WAGENINGEN UR
For quality of life

C&E data in the documentation system of CGN

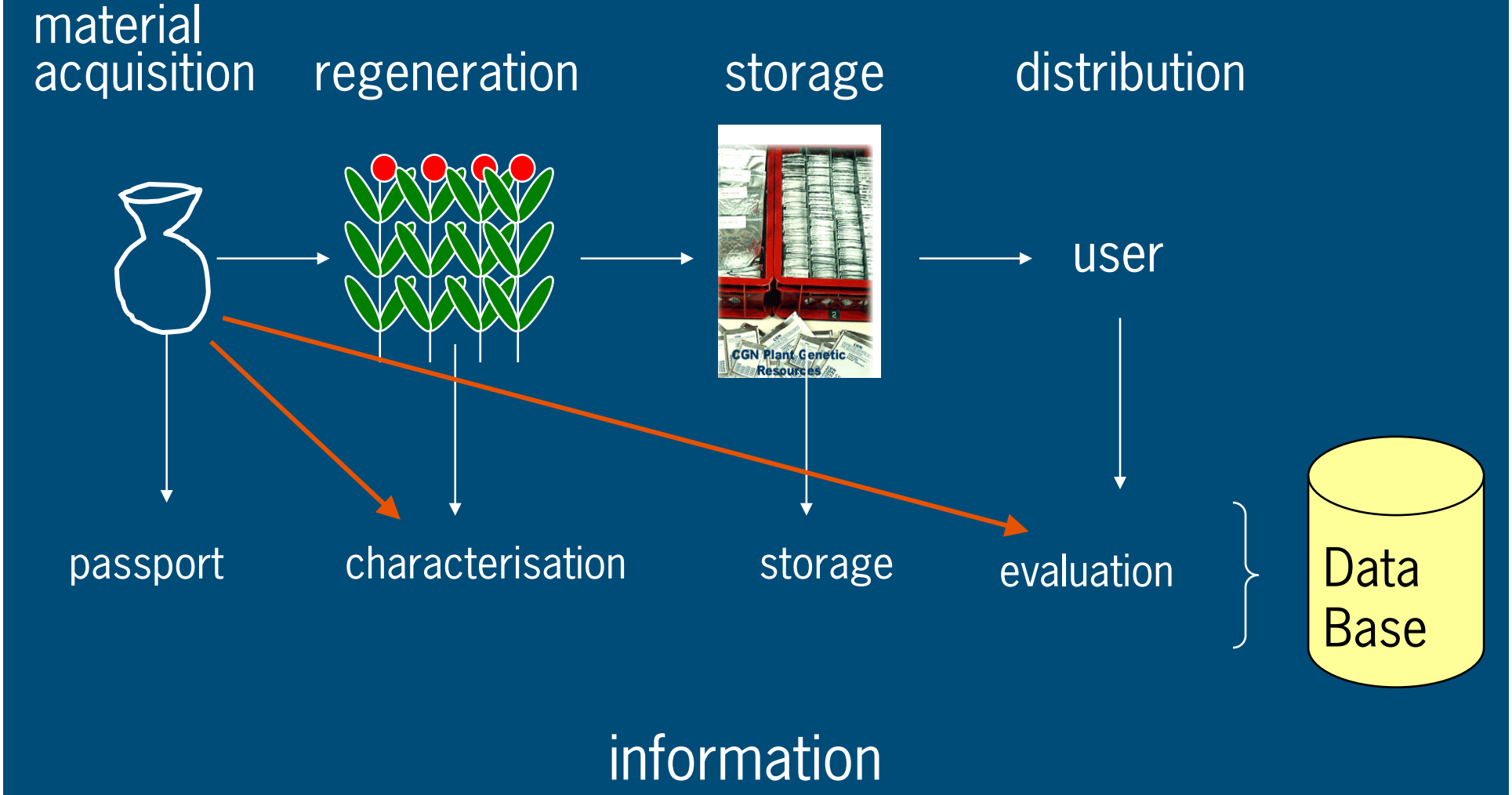
R. Hoekstra, Th.J.L. van Hintum, F. Menting

Centre for Genetic Resources, the Netherlands

outline

- acquisition of C&E data
  - descriptor lists, sources C&E data
- documentation of the scores
- public access to the C&E data
  - download facility
  - processing scores for on-line search

acquisition of descriptive (C&E) data



Centre for Genetic Resources, the Netherlands

WAGENINGEN UR
For quality of life

C&E data per experiment year

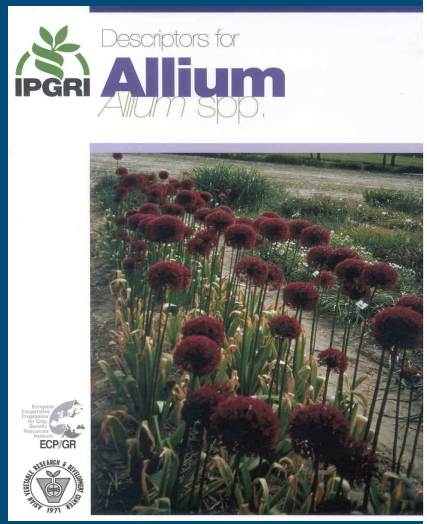
CGN started 1985

- 13% of C&E data received with germplasm acquisition
  - from previous collections: local breeding institutes or university

E_year	wheat	barley	flax	peas	lettuce	cruciferae	maize	faba beans	clover	spinach	allium	lupin	tomato	pepper	eggplant	cucumber	potato	melon	Total
64-'84		547			946	838					698		22671	17752	4342		6821		54615
85-'89	26479	37816		9010	22127	10179	4876	4775	5135	3261			10578	5678	3456		2037		145407
90-'94	1020	1335	222	3398	21482	5340	228		2713	9492	4302	109	8265	686	2587		2024		63203
95-'99	405	2241	10914	742	18531	1291			983	1067	2853		308	5091	2	6875	734		52037
00-'04	13737	3872	3267	2805	25067	6040				725	3189		1301	16199	9096	5586	31		90915
05-'08	1984	501		1044	9850	2396					310		4222	1384	170	344	113	1402	23720
Total	43625	46312	14403	16999	98003	26084	5104	4775	8831	14545	11352	109	47345	46790	19653	12805	11760	1402	429897

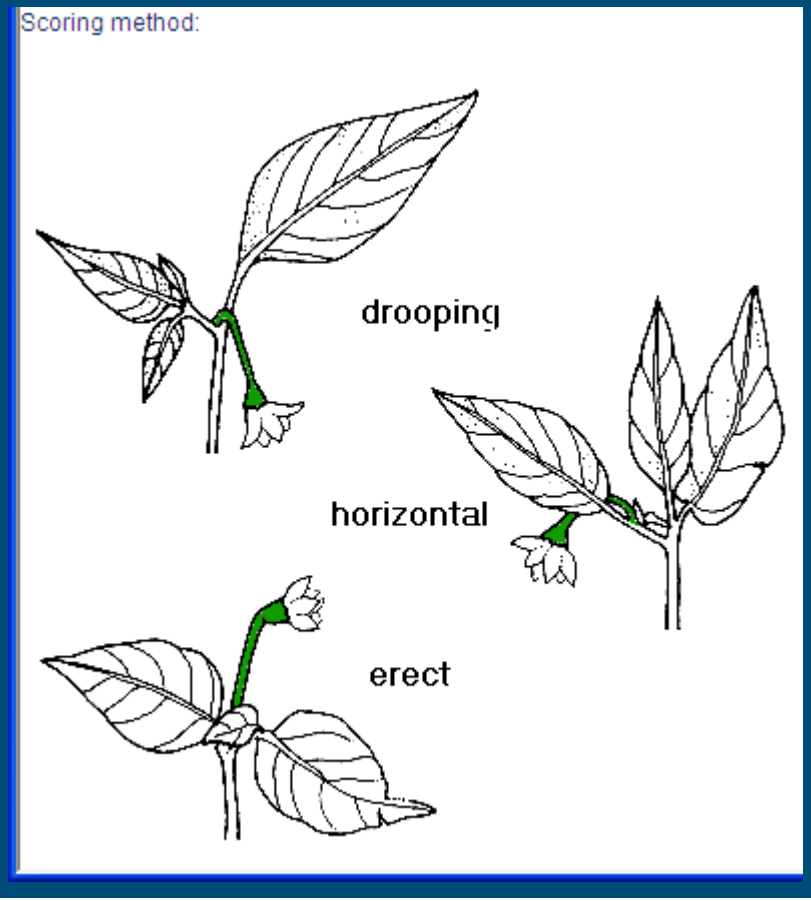
descriptor lists

help to determine what & how to screen PGR



- Existing descriptor lists
  - Standard lists of Bioversity
  - UPOV lists
- Own developed descriptor lists
  - descriptors partly based on Bioversity & UPOV lists & partly on own developed descriptors
    - in consultation with crop specialists and breeders

Descriptors: characterization data

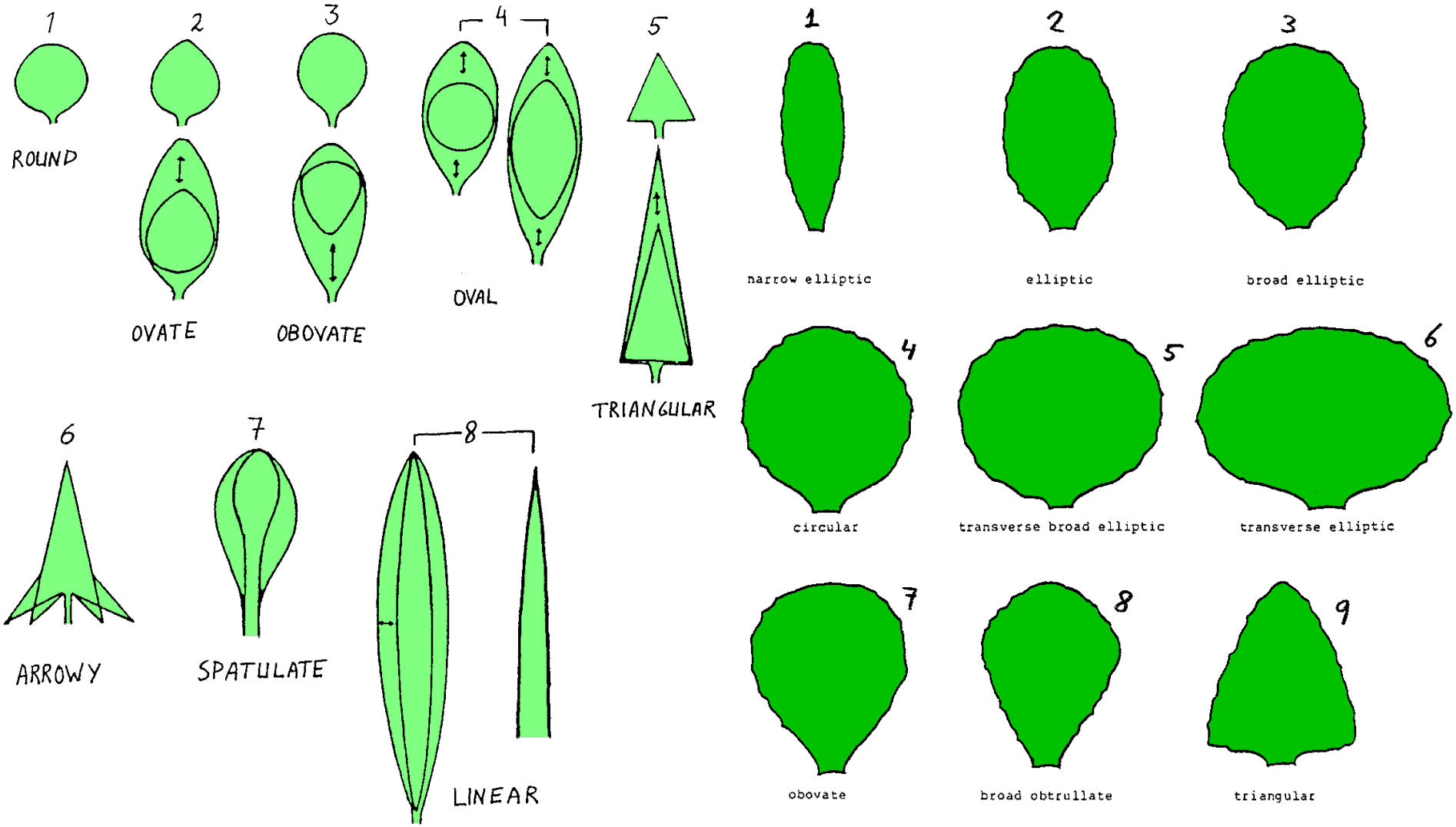


- description of simple more taxonomically related characters
- traits often highly heritable & expressed in all environments

descriptor for flower attitude of sweet pepper

Bruinsma Wonder=7 (1=very drooping, 3=drooping, 5=horizontal, 7=semi-erect, 9=erect)

descriptor for leaf shape lettuce: wild & cultivated



Centre for Genetic Resources, the Netherlands

WAGENINGEN UR
For quality of life

Descriptors: preliminary evaluation

- traits with agronomic value
  - e.g. lodging, tillering, field resistance for pests & diseases
- often observed during a regeneration
- sometimes special trials and knowledge required

screening resistance to *Xanthomonas campestris* (blackrot) in Crucifers

promote screening of PGR

impossible to conduct all type of descriptions by the genebanks themselves

- where possible, participate in relevant EU projects
- contacts with private breeders
  - CGN has 9 crop committee's with crop specialists of companies & the university. The breeding companies conduct screening programmes for important traits
    - and also participate in regeneration activities (input in kind)

2-3 years after distribution of PGR, screening results are requested from the users

- documentation of the scores

type of traits

- qualitative traits
  - mostly little influenced by environmental factors
  - e.g. flower color, leaf shape, chromosome no.
    - scores: 1=red, 2=white, 3=yellow
- quantitative traits
  - usually highly influenced by environmental factors
    - e.g. plant height
      - score: continuous scale (1-9 or absolute number)
- reaction to diseases or pests
  - pathotypes, different types of reaction
    - mostly on 1-9 scale
- fingerprinting data
  - not include in Genis yet

documentation of the scores

descriptors according to descriptor list

- descriptor states, way of scoring
  - absolute figures: length (cm), weight (g), percentages
  - scales: 1-5, 1-9, 1-5-9
  - symbols: R, M, S, I, H
- info about experiment & method required

examples of method (& trait) descriptions

- most common units: Stand (1-9) and State

	A	B	C	D	E	F	G
1	**CNR**	**ENR**	**TRT**	**TRAIT**	**METH**	**UNIT**	**M_DESCR**
2	1	18	22	growth height	62	cm	Average of 10 plants, at post anthesis to maturity
3	4	618	7	thousand grain weight	4	g	Old seeds, at least 100 seeds were weighed
4	1	18	80	flag leaf width	61	mm	Measure at widest point, average of 10 plants, at post anthesis to early maturation
5	4	314	3	flower insertion height	7	Num	Count nodes on main stem downwards, including 1st flowering node, not including scale nodes
6	1	4	18	spike density	8	Stand	(3=lax, 7=dense)
7	1	4	22	growth height	10	Stand	Measurement, including spike, excluding awns, related to standards
8	2	226	19	lodging susceptibility	20	Stand	-
9	3	569	69	harvest time	19	Stand	(3=early, 7=late)
10	4	2	5	pod length	18	Stand	Measured arount harvest, average of five pods, related to the population
11	37	339	386	fruit fleshiness	508	Stand	Sonatine=3
12	2	680	21	row number	9	State	(1=2-rowed, 9=4- or 6-rowed)
13	2	46	24	lemma color at maturity	12	State	(1=yellow/white, 5=red/purple, 9=blue/black, 0=other)
14	2	196	339	yellow dwarf virus resist	24	State	Percentage of leaf infected (0=0%, 1=0-1%, 2=1-5%, 3=5-10%, 4=10-20%, 5=20-30%, 6=30-40%, 7=40-60%, 8=60-80%, 9=>80%)R
15	4	8	8	testa color	1	State	Old seeds (1=colorless, 2=single colored, 3=marbled, 4=dotted, 5=marbled and dotted)

currently in use: 49 crops, 615 Traits, 907 methods, 764 Experiments

scoring heterogeneity

- score=A, all plants were scored A
- score=BC, little more than 50% score B, and little less than 50% score C
- score=BxC, about 70% of the plants score B, 30% of the plants score C
- score=BxxC, about 90% of the plants score B, 10% of the plants score C

also allowed complex scores like: AxBxxC

C & E data storage

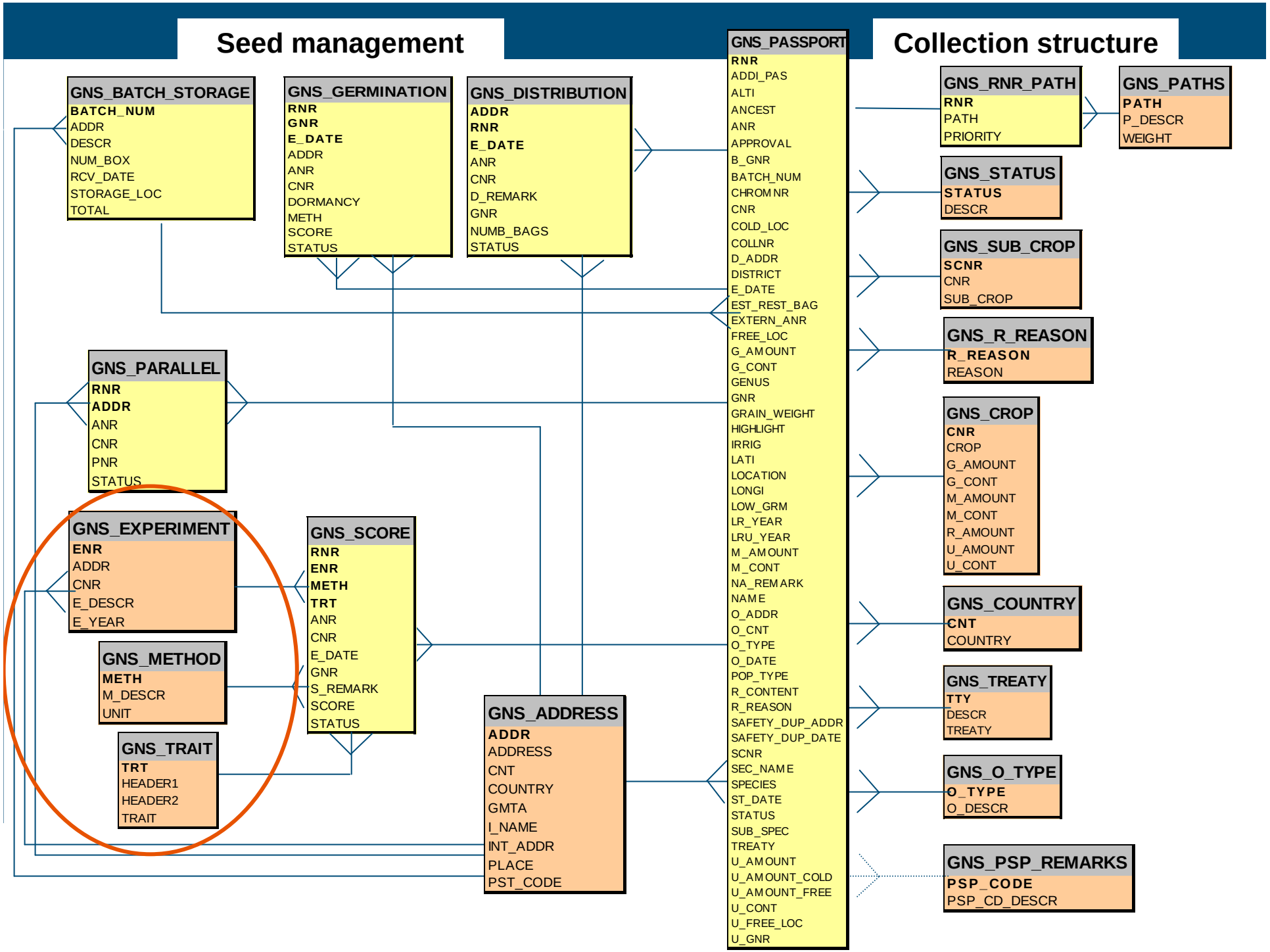
- per accession
  - Trait
  - Method
  - Experiment (under which conditions; by whom)
  - Score
- the results are stored score oriented

ACC	TRT	METH	SCORE
3251	plant height	lenghth (cm)	60
3251	plant height	1= short, 9 = long	5
3251	flower colour	A=white, B=red	A
3251	spikes	count	25
3251	fruit colour	A=white, B=red	B

C & E data storage coded

ACC	TRT	METH	EXP	SCORE
3251	15	12	1	60
3251	15	15	1	A
3251	4	8	1	B
3251	20	55	58	1
3251	7	55	77	1

- decoding tables required for Trait, Experiment & Method
- the numbers for new Traits, Experiment or Methods are provided by the documentation section


Seed management
Collection structure
GNS_BATCH_STORAGE
BATCH_NUM
ADDR
DESCR
NUM_BOX
RCV_DATE
STORAGE_LOC
TOTAL
GNS_GERMINATION
RNR
GNR
E_DATE
ADDR
ANR
CNR
DORMANCY
METH
SCORE
STATUS
GNS_DISTRIBUTION
ADDR
RNR
E_DATE
ANR
CNR
D_REMARK
GNR
NUMB_BAGS
STATUS
GNS_PASSPORT
RNR
ADDI_PAS
ALTI
ANCEST
ANR
APPROVAL
B_GNR
BATCH_NUM
CHROMNR
CNR
COLD_LOC
COLLNR
D_ADDR
DISTRICT
E_DATE
EST_REST_BAG
EXTERN_ANR
FREE_LOC
G_AMOUNT
G_CONT
GENUS
GNR
GRAIN_WEIGHT
HIGHLIGHT
IRRIG
LATI
LOCATION
LONGI
LOW_GRM
LR_YEAR
LRU_YEAR
M_AMOUNT
M_CONT
NA_REMARK
NAME
O_ADDR
O_CNT
O_TYPE
O_DATE
POP_TYPE
R_CONTENT
R_REASON
SAFETY_DUP_ADDR
SAFETY_DUP_DATE
SCNR
SEC_NAME
SPECIES
ST_DATE
STATUS
SUB_SPEC
TREATY
U_AMOUNT
U_AMOUNT_COLD
U_AMOUNT_FREE
U_CONT
U_FREE_LOC
U_GNR
GNS_RNR_PATH
RNR
PATH
PRIORITY
GNS_PATHS
PATH
P_DESCR
WEIGHT
GNS_STATUS
STATUS
DESCR
GNS_SUB_CROP
SCNR
CNR
SUB_CROP
GNS_R_REASON
R_REASON
REASON
GNS_CROP
CNR
CROP
G_AMOUNT
G_CONT
M_AMOUNT
M_CONT
R_AMOUNT
U_AMOUNT
U_CONT
GNS_COUNTRY
CNT
COUNTRY
GNS_TREATY
TTY
DESCR
TREATY
GNS_O_TYPE
O_TYPE
O_DESCR
GNS_PSP_REMARKS
PSP_CODE
PSP_CD_DESCR
GNS_PARALLEL
RNR
ADDR
ANR
CNR
PNR
STATUS
GNS_EXPERIMENT
ENR
ADDR
CNR
E_DESCR
E_YEAR
GNS_METHOD
METH
M_DESCR
UNIT
GNS_TRAIT
TRT
HEADER1
HEADER2
TRAIT
GNS_SCORE
RNR
ENR
METH
TRT
ANR
CNR
E_DATE
GNR
S_REMARK
SCORE
STATUS
GNS_ADDRESS
ADDR
ADDRESS
CNT
COUNTRY
GMTA
I_NAME
INT_ADDR
PLACE
PST_CODE

- public access to the C&E data
  - downloadable Excel files (zipped)

procedure export Oracle → Excel

C&E data downloadable since January 2001

- per crop/trait
  - export raw data into an Ascii file using SQL
    - incl. description of experiments & methods
  - import from Ascii file into Excel file using a VisBas tool
  - zip with ZipGenius

raw C&E data downloadable as Excel file

- from: www.cgn.wur.nl/uk
  - the user first selects a crop

CGN downloadable data sets

CGN maintains several databases. The most important, GENIS, contains information on the CGN collections. Since not all information can be searched on-line, it is possible to download all passport and all characterisation / evaluation data.

Data on the following crops can be downloaded

- agrostis
- allium
- apple
- barley
- clover
- cocksfoot
- cruciferae
- cucumber
- eggplant
- faba beans
- fescue
- flax
- lettuce
- lily
- lolium
- lupin
- maize
- meadow grass (Poa)
- melon
- oats
- peas
- pepper
- potato
- spinach
- timothy
- tomato
- wheat

WAGENINGEN UR
For quality of life

selected:

cruciferae

- then selects a trait

CGN downloadable data

Downloadable data for cruciferae

The complete passport data can be downloaded in Zipped Excel format.

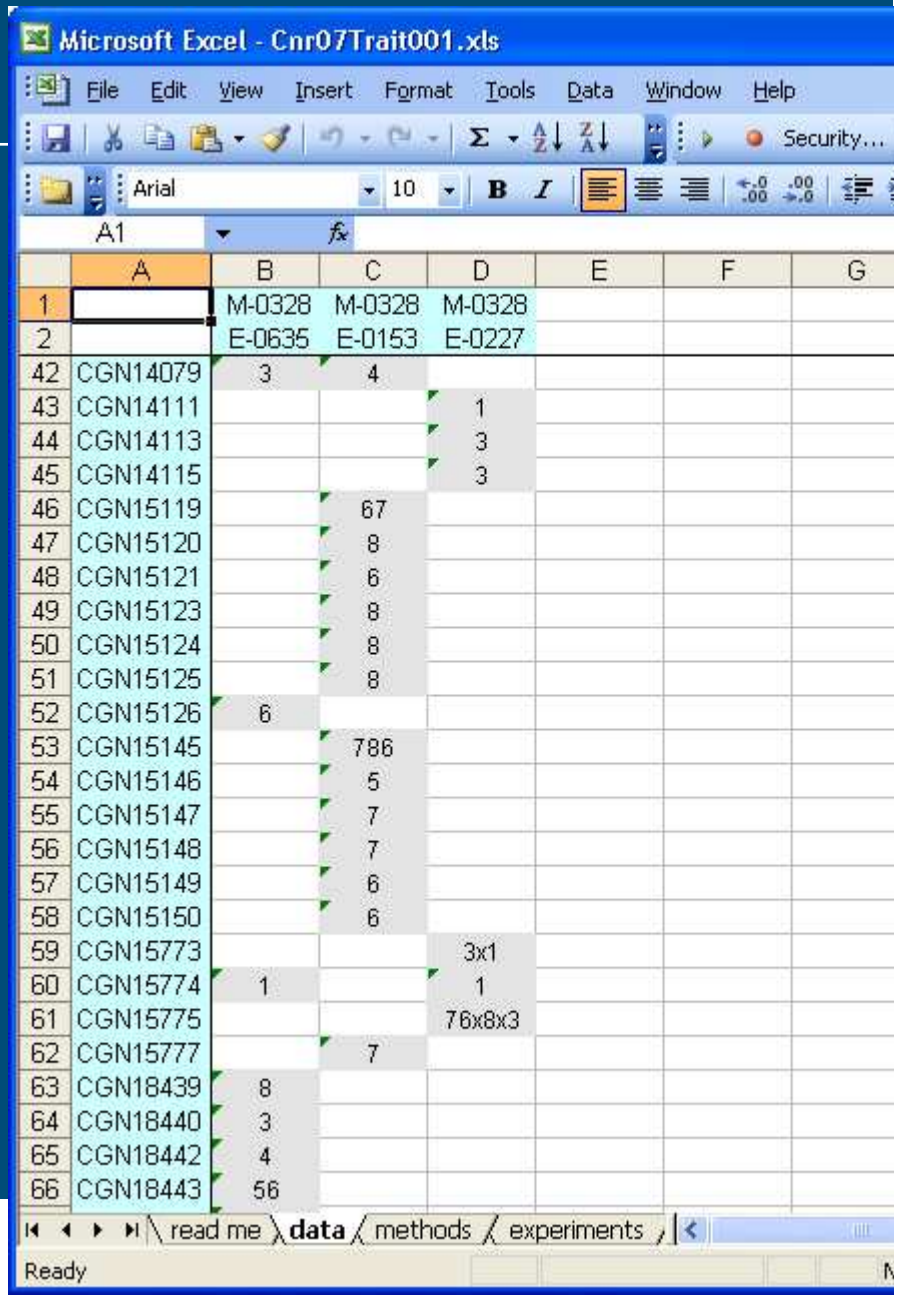
The evaluation data can be downloaded per trait in Zipped Excel format.

internode length	79 observations on 71 accessions
flower color	106 observations on 106 accessions
thousand grain weight	52 observations on 52 accessions
lodging susceptibility	18 observations on 18 accessions
growth height	768 observations on 667 accessions
winter susceptibility	107 observations on 107 accessions
flowering time begin	3 observations on 3 accessions
ripening time	24 observations on 24 accessions
harvest time	777 observations on 623 accessions
oil content	170 observations on 165 accessions
growth habit	3 observations on 3 accessions
flowering time	19 observations on 19 accessions
leaf number	311 observations on 238 accessions
leaf width	12 observations on 12 accessions
leaf length	131 observations on 131 accessions
stem length	97 observations on 89 accessions
leaf attitude	652 observations on 570 accessions
seed protein content	170 observations on 165 accessions
leaf shape	729 observations on 638 accessions
Heterodera schachtii	113 observations on 57 accessions
leaf color	1072 observations on 803 accessions
leaf hairiness	106 observations on 106 accessions

selected:

internode length

- 3 experiments
- only 1 method (fortunately)



	A	B	C	D	E	F	G
1		M-0328	M-0328	M-0328			
2		E-0635	E-0153	E-0227			
42	CGN14079	3	4				
43	CGN14111			1			
44	CGN14113			3			
45	CGN14115			3			
46	CGN15119		67				
47	CGN15120		8				
48	CGN15121		6				
49	CGN15123		8				
50	CGN15124		8				
51	CGN15125		8				
52	CGN15126	6					
53	CGN15145		786				
54	CGN15146		5				
55	CGN15147		7				
56	CGN15148		7				
57	CGN15149		6				
58	CGN15150		6				
59	CGN15773			3x1			
60	CGN15774	1		1			
61	CGN15775			76x8x3			
62	CGN15777		7				
63	CGN18439	8					
64	CGN18440	3					
65	CGN18442	4					
66	CGN18443	56					

A1 Experiment

	A	B	C
1	Experiment	Year	Description
2	E-0153	1989	Borecole, marrow stem kale, unspecified kale, wild species, thousand headed kale and perennial kale. Field trail. Sowing date 21/6/89, planting date 18/7/89, planting distance 65*50 cm. 44 Plants per object. IVT grounds Wageningen.
3	E-0227	1991	Field trail several cabbages and kales, rnr 900437 & anr 7029: sowing date 7/3/91, planting date 25/4, distance 60x60cm; rest of objects: sowing date 20/6, planting date 23/7, distance 50x75cm. Location de Goor, Wageningen.
4	E-0635	2002	Characterisation borecole (sowing date: 20 June, planting date: 22 July, 65 x 50 cm) and kale (sowing date: 5 March, planting date: 25 April, 65 x 65 cm), Nergena, Wageningen.
5			

data methods **experiments**

A1 Method

	A	B	C
1	Method	Unit	Description
2	M-0328	Stand	(3=short, 5=medium, 7=long, 9=very long)
3			

data **methods** experiments

methods per crop/trait combination

- high number of methods per crop/trait combination always on different pathotypes for a diseases

A	B	C
no. distinct methods	**trt/cnr combi-nations**	
1	552	
2	141	
3	19	
4	7	
5	5	
6	2	
7	3	
8	2	disease resistance, different pathotypes
18	1	potato: G.pallida resistance (different pathotypes)
20	1	barley: powdery mildew infection rates and presence of specific R genes
63	1	lettuce: Bremia resistance (different pathotypes)

	A	B
1	**METH**	**M_DESCR**
2	23	Infection type (1=O, 3=I, 5=II, 7=III, 9=IV)
3	24	Percentage of leaf infected (0=0%, 1=0-1%, 2=1-5%, 3=5-10%, 4=10-20%, 5=20-30%, 6=30-40%, 7=40-60%, 8=60-80%, 9=>80%)R
4	550	Determined at natural infection (0=no symptoms, .., 9=serious symptoms)
5	614	+ = presence of MI-(41/145)
6	615	+ = presence of MI-(CP)
7	616	+ = presence of MI-(Wo)
8	617	+ = presence of MI-a14
9	618	+ = presence of MI-a6
10	619	+ = presence of MI-a7
11	620	+ = presence of MI-a8
12	621	+ = presence of MI-a12
13	622	+ = presence of MI-a9
14	623	+ = presence of MI-g
15	624	+ = presence of MI-h
16	625	+ = presence of MI-k
17	626	+ = presence of MI-v
18	627	+ = presence of MI-(AB)
19	628	+ = presence of MI-a13
20	629	+ = presence of MI-a3
21	630	+ = presence of MI-o

Ce

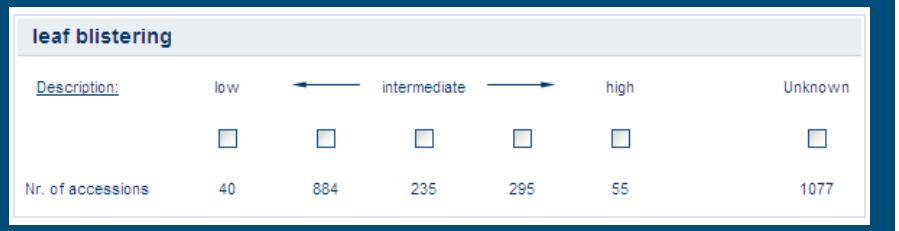
WAGENINGEN UR
For quality of life

- public access to the C&E data
  - on-line search

the road to 1 score per trait/acc

the data sets need to be:

- analysed
- transformed (making data sets comparable)
- validated (abnormalities reported to curator)
- combined into 1 single score (per trait/accession)
- re-scaled into 1 - 5 scale (max. 5 check-boxes)



leaf blistering						
Description:	low	←	intermediate	→	high	Unknown
	☐	☐	☐	☐	☐	☐
Nr. of accessions	40	884	235	295	55	1077

Obvious problems

for the aggregation of results from several data sets into 1 score per trait / accession

- experiment mostly not replicated
- no standards included
- different environments
- different methods
- unknown pathotypes
- typo's

www.cgn.wur.nl/uk

CGN Crop-specific search form

Other searchpages: All crops Go! **Manage shopping cart**

Crop:	**cucumber**
Number of accessions:	**922**

Searches can be made based on passport data and characterization / evaluation data or both. Only a selection of traits is on-line searchable, however all data are downloadable.
Furthermore it is possible to create core selections to limit the number of selected accessions.

Search specific passport fields

Search for text in the genus, species, subtaxon, accession name, collection site, ancestor and remarks field.
This field accepts only one keyword, which can consist of multiple terms (i.e. 'long red').

Search specific passport fields

Genus:	All
Species:	All
Sub-taxa:	All
Sub-crop type:	All
(Part of) the accession name:	
(Part of) the accession number:	
Country of origin:	All
Sample status:	All

Pictures

☐ Check if you only want to see accessions with pictures.

Core selection

Enter the maximum number of accessions to be selected.

Sorting

◉ Scientific name ○ CGN Number ○ Country of origin

Submit the form, or reset the form to its default settings. Search Reset form

Characterization/evaluation data - Available traits

Check the boxes below to select only the accessions with the chosen characteristics. The description of each trait links to some background information including a link to the downloadable raw characterisation/evaluation data for that trait.

Plant characteristics
- growth habit
- stem length
- leaf color
- leaf size

Flowering characteristics
- sex expression (portion of female flowers)
- parthenocarpy
- hermaphroditism

Fruit characteristics
- fruit type
- fruit length (at market stage)
- length/diameter ratio
- fruit skin ground color (at market stage)
- final fruit skin ground color (at seedripe stage)
- fruit skin mottling
- fruit skin striping
- fruit ribbing (at market stage)
- fruit shape stem end
- fruit vestiture type
- fruit vestiture color
- cotyledon bitterness
- fruit bitterness stem end

Plant characteristics

growth habit

Description:	determinate	indeterminate	mixture	Unknown
	☐	☐	☐	☐
Nr. of accessions	48	435	2	437

Currently on-line searchable crops (C&E data)

- potato 1 – 2001 (using Oracle - SQL)
- lettuce 6 – 2001 (using Excel), update: 6 - 2008

  software development (VBA) started 08 - 2002
- cucumber 10 - 2003
- sweet pepper 12 - 2003
- tomato 3 - 2004
- spinach 8 - 2004
- wheat 7 – 2006
- barley 12 – 2008

next: *Cruciferae*, ...

Example: *Globodera pallida*

3 pathotypes presented as 2 traits

- Pa1 resistance determined by 1 dominant gene
- Pa2 and Pa3 resistance are correlated and polygenic
  - combined into 1 trait

PA 1 (*Globodera pallida*)

Description:	susceptible		medium		resistant	Unknown
	☐	☐	☐	☐	☐	☐
Nr. of accessions	0	5	3	25	0	1346

PA 2,3 (*Globodera pallida*)

Description:	susceptible		medium		resistant	Unknown
	☐	☐	☐	☐	☐	☐
Nr. of accessions	8	218	98	95	2	958

Centre for Genetic Resources, the Netherlands

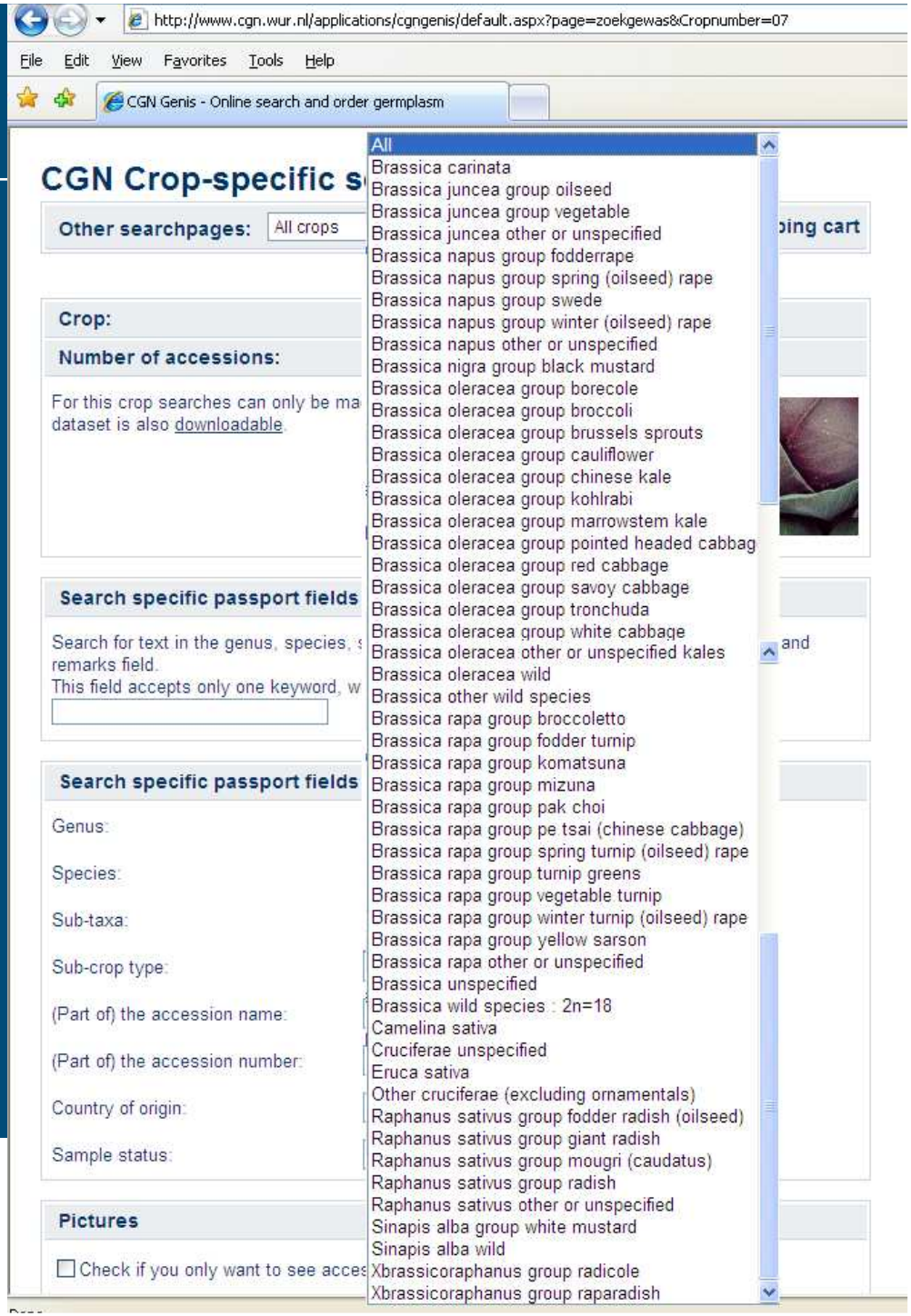
lettuce – reaction to *Bremia lactucae* (54 pathotypes)

- The classification for on-line search is depending on the no. of races tested for and the % to which the acc shows resistance
  - for on-line search split into 3 sub-traits: >=20 ; 10-19 ; 2-9 races

Classes	>= 20 races tested. % resistance	11 - 19 races tested. % resistance	5 - 9 races tested. % resistance	2 - 4 races tested. % resistance
high potential	>= 85 %	>= 85 %	>= 90 %	-
medium-high	>= 65 %	>= 65 %	>= 70 %	> 90 %
medium	>= 45 %	>= 45 %	>= 50 %	>= 70 %
low-medium	>= 30 %	>= 30 %	>= 30 %	> 50 %
low potential	< 30 %	< 30 %	< 30 %	<= 50 %

the next challenge

- Cruciferae, with all its sub-crops



WAGENINGEN UR
For quality of life

- Thanks for your attention