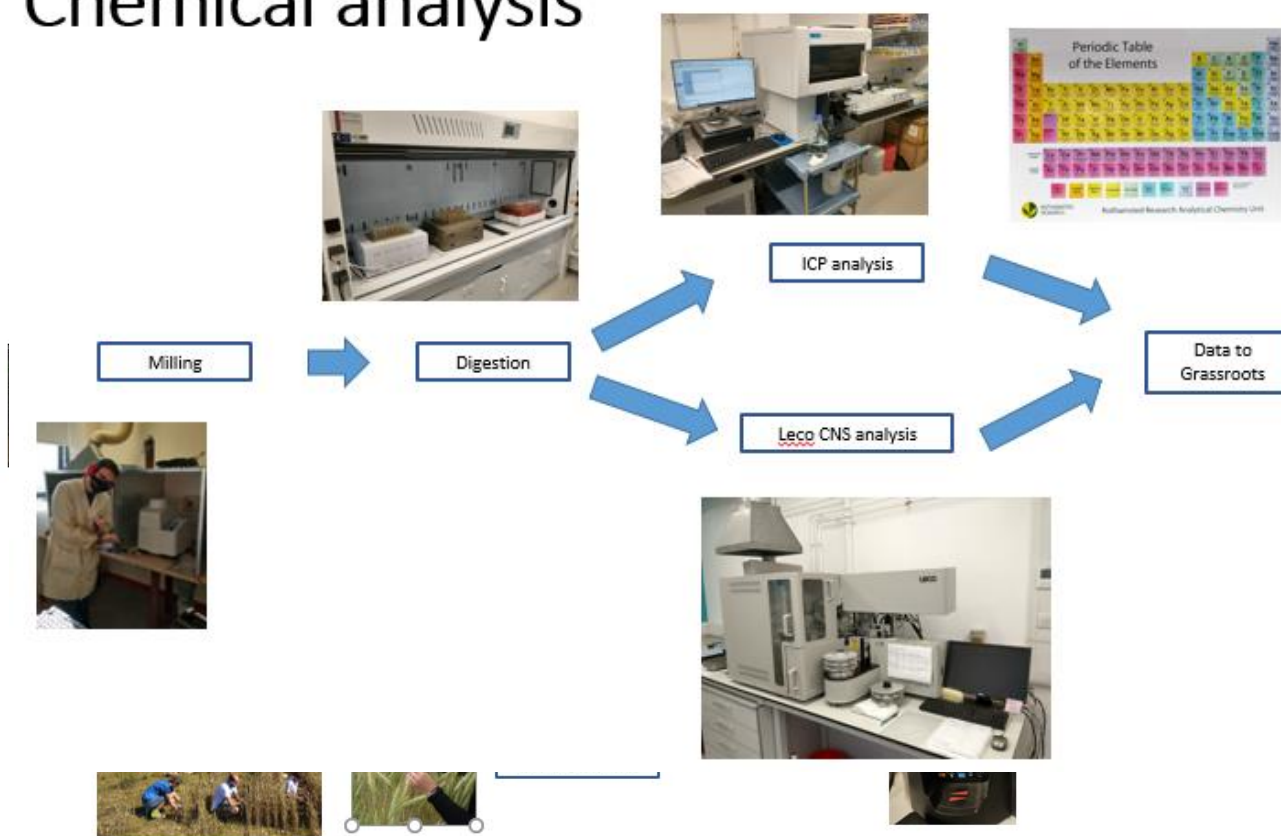


Data pipelines

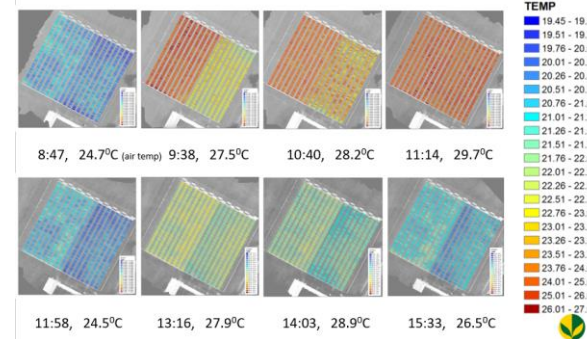
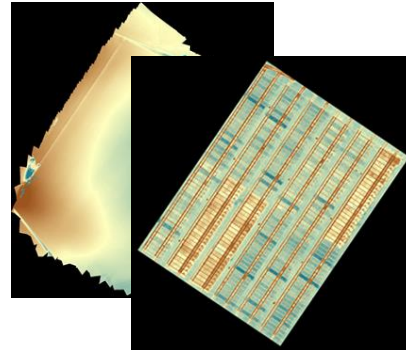
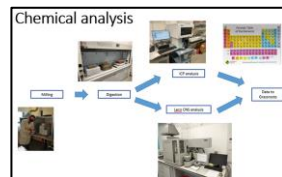
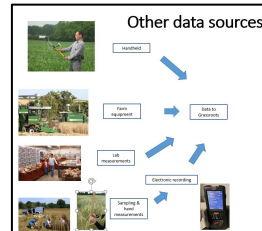
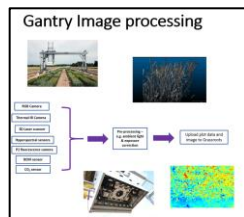
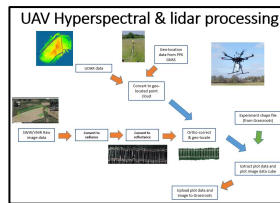
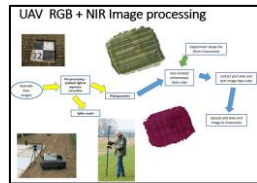
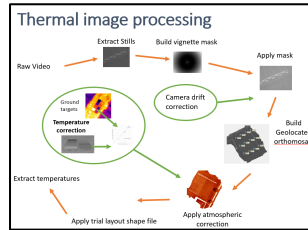


Chemical analysis

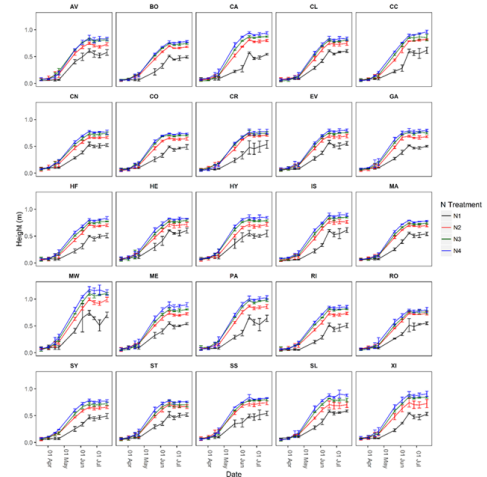
Other data sources



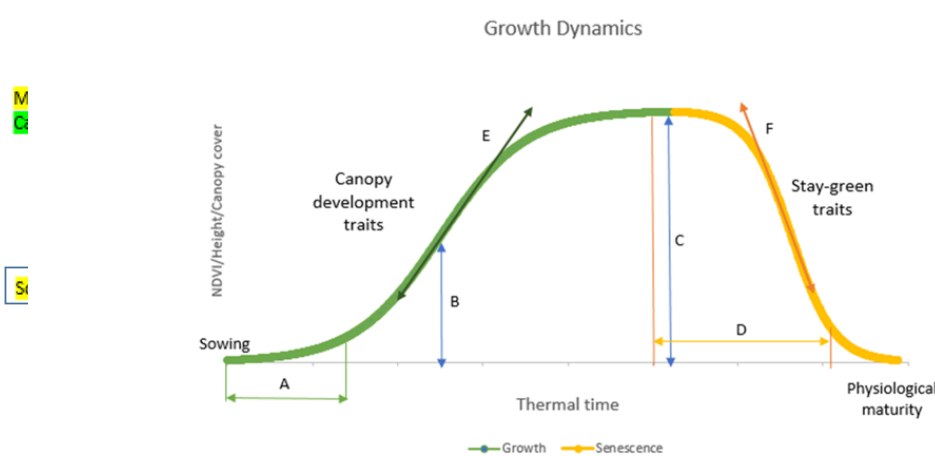
Data pipelines



Trait and metadata into Grassroots.... The end....

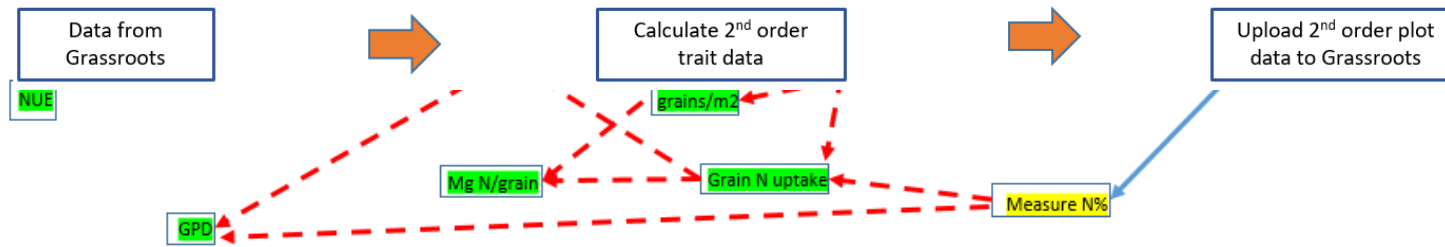


Second order traits



- Rates of increase and decrease
- Maxima & minima
- Time to reach key points

Additional metadata
Grassroots



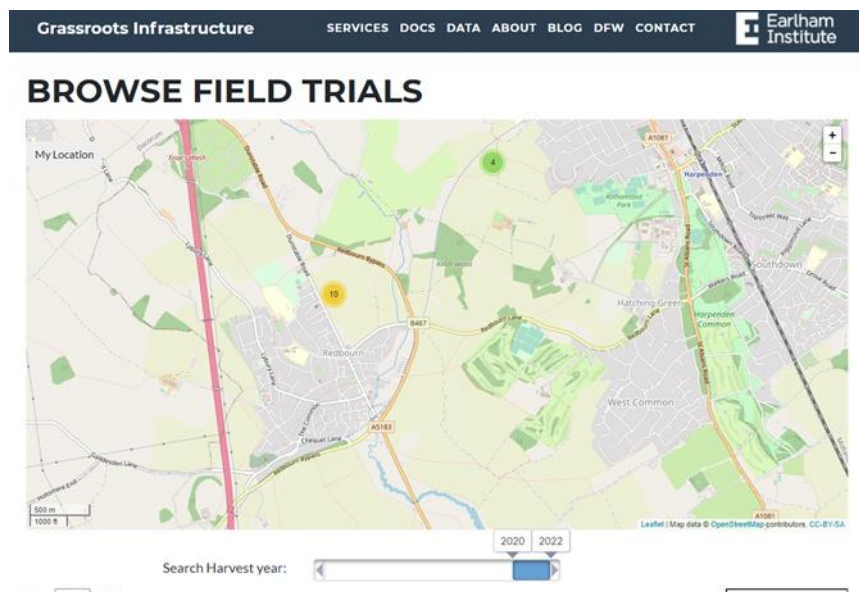
Grassroots online workshop



Tuesday 7th December
10:30 – 12:30

Email invite to all DFW

- Go through what grassroots offers
- How to upload data to Grassroots
- How to download data from grassroots
- Meeting to be recorded, and hoping for feedback on what works and what needs changing or adding



ROTHAMSTED
RESEARCH