

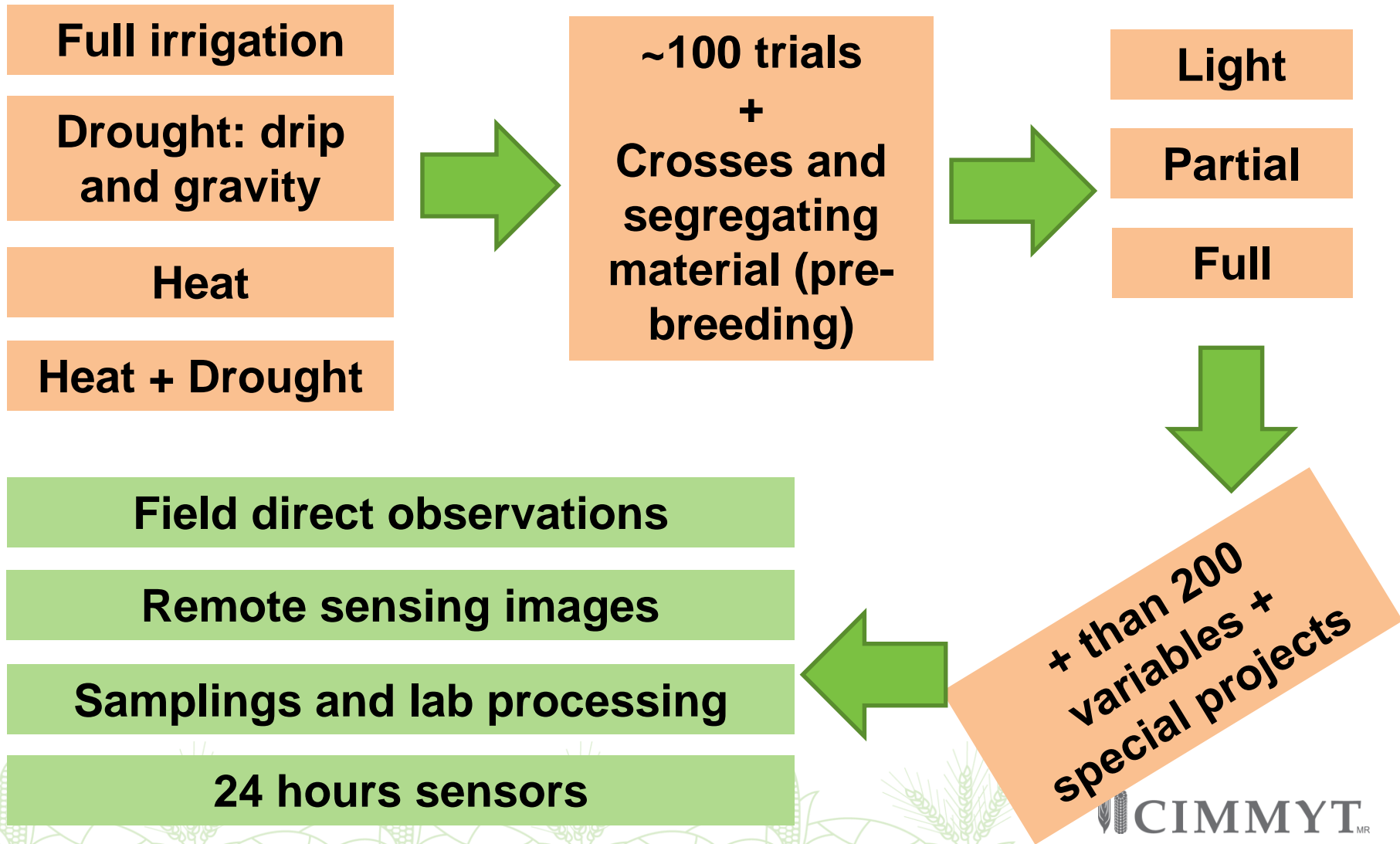
Data management at CIMMYT Wheat Physiology

Carolina Rivera

CSIRO-CIMMYT Collaborative workshop

12th June 2017

CIMMYT Wheat Physiology current capacity



Other types of data (metadata or information)

Environmental



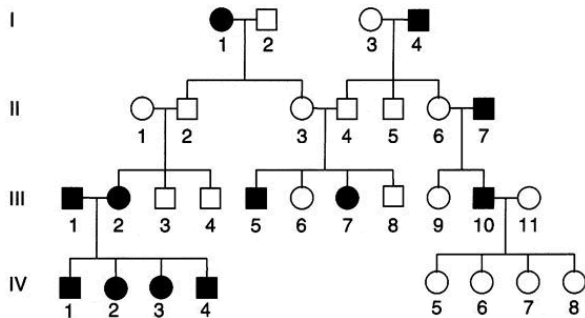
Trial management/Exp designs



Seed inventories



Pedigree/Genealogy



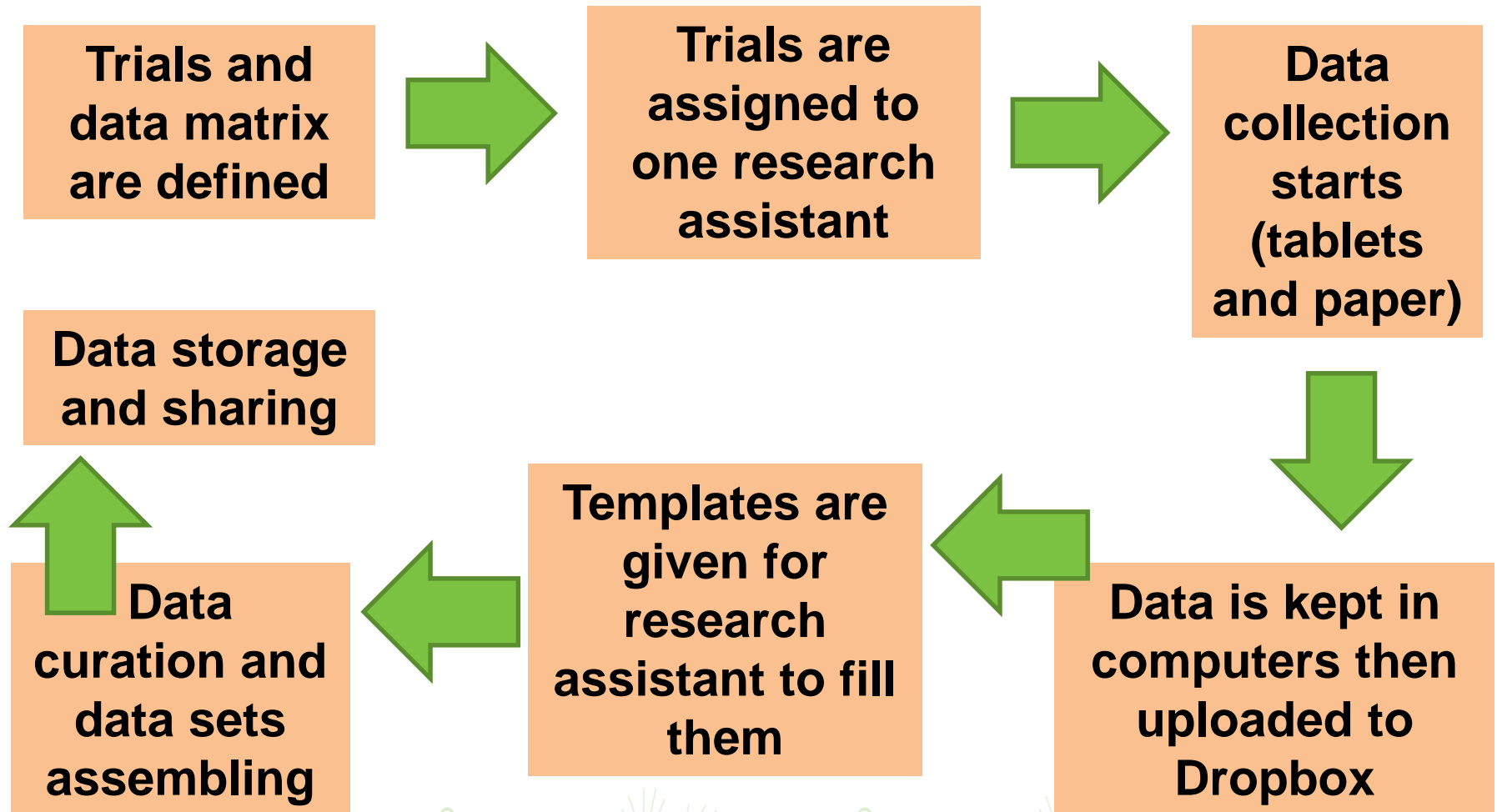
CID
SID
GID
Cross name
Selection history

Historical data

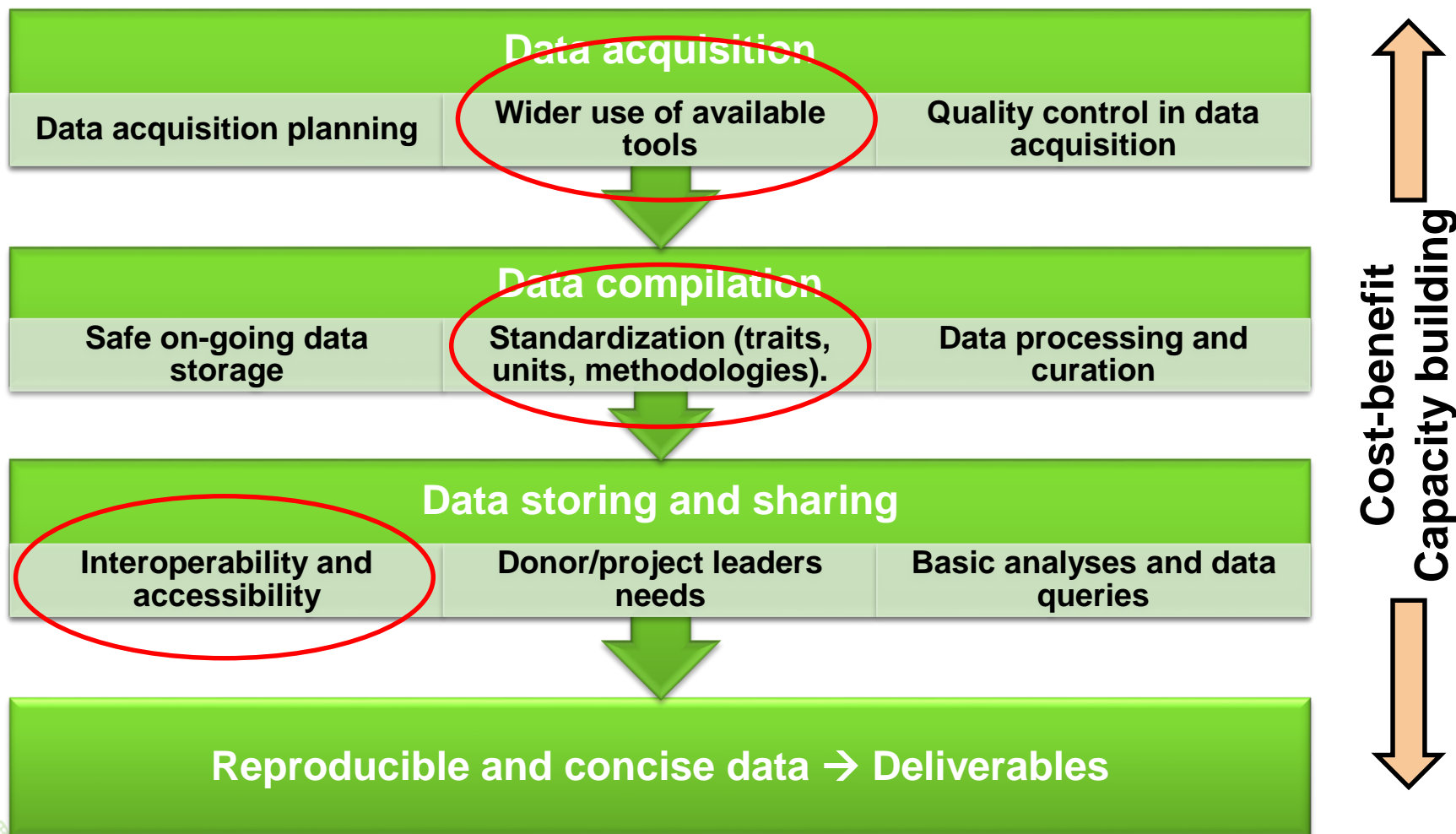


Methodologies

General data acquisition and management pipeline at CIMMYT WP



Data acquisition and management overview of opportunities for improvement



Specific challenges/opportunities in wheat data acquisition

- Centralized planning of measurements
- Wider use of available tools
- On going data quality control
- Capacity building (primary sources)

Increases accuracy and optimizes use of resources



Case: planning measurements and on-going data quality control

PS TAILS-RESPIRATION SET Y17_esp_1 - Excel

A1	PS TAILS RESPIRATION SET Y174M 2HIL												
	A	B	C	D	E	F	G	H	I	J	K	L	
1	PS TAILS RESPIRATION SET Y174M 2HIL						TIPO DE CORTE:		Marco=50 cm	marco de 50 cm			
2	ENCARGADO: PIMI							X	solo 2 reps (1y 2)	CORTADAS		PS F-PS S- 50 tallos subm y # espigas	
3	ESPIGAMIENTO						PART		Rep 3 no se corta	PROGRAMADAS			
4							CHOS			PROGRAMADAS			
6	TrialName	Plot	R	T	SubBlo	En	PlantingDa	Inicio de embuc	Espigamien	Fecha espigamien	Fecha espigamien +1	Fecha programada	Comments
7	PS TAILS-RESPIRATION SET Y17 - RGO	116	2	5	52	12/8/2016	43	55	1-Feb-17	13-Feb-17	13-Feb-17	13-Feb-16	
9	PS TAILS-RESPIRATION SET Y17 - RGO	52	1	7	52	12/8/2016	46	58	4-Feb-17	16-Feb-17	15-Feb-17		
10	PS TAILS-RESPIRATION SET Y17 - RGO	56	1	7	56	12/8/2016	50	58	4-Feb-17				
11	PS TAILS-RESPIRATION SET Y17 - RGO	142	2	8	44	12/8/2016	51	58	4-Feb-17				
12	PS TAILS-RESPIRATION SET Y17 - RGO	44	1	6	44	12/8/2016	48	59	5-Feb-17				
13	PS TAILS-RESPIRATION SET Y17 - RGO	104	2	3	56	12/8/2016	46	59	5-Feb-17				
17	PS TAILS-RESPIRATION SET Y17 - RGO	83	2	1	47	12/8/2016	50	60	6-Feb-17				
18	PS TAILS-RESPIRATION SET Y17 - RGO	91	2	2	55	12/8/2016	54	60	6-Feb-17				
19	PS TAILS-RESPIRATION SET Y17 - RGO	160	2	10	71	12/8/2016	52	60	6-Feb-17				
20	PS TAILS-RESPIRATION SET Y17 - RGO	55	1	7	55	12/8/2016	51	61	7-Feb-17				
21	PS TAILS-RESPIRATION SET Y17 - RGO	72	1	9	72	12/8/2016	51	61	7-Feb-17				
22	PS TAILS-RESPIRATION SET Y17 - RGO	126	2	6	54	12/8/2016	51	61	7-Feb-17				
23	PS TAILS-RESPIRATION SET Y17 - RGO	128	2	7	72	12/8/2016	52	61	7-Feb-17				



CORTES MARZO 2016 - Excel

	A	B	C	D	E	F	G	H	I	
1	Cortes del VIERNES 3 de marzo de 2016									
2										
3	ENSAYO	ENCARGADO	PLOT	FECHA ESPIGIANT	FECHA ESP+12+10IANT+7	DIA DE CORTE	41+XX SSM BIOMASA	20 CAMPO PART	31 CAMPO CHOS	
22	BACANORA-WEEBIL - RGO	RAYISAM	534			3-Mar-17	X			
23	BACANORA-WEEBIL - RGO	RAYISAM	531			3-Mar-17	X			
24	BACANORA-WEEBIL - RGO	RAYISAM	535			3-Mar-17	X			
25	BACANORA-WEEBIL - RGO	RAYISAM	536			3-Mar-17	X			
26	BACANORA-WEEBIL - RGO	RAYISAM	538			3-Mar-17	X			
27	BACANORA-WEEBIL - RGO	RAYISAM	564			3-Mar-17	X			
28	BACANORA-WEEBIL - RGO	RAYISAM	567			3-Mar-17	X			
29	BACANORA-WEEBIL - RGO	RAYISAM	571			3-Mar-17	X			
30	BACANORA-WEEBIL - RGO	RAYISAM	586			3-Mar-17	X			
31	BACANORA-WEEBIL - RGO	RAYISAM	535			3-Mar-17	X			
32										
33	HIBAPY16 (SEL GEMMA)	JAZMIN	FINALIZADO							
34										
35	PADMAPY14 - RGO	MIRIAM	FINALIZADO							
36										
37	QeRIP - RGO	GERARDO	NO HAY							
38										
39	PNL HAR - RGO	SYHARA	337		3-Mar-17	3-Mar-17		X		
40	PNL HAR - RGO	SYHARA	338		3-Mar-17	3-Mar-17		X		
41	PNL HAR - RGO	SYHARA	470		3-Mar-17	3-Mar-17		X		
42	PNL HAR - RGO	SYHARA	439		3-Mar-17	3-Mar-17		X		
43	PNL HAR - RGO	SYHARA	718		4-Mar-17	3-Mar-17		X		
44										
45	4 WYCYT + PDS 4M2 HLRGO	ARTURO	FINALIZADO							
46										
47	L3 WYCYT + PADRES - RGO	SYHARA	NO HAY			3-Mar-17	X	X	X	
48	PS TAILS-RESP SET Y17 - RGO - REPS 1Y 2									

Opportunities:

- Follow this scheme for other measurements based on phenology
- Create more automated spreadsheets/templates



Challenges/opportunities in wheat data processing/curation

Accelerate and improve quality in data processing and curation of several physiological traits.

- **Quality check and basic curation done mainly in Excel for general physiological traits → KDXplore? R? Interfaces?**
- **Access to data visualization**



Case: NDVI (GB) data fast track processing

5 satin rgo 06 12 16 - Notepad

File Edit Format View Help

```
Time(ms),Plot,Count,NDVI,VI_2
1487510, 1, 1, 0.258, 0.590
1487610, 1, 2, 0.238, 0.616
1487710, 1, 3, 0.221, 0.638
1487810, 1, 4, 0.222, 0.637
1487910, 1, 5, 0.218, 0.642
1488010,
1488110,
1488210,
1488310,
1488410,
1488510,
1488610,
1488710,
1488810,
1488910,
1489010,
1489110,
1489210,
1490710,
1490810,
1490910,
1491010,
1491110
```

Set parameters to discard outliers



*new 1 - Notepad++

File Edit Search View Encoding Language Settings Run TextFX Plugins Window ?

```
new 1
1 timemax<-100 # number of frames (and observations - so no interpolation needed)
2 setwd("C:/Users/Robert/Documents/animation/")
3 graphdata<-read.csv("filling_line_data.csv")
4 attach(graphdata)
5 gname<-paste("g",time,".tif", sep="") # holds the names of the picture files
6
7 # draw graphs
8 for (i in 1:timemax) {
9   tiff(gname[i],width=480)
10  plot(time[i],value[i],type="l",ylim=c(0,110),xlim=c(0,100),
11       ylab="",xlab="time")
12 }
13
14
15
16
17
18
19
```

										NDVI									
										NDVIvg1	NDVIvg2	NDVIvg3	NDVIvg4	NDVIvg5	NDVIg1	NDVIg2	NDVIc	NDVIvgAV	NDVIgAV
										11-Dec	17-Dec	26-Dec	4-Jan	5-Feb	23-Feb	14-Mar	Av.	Av.	Av.
1	1	1	1	459206	190	5077000	CIRNO C 2008	0.140	0.141	0.185	0.309	0.833	0.777	0.784	0.141	0.442	0.781		
2	2	1	1	534690	24	6922442	C80.1/3*QT4118/K	0.165	0.200	0.347	0.675	0.861	0.787	0.720	0.183	0.628	0.754		
3	3	1	1	6831	33	775	SIETE CERROS T66	0.158	0.199	0.415	0.750	0.851	0.773	0.784	0.179	0.672	0.779		
4	4	1	1	7624	7	2465	PAVON F 76	0.177	0.196	0.411	0.633	0.836	0.771	0.706	0.186	0.627	0.738		
5	5	1	1	7691	50	3895	SERI M 82	0.164	0.168	0.298	0.622	0.870	0.762	0.703	0.166	0.597	0.732		
6	6	1	1	7896	254	16122	BACANORA T 88	0.159	0.149	0.256	0.491	0.807	0.728	0.695	0.154	0.518	0.711		
7	7	1	1	8890	34	41948	ANTILLA	0.173	0.185	0.345	0.716	0.850	0.772	0.686	0.179	0.637	0.729		
8	8	1	1	8626	465	447647	BAVIACORA M 92	0.149	0.136	0.166	0.287	0.761	0.756	0.683	0.142	0.404	0.719		
9	9	1	1	160278	44	371007	SERI/RAYON	0.167	0.196	0.324	0.620	0.838	0.735	0.719	0.181	0.594	0.727		
10	10	1	1	448418	52	5397958	BRBT1*2/KIRITATI	0.172	0.192	0.440	0.649	0.826	0.689	0.706	0.182	0.638	0.697		
11	11	1	2	448436	114	5397748	PFAU/WEAVER*2//	0.158	0.162	0.207	0.403	0.822	0.749	0.670	0.160	0.477	0.710		
12	12	2	2	390523	0	2447427	KRICHAUFF	0.189	0.210	0.413	0.562	0.740	0.653	0.650	0.200	0.572	0.651		
13	13	1	2	448409	101	5398160	BECARD	0.155	0.217	0.427	0.633	0.843	0.728	0.657	0.186	0.634	0.692		
14	14	4	1	517040	33	6178783	SAUAL/WHEAR//S	0.167	0.178	0.398	0.641	0.827	0.747	0.719	0.173	0.622	0.733		
15	15	5	1	516641	60	6179253	WBLL1*2/4/BABAX	0.172	0.265	0.521	0.821	0.859	0.782	0.757	0.219	0.734	0.770		
16	16	1	2	521064	14	6179559	KFA/3/PFAU/WEAV	0.156	0.163	0.224	0.442	0.798	0.700	0.655	0.159	0.488	0.678		
17	17	1	2	520543	41	6177599	KINGBIRD	0.163	0.190	0.402	0.720	0.854	0.796	0.731	0.176	0.659	0.764		
18	18	1	2	516615	84	6176178	UP2338*2/4/SNI/T	0.170	0.161	0.248	0.566	0.840	0.734	0.737	0.166	0.552	0.735		
19	19	1	2	485004	542	6171893	CMH79A.955/4/AG	0.163	0.189	0.335	0.614	0.803	0.700	0.712	0.176	0.584	0.706		
20	20	1	2	534690	11	6489912	C80.1/3*QT4118/K	0.172	0.188	0.360	0.610	0.857	0.750	0.646	0.180	0.609	0.698		
21	21	1	3	550040	35	7129731	SOKOLL/PUB94.15	0.179	0.205	0.451	0.702	0.854	0.712	0.690	0.192	0.669	0.701		

R program

Legend HIBAPY15-16 RGO_raw HIBAPY15-16 RGO_XX Field Management

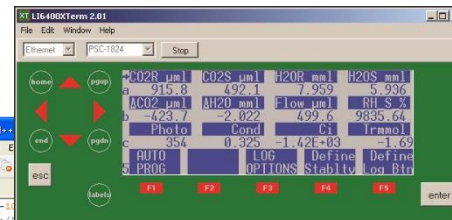
Case: IRGAS data processing



OPEN 6.1.4

Obs in	HHMMSS in	FTime in	EBal? in	Photo out	Cond out	Cl out	FCnt in	DCnt in	Fo in	Fm in	Fo' in	Fm' in
1	11:35:02	1489.5	0	17.1629804	0.14561462	171.856627	0	0	0	0	0	0
2	11:38:42	1709.5	0	-22.471994	0.14857981	668.430939	0	0	0	0	0	0
3	11:43:05	1973.5	0	24.4790586	0.18563014	135.042287	0	0	0	0	0	0
4	11:47:01	2210	0	-38.335218	0.31298855	656.29832	0	0	0	0	0	0
5	11:51:39	2488	0	-0.1329091	0.12179614	387.83257	0	0	0	0	0	0
6	11:55:14	2703.5	0	15.6575199	0.07591813	33.4704323	0	0	0	0	0	0
7	11:58:51	2921	0	-16.559563	0.11466133	647.188024	0	0	0	0	0	0
8	12:01:47	3097.5	0	15.5438864	0.04699411	-162.68464	0	0	0	0	0	0
9	12:06:07	3357.5	0	-26.134874	0.09427793	876.045994	0	0	0	0	0	0
10	12:06:57	3407.5	0	1.21021317	0.06010505	355.173074	0	0	0	0	0	0
11	12:06:58	3408.5	0	4.47943261	0.0572992	256.732398	0	0	0	0	0	0
12	12:06:59	3409.5	0	7.39563145	0.05539713	163.606188	0	0	0	0	0	0
13	12:07:02	3412.5	0	13.8416555	0.05497848	-35.059152	0	0	0	0	0	0

PLOT	ENTRY	REP	SUBREP	BLOCK	CrossName	Date_FL_pre	A_FL_Pre	gs_FL_Pre	Cl_FL_Pre	Trmmol_FL	VpdL_FL	P_Tail	FL_Pn
4	1	1	1	1	Jerez 37	17-jun-16	22.8	0.388	281.0	4.31	1.12	24.2	
4	1	1	2	1	Jerez 38	17-jun-16	24.6	0.352	261.5	4.20	1.20	24.2	
6	1	2	1	2	Jerez 39	17-jun-16	22.0	0.376	281.8	4.17	1.12	24.1	
6	1	2	2	2	Jerez 40	17-jun-16	21.9	0.315	263.9	3.59	1.14	24.0	
6	1	2	2	2	Jerez 41								
13	1	3	1	3	Jerez 42								
13	1	3	2	3	Jerez 43								
13	1	3	1	3	Jerez 44								
13	1	3	2	3	Jerez 45								
5	2	1	1	1	Blanqueta	22-jun-16	30.3	0.801	306.4	8.66	1.14	25.8	
5	2	1	2	1	Blanqueta	22-jun-16	28.5	0.714	304.5	8.04	1.18	25.5	
8	2	2	1	2	Blanqueta	22-jun-16	27.9	0.588	292.0	7.44	1.47	26.1	
8	2	2	2	2	Blanqueta	22-jun-16	29.8	0.585	283.2	7.44	1.44	26.0	
12	2	3	1	3	Blanqueta	22-jun-16	26.0	0.570	295.2	7.44	1.44	26.1	
12	2	3	2	3	Blanqueta	22-jun-16	16.4	0.230	259.9	6.93	1.88	26.0	
12	2	3	3	3	Blanqueta	22-jun-16	28.1	0.550	286.4	7.84	1.47	26.1	
2	3	1	1	1	Anton	17-jun-16	27.4	0.472	277.4	4.36	0.94	23.9	
2	3	1	2	1	Anton	17-jun-16	25.0	0.462	287.4	4.32	0.95	23.9	
9	3	2	1	2	Anton	17-jun-16	7.5	0.090	253.4	1.32	1.42	24.0	
9	3	2	2	2	Anton	17-jun-16	21.6	0.381	285.4	3.84	1.02	24.1	



```

File Edit Search View E
new 1 Notepad+
File Edit Search View E
1 timestart=1
2 setwd("C:/Users/robert/Desktop/IRGAS/lineR")
3 graphdata<-read.csv("filling_line_data.csv")
4 attach(graphdata)
5 gname<-paste("g",time,".tif", sep="") # holds the names of the picture files
6
7 # draw graphs
8 for (i in 1:timestart) {
9   tiff(gname[i],width=480)
10  plot(time[i],value[i],type="l",ylim=c(0,100),
11       ylab="",xlab="time")
12  polygon(c(1:i,1:i),c(upper[1:i],lower[1:i])
13         lines(time[1:i],value[1:i])
14         dev.off(dev.cur())
15 }
16
17 # call FFmpeg and make the video
18 shell("C:/ffmpeg/ffmpeg.exe -report -i g%d.tif -biv c08k filling_lineR.mpg,mustWork=FALSE)
19
R programming language length: 677 lines: 19
Ln: 19 Col: 1 Sel: 0
Dos/Windows ANSI
  
```

Codes and/or programming equipment

Image and screen prints: Rut Sanchez

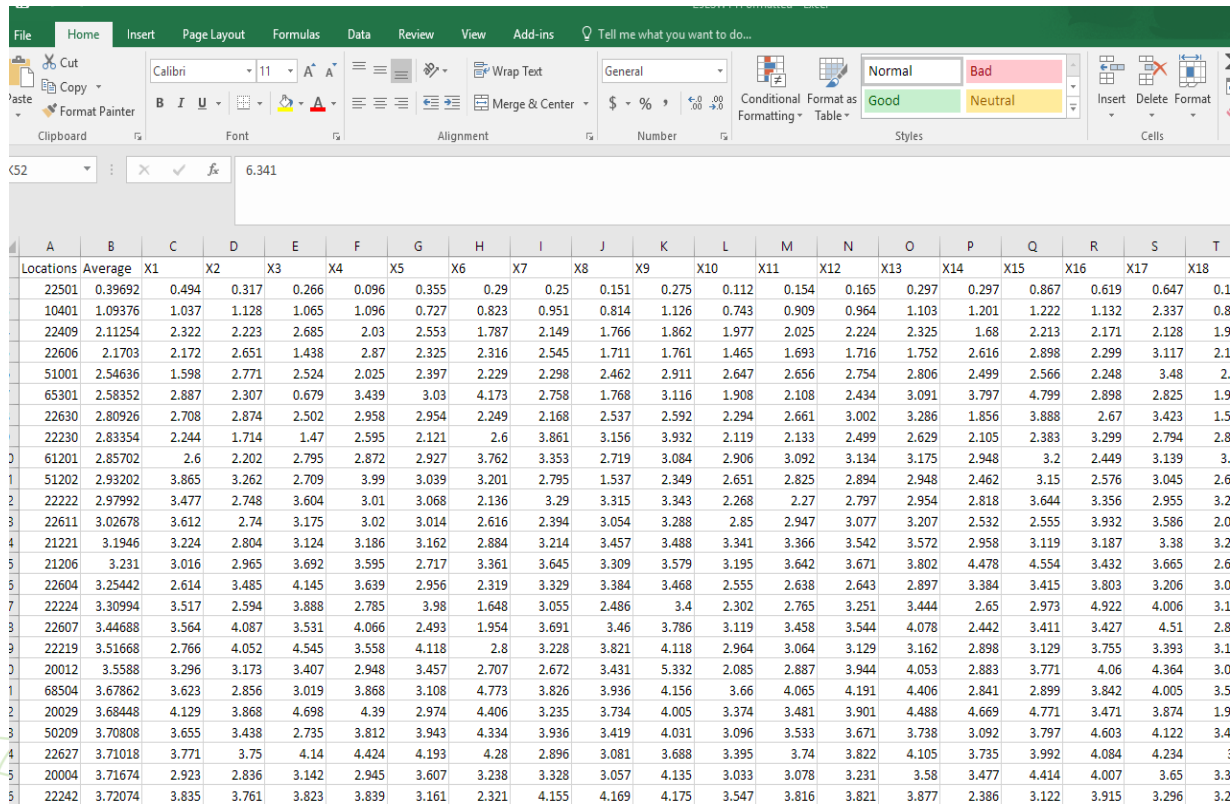
Case: historical wheat research data

There is a need to migrate historical data from older software (e.g. UQ123) to more accessible and compatible databases

In the process, we could arrange, process and analyze the data to make it accessible to collaborators and scientist for publications.

- ESWYT historical yield and TGW data, 10 years, ~60 locations, 50 genotypes.
- Data arrangement for complex statistical analyses.
- Manual → R

**Sivakumar Sukumaran
and Carolina Rivera**



The screenshot shows an Excel spreadsheet with the following structure:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
	Locations	Average	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18
1	22501	0.39692	0.494	0.317	0.266	0.096	0.355	0.29	0.25	0.151	0.275	0.112	0.154	0.165	0.297	0.297	0.867	0.619	0.647	0.1
2	10401	1.09376	1.037	1.128	1.065	1.096	0.727	0.823	0.951	0.814	1.126	0.743	0.909	0.964	1.103	1.201	1.222	1.132	2.337	0.8
3	22409	2.11254	2.322	2.223	2.685	2.03	2.553	1.787	2.149	1.766	1.862	1.977	2.025	2.224	2.325	1.68	2.213	2.171	2.128	1.9
4	22606	2.1703	2.172	2.651	1.438	2.87	2.325	2.316	2.545	1.711	1.761	1.465	1.693	1.716	1.752	2.616	2.898	2.299	3.117	2.1
5	51001	2.54636	1.598	2.771	2.524	2.025	2.397	2.229	2.298	2.462	2.911	2.647	2.656	2.754	2.806	2.499	2.566	2.248	3.48	2.
6	65301	2.58352	2.887	2.307	0.679	3.439	3.03	4.173	2.758	1.768	3.116	1.908	2.108	2.434	3.091	3.797	4.799	2.898	2.825	1.9
7	22630	2.80926	2.708	2.874	2.502	2.958	2.954	2.249	2.168	2.537	2.592	2.294	2.661	3.002	3.286	1.856	3.888	2.67	3.423	1.5
8	22230	2.83354	2.244	1.714	1.47	2.595	2.121	2.6	3.861	3.156	3.932	2.119	2.133	2.499	2.629	2.105	2.383	3.299	2.794	2.8
9	61201	2.85702	2.6	2.202	2.795	2.872	2.927	3.762	3.353	2.719	3.084	2.906	3.092	3.134	3.175	2.948	3.2	2.449	3.139	3.
10	51202	2.93202	3.865	3.262	2.709	3.99	3.039	3.201	2.795	1.537	2.349	2.651	2.825	2.894	2.948	2.462	3.15	2.576	3.045	2.6
11	22222	2.97992	3.477	2.748	3.604	3.01	3.068	2.136	3.29	3.315	3.343	2.268	2.27	2.797	2.954	2.818	3.644	3.356	2.955	3.2
12	22611	3.02678	3.612	2.74	3.175	3.02	3.014	2.616	2.394	3.054	3.288	2.85	2.947	3.077	3.207	2.532	2.555	3.932	3.586	2.0
13	21221	3.1946	3.224	2.804	3.124	3.186	3.162	2.884	3.214	3.457	3.488	3.341	3.366	3.542	3.572	2.958	3.119	3.187	3.38	3.2
14	21206	3.231	3.016	2.965	3.692	3.595	2.717	3.361	3.645	3.309	3.579	3.195	3.642	3.671	3.802	4.478	4.554	3.432	3.665	2.6
15	22604	3.25442	2.614	3.485	4.145	3.639	2.956	2.319	3.329	3.384	3.468	2.555	2.638	2.643	2.897	3.384	3.415	3.803	3.206	3.0
16	22224	3.30994	3.517	2.594	3.888	2.785	3.98	1.648	3.055	2.486	3.4	2.302	2.765	3.251	3.444	2.65	2.973	4.922	4.006	3.1
17	22607	3.44688	3.564	4.087	3.531	4.066	2.493	1.954	3.691	3.46	3.786	3.119	3.458	3.544	4.078	2.442	3.411	3.427	4.51	2.8
18	22219	3.51668	2.766	4.052	4.545	3.558	4.118	2.8	3.228	3.821	4.118	2.964	3.064	3.129	3.162	2.898	3.129	3.755	3.393	3.1
19	20012	3.5588	3.296	3.173	3.407	2.948	3.457	2.707	2.672	3.431	5.332	2.085	2.887	3.944	4.053	2.883	3.771	4.06	4.364	3.0
20	68504	3.67862	3.623	2.856	3.019	3.868	3.108	4.773	3.826	3.936	4.156	3.66	4.065	4.191	4.406	2.841	2.899	3.842	4.005	3.5
21	20029	3.68448	4.129	3.868	4.698	4.39	2.974	4.406	3.235	3.734	4.005	3.374	3.481	3.901	4.488	4.669	4.771	3.471	3.874	1.9
22	50209	3.70808	3.655	3.438	2.735	3.812	3.943	4.334	3.936	3.419	4.031	3.096	3.533	3.671	3.738	3.092	3.797	4.603	4.122	3.4
23	22627	3.71018	3.771	3.75	4.14	4.424	4.193	4.28	2.896	3.081	3.688	3.395	3.74	3.822	4.105	3.735	3.992	4.084	4.234	3.
24	20004	3.71674	2.923	2.836	3.142	2.945	3.607	3.238	3.328	3.057	4.135	3.033	3.078	3.231	3.58	3.477	4.414	4.007	3.65	3.3
25	22242	3.72074	3.835	3.761	3.823	3.839	3.161	2.321	4.155	4.169	4.175	3.547	3.816	3.821	3.877	2.386	3.122	3.915	3.296	3.

Challenges and opportunities in data compilation-after data collection

- Use of templates for data compilation
- Defined protocols for data compilation
- Short term storage, allowing to store different versions of data sets: **Dropbox, other clouds, KDXplore**
- Define quality control check points when compiling data
- Capacity building (to speed up the data sets compilation process and improve data quality)



Challenges and opportunities in data storing and sharing

- More extensive use of crop ontology
- Get as close as possible to donors needs
- Achieve efficient long-term storage databases
 - Centralize data within the institution in an accessible way
 - Find data easily inside and outside the institution
 - Store different types of data/information
 - Make data queries and visualize results
 - Compatibility with other software/databases
- Assure computational capacity



Examples of institutional platforms for sharing, storing and visualizing data

CIMMYT
International Maize and Wheat Improvement Center

Storing and publishing files

CIMMYT Dataverse Network

POWERED BY THE **Dataverse Network** PROJECT v. 3.0

Search Studies [Go] Advanced Search Tips

Released Dataverses

Name	View Info [+]	Owner
International Wheat Yield Partnership Research Data	View Info [+]	CIMMYT o
Cereal Systems Initiative for South Asia (CSISA) Research Data	View Info [+]	CIMMYT, I
CIMMYT Research Software	View Info [+]	CIMMYT
CIMMYT Seeds of Discovery	View Info [+]	CIMMYT
CIMMYT Research Data	View Info [+]	CIMMYT

Can we merge open access platforms for projects partners and CIMMYT staff?

Not secure | germinate.seedsfordiscovery.org/wheat/

MasAgro SAGARPA SECRETARÍA DE AGRICULTURA, GANADERÍA, DESARROLLO RURAL, PESCA Y ALIMENTACIÓN

CIMMYT International Maize and Wheat Improvement Center

Contact us: seed@masagro.org

INTRO DATA + ENVIRONMENT + GROUPS ABOUT + SEARCH LOGOUT

SeeD Wheat Germinate

Login Register

Enter your username and password below to access this site. If you have forgotten your login details please consult the help page.

Username:

Password:

Login

Data warehouse allowing querying and visualizing of several types of data)