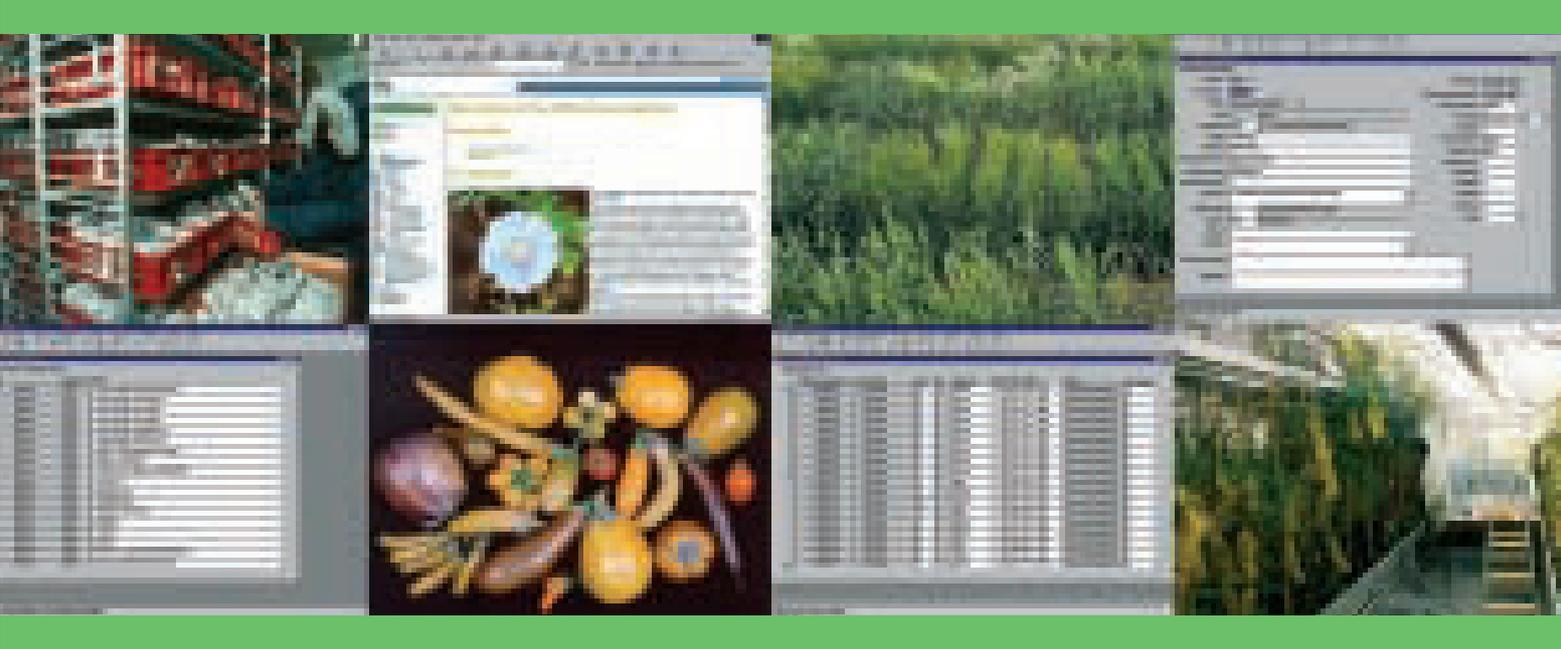


# Genis Data Dictionary

F.B.J. Menting & Th.J.L. van Hintum



Note 254





# Genis Data Dictionary

F.B.J. Menting & Th.J.L. van Hintum

**Centre for Genetic Resources, The Netherlands (CGN)**

P.O. Box 16

6700 AA Wageningen

The Netherlands

Email: [CGN@wur.nl](mailto:CGN@wur.nl)

# Table of contents

	page
1. Introduction	1
2. Structure	3
2.1 Information	3
2.1.1 User information	3
2.1.2 Decode information	4
2.1.3 Process information	4
2.2 Users	5
2.3 Processes	6
2.3.1 The status of the material	6
2.3.2 Requests	7
2.3.3 Data entry	7
2.3.4 Labels	7
3. Constituting elements	9
3.1 Tables and views with their fields	9
3.2 Menu and modules	19
4. Detailed description of some tables	23
4.1 The GNS_PASSPORT table	23
4.2 The GNS_PARALLEL table	30
4.3 The GNS_SCORE table	30
4.4 The GNS_PATHS and GNS_RNR_PATH tables	31
5. References	33
Appendix I. Codes used in CHI	1 p.
Appendix II. Symbols and abbreviations	1 p.
Appendix III. Privileges on tables/views for Oracle User Names	1 p.
Appendix IV. System for scoring heterogeneous populations	2 pp.
Appendix V. Genis on the Internet	1 p.
Appendix VI. Major changes since previous version	1 p.



# 1. Introduction

The information system of the Center for Genetic Resources, the Netherlands is called Genis (GENetic resources Information management System). It is based on the relational database management system Oracle. The latest version of Genis, version 2.1, is based on Oracle version 7.1 and was build using Oracle Designer 2000 and Oracle Developer 2000.

Genis version 2.1 is a client server application. This means that the software of the application is located at several locations. The database, the database software and the screen software are located at the sever side; the software to communicate with the screens is located at the client side. At the client side the hardware consists of PC's using Windows NT as operating system, and at the serverside Genis is run on a Windows 2000 server.

A database management system, such as Oracle, is a package of tools that can be used to manage large data sets. The owner of the database can define privileges for different kinds of users. Per (part of a) table can be defined what a user is allowed do: query, add, update or delete information.

All the information is stored in two-dimensional tables. The tables are linked using key-fields. This way it is possible to store different types of data in different tables and bring it together when necessary. For example there are separate tables for passport information, addresses of institutes, and germination results.

The basic structure of Genis was created in 1985 and became operational in 1986. After five years of experience with this version the system was revised in 1991 (Hintum & Hazekamp, 1992). To anticipate on the developments in computer technology and genebank documentation, a second revision took place in 1999. The last revision, that released Genis version 2.1, took place in 2002 and was an adaptation to changed genebank procedures.

This data dictionary will give an overview of all aspects of the information system important for understanding how Genis works. First the logical structure will be described: the different types of information in relation to the tables, the users with their privileges, and the processes. Next, a complete list of all elements in Genis will be given including the tables with their fields, the menu's, the screens and finally the procedures implemented in the system. Since the use of a table is not defined by formal format rules only, the tables with descriptive information on the accessions are described in some detail.

The information in Genis is largely available via the Internet. The system allowing this access is completely outside the Genis environment, the only connection is a regular upload of Genis data into the Internet application. Therefore the Internet application is only briefly described in an appendix.



## 2. Structure

In Genis three types of information are distinguished (see 2.1):

- User information, describing properties of the accessions in the genebank: the history, identity, genetic characteristics etc. of the accessions, and seed management information about the physical seed (location, quantity, distribution, germinability, etc).
- Decode information, giving an explanation of all codes used in the system.
- Process information, temporary stored information necessary for processes like accession of samples, or generation of labels, etc.

Four types of users of the information system are distinguished (see 2.2):

- End-users of the data, including plant breeders and scientists, but also the curators.
- Data typists, usually the assistants of the curators and database administrator.
- The seed storage manager.
- The database administrator.

All users have different privileges, depending on how they are supposed to use the system; the end-users can query most data, the documentation assistant and the seed manager can query all data, and add or change some specific types of data, the database administrator can do anything.

Basic processes supported by Genis are (see 2.3):

- Change of status of samples: the way the material is accessed, not accessed or archived in the CGN collections.
- Distribution of samples to users.
- Entering characterisation / evaluation data into the system.
- Generating labels for the seed bags.

### 2.1 Information

The three types of information distinguished in Genis, i.e., user-information, decode information, and process information, will be discussed in some detail below. References will be made to the tables in Genis (in capitals with names such as GNS\_PASSPORT) whose formal definitions are given in Chapter 3.

#### 2.1.1 User information

The ‘user information’ in Genis describes properties of the accessions in the genebank, and is considered to consist of two types:

- Descriptive information: history, identity, genetic characteristics etc. of the material, etc.
- Seed management information about the physical seed: location, quantity, distribution, germinability, etc.

##### 2.1.1.1 Descriptive information

Descriptive information describes the history, identity, genetic characteristics etc. of the accessions. This information is partitioned in:

- Passport information stored in the GNS\_PASSPORT table, describing identity, origin, history and taxonomy of the accession.

- Parallel numbers, in the GNS\_PARALLEL table, giving the accession numbers of the CGN material in other genebanks.
- Characterization and evaluation data, in the GNS\_SCORE table, describing the phenotypes of the accessions.

#### 2.1.1.2 Seed management information

The amount and type of information needed for seed management of a genebank depends on the procedures of that genebank, and is for that reason very genebank specific. Seed management information describes the physical seeds: their location in the seed storage, the quantity of the available seeds, the distribution to users and the germinability.

In Genis the administration of the seed storage is stored in the GNS\_PASSPORT table, the germinability data, in the GNS\_GERMINATION table, and the information about the distribution, in the GNS\_DISTRIBUTION table.

Safety duplication of CGN accessions sent to other genebanks is documented per accession in GNS\_PASSPORT. Safety duplication information received from other genebanks and information about batches of archived CGN-material (see 2.3.1) is stored in GNS\_BATCH\_STORAGE. Only information on the number and location of received safety duplicates are stored, passport information about these samples is not available in Genis.

### 2.1.2 Decode information

To avoid unnecessary duplication of information, and to safeguard the integrity of the database, many items in Genis are coded. These codes are decoded in decode tables storing the code in the first column and decode information in the following column(s).

Addresses of institutes, persons and expeditions are stored in the GNS\_ADDRESS table. The names of the crops can be found, with some additional storage information, in the GNS\_CROP table, the names of the sub crops in GNS\_SUB\_CROP. Traits scored during evaluation or characterization are stored in GNS\_TRAIT, while descriptions of the methods that have been used are stored in GNS\_METHOD. Information about these screenings can be found in GNS\_EXPERIMENT. The country codes are explained in the decode-table GNS\_COUNTRY. Origin types are described in GNS\_O\_TYPE. Codes used in the GNS\_PASSPORT field ADDI\_PAS are embedded in text, not in a separate column; they are decoded in GNS\_PSP\_REMARKS. The reasons for deciding not to access populations in the CGN collections are stored in GNS\_R\_REASON. The different statuses of the material are decoded in table GNS\_STATUS. Codes for information on access and benefit sharing of the seed samples are stored in GNS\_TREATY.

### 2.1.3 Process information

Process information is the information needed to perform processes such as generation of labels for seed bags (see 2.3) and processing distributed seed samples. This type of information is usually temporal, as soon as the process is finished the information is removed from the system.

There are several process tables in Genis:

GNS\_PROC\_ACCESS:

to assign accession numbers to material that will form part of the CGN collection, i.e., whose status will be changed from 'Received' to 'Accessed'.

**GNS\_PROC\_ARCHIVE:**

information on seed samples that are to be archived, i.e., whose status will be changed from 'Received' or 'Accessed' to 'Archived'.

**GNS\_PROC\_DISTRIBUTION:**

to store the requests for distribution of seeds.

**GNS\_PROC\_REJECT:**

information on seed samples that are not going to be accessed, i.e., whose status will be changed from 'Received' to 'Not Accessed'.

**GNS\_PROC\_OBJ\_SEQ and GNS\_PROC\_COLS:**

to create a format for efficiently entering characterization / evaluation data into GNS\_PROC\_SCORE per experiment.

**GNS\_PROC\_SCORE:**

to be able to check data before input in the GNS\_SCORE table.

**GNS\_PROC\_LABELS:**

to store information for the generation of labels.

## 2.2 Users

There are four roles defined in Genis.

**End-users of the information system (CGNUSR):**

plant breeders or researchers searching for suitable material for inclusion in their programs. Also genebank curators who have to base many of their decisions on information in Genis, like what to regenerate, what to include or archive in the collections, where to plan a collection mission, etc. The members of this group are able to query all data in the system, and generate some output-listings in a user-friendly way. They are not able to change or delete anything, not even by mistake. They are allowed to place orders for material.

**Data typists (CGNADM):**

are able to query all data and to add and correct all accession and certain process and decode information. The data typists are allowed to make basic output lists and files.

**The seed manager (CGNSMR):**

is able to query all data, to add and correct seed management data, and to place orders for labels. The seed manager is also able to generate relevant output.

**The database administrator (CGNDBA):**

owns the database and can do anything with it: adding, changing and deleting tables, granting or revoking privileges to users, etc.

For a detailed overview of all privileges of the different users per table see Appendix III.

## 2.3 Processes

As mentioned before, there are a few processes supported by Genis. An important process is the change of status of the material, from 'Received' to 'Accessed', 'Not Accessed' or 'Archived'. Other processes are the way the distribution of the material is administrated and the way characterization / evaluation data are entered into the system.

### 2.3.1 The status of the material

The status of a seed sample is given in the STATUS field of the GNS\_PASSPORT table. This field is duplicated in several user tables to improve performance and facilitate querying.

When seed samples are received they get a six digits number: the receipt number (RNR). The first two digits represent the year; the last four are a sequential number. All data that goes with this seed sample are now connected to this receipt number, which will stay the key to the sample from now on. The status of a received seed sample is 'R' ('Received').

Some tests can be performed on these received seed samples, to determine if they should be accessed: germination tests, resistance for diseases etc. But also the passport data are studied to determine if the sample would improve the composition of the collection. When it is decided to access a received sample the seed manager assigns an accession number (ANR) to it. The status changes from 'R' ('Received') to 'A' ('Accessed').

When it is decided not to access a received sample the sample gets the 'N' status ('Not Accessed'). Codes for the reasons for not accessing are given in the R\_REASON column of the GNS\_PASSPORT table. These codes are decoded in the GNS\_R\_REASON table. An additional remark about not accessing can be made in the field NA\_REMARK in GNS\_PASSPORT. The date of the change of status is stored in the GNS\_PASSPORT table in the ST\_DATE column.

It is also possible in exceptional cases that the status of an accession changes from 'A' to 'N'. For example if there appeared to be no viable seed. The result is a not accessed seed sample with an accession number and a receipt number.

The fourth status is the 'Archive'-status ('C'). Material with a 'C' status is stored under optimal conditions with zero input (no germination tests, distribution or multiplication). The 'C' status makes it easier to decide for example to remove probably redundant accessions from the collection or not to access candidate accessions. Archived material can be stored in batches with no information on individual samples (e.g. new, not yet interesting, collections), or as individual samples (e.g. removed accessions). Information on batches of archived material is stored in table GNS\_BATCH\_STORAGE. Information on individual samples with a 'C' status is stored in GNS\_PASSPORT. Archived material can only be activated on initiative of the curators.

The 'E' ('External') status allows the documentation of material that is under CGN's responsibility but is not physically stored in the CGN genebank. An example is the apple collection that is maintained as trees (field genebank) outside CGN. These accessions will always get an external accession number (EXTERN\_ANR) instead of an accession number (ANR).

### **2.3.2 Requests**

Users can request material with an 'A' status. When the accession numbers are entered in the system, the system checks whether there is still enough seed for distribution. If there is enough seed the requested sample is written off in the relevant table and registered as requested. The seed storage manager can ask once in a while for a list of requests, with the storage location of the requested seeds. The seed is then taken from the storage rooms and sent to the person requesting it. A list with passport data is generated automatically for each request.

### **2.3.3 Data entry**

There are several ways of entering data in Genis: directly in the Genis screens, or with tools such as SQL\*plus, which is a line oriented SQL interface, or SQL\*XL, an SQL extension of the commonly used Excel spreadsheet. It will depend on ones privileges and experience, and the type and amount of data which option is best.

Germination results, evaluation data, and passport data are loaded in Genis using SQL\*XL and Excel templates. These processes follow strict procedures to avoid mistakes. Original data are archived.

For entering evaluation data via the Genis screens, a process has been created in which it is possible to enter the objects included in an experiment and the trait / method combinations in that experiment. On the basis of these data a set of default records is created in a temporal table (GNS\_PROC\_SCORE) which scores only have to be updated, checked and transferred to the 'real' GNS\_SCORE table.

### **2.3.4 Labels**

A last process to be mentioned is the generation of labels using the process table GNS\_PROC\_LABELS: if the seed manager needs labels, or when new accessions have been made, labels can be ordered. A file with information that can be printed on labels is generated automatically. In this file information of several tables is combined.



## 3. Constituting elements

Genis consists of many programmes, structures, data elements, etc. In this chapter the most important of these elements will be listed and described briefly:

- Tables and views with their fields (3.1),
- Menu's and modules (3.2)

### 3.1 Tables and views with their fields

All tables and views of Genis are listed below with their fields. First the name of the table or view is given (in bold) followed by a short description. Views can be recognised by the ‘\_VW’ as suffix in the view name. The fields in the table are listed, with the type and length of the fields, an indication of the key fields, the table where certain fields are decoded, the ‘not null fields’ and a short description.

*Legend:* Type (= field type): V = Varchar2, N = Number, D = Date; Len: Length of the field; Key: P = Primary key, R = Referential integrity key, decoded in another table; Null: Y = field can be empty, N = field must contain a value

<b>GNS_ADDRESS</b>		Decode table for address codes				
Column	Type	Len	Key	Decoded in	Null	Description
ADDR	V	6	P		N	Address code of a person, institute, company, expedition, etc.
ADDRESS	V	75			Y	Street and number
CNT	V	3	R	GNS_COUNTRY	N	Country code
COUNTRY	V	25			Y	Country name
GMTA	V	1			Y	‘Y’ if the institution has signed a general material transfer agreement, ‘N’ if not
I_NAME	V	75			N	Full name of the institute
INT_ADDR	V	12			Y	Corresponding international FAO institute code
PLACE	V	40			Y	City, town or village
PST_CODE	V	10			Y	Postal code

<b>GNS_BATCH_STORAGE</b>		Information on received safety duplicate parcels and batches of archived material				
Column	Type	Len	Key	Decoded in	Null	Description
ADDR	V	6	R	GNS_ADDRESS	N	Address code of the institute that send the safety duplicates (archived material = ‘CGN’)
BATCH_NUM	V	10	P		N	Batch number
DESCR	V	255			Y	Description of the content of the batch
NUM_BOX	N				N	Number of boxes within the batch
RCV_DATE	D				N	Date that the batch was received or created
STORAGE_LOC	V	10			N	Storage location of the batch
TOTAL	N				N	Number of seed samples in the batch

<b>GNS_COUNTRY</b>		Decode table for country codes				
Column	Type	Len	Key	Decoded in	Null	Description
CNT	V	3	P		N	Code for the country
COUNTRY	V	25			N	Country name

<b>GNS_CROP</b>		Decode table for crop codes				
Column	Type	Len	Key	Decoded in	Null	Description
CNR	N		P		N	Code for the crop
CROP	V	25			N	Crop name
G_AMOUNT	N				Y	Default number of bags for germination tests
G_CONT	N				N	Default number of seeds in a bag for germination tests
M_AMOUNT	N				Y	Default number of bags for regeneration
M_CONT	N				N	Default number of seeds in a bag for regeneration
R_AMOUNT	N				Y	Default number of bags with rest seeds
U_AMOUNT	N				Y	Default number of bags for distribution
U_CONT	N				N	Default number of seeds in a bag for distribution

<b>GNS_DIST_CUR_VW</b>		Distinct addresses for module odist (menu item: order_seed). Based on table GNS_PROC_DISTRIBUTION				
Column	Type	Len	Key	Decoded in	Null	Description
ADDR	V	6			N	Address code

<b>GNS_DISTRIBUTION</b>		Information about distribution of material to users				
Column	Type	Len	Key	Decoded in	Null	Description
ADDR	V	6	PR	GNS_ADDRESS	N	Address code of the institute to which the seed sample was sent
ANR	N				N	Accession number in the CGN collection
CNR	N		R	GNS_CROP	N	Code for the crop
D_REMARK	V	40			Y	Remark about the distribution of the accession
E_DATE	D		P		N	Date of entry in the table
GNR	V	5			Y	Generation number of the accession
NUMB_BAGS	N				N	Number of bags distributed
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number
STATUS	V	1		GNS_STATUS	N	Status of the seed sample

<b>GNS_EXPERIMENT</b>		Decode table for experiment codes				
Column	Type	Len	Key	Decoded in	Null	Description
ADDR	V	6	R	GNS_ADDRESS	N	Address code of the institute that performed the experiment
CNR	N		R	GNS_CROP	N	Code for the crop
E_DESCR	V	240			Y	Description of the experiment
E_YEAR	V	4			N	Year in which the experiment was performed
ENR	N		P		N	Code for the experiment

<b>GNS_FBSCORE_VW</b>		Experiments in fieldbook to be used in module cpyexp (menu item: move), based on table GNS_PROC_SCORE				
Column	Type	Len	Key	Decoded in	Null	Description
CNR	N				N	Shows the crop number
ENR	N				N	Shows the experiments in the fieldbook

<b>GNS_GERMINATION</b>		Results of germination tests				
Column	Type	Len	Key	Decoded in	Null	Description
ADDR	V	6	R	GNS_ADDR	N	Address code of the institute that performed the test
ANR	N				Y	Accession number in the CGN collection
CNR	N		R	GNS_CROP	N	Code for the crop
DORMANCY	N				Y	Percentage seeds with dormancy
E_DATE	D		P		N	Date of entry in the table
GNR	V	5	P		N	Generation number of the accession
METH	N		R	GNS_METHOD	N	Code of the used method
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number
SCORE	N				N	Germination result (percentage)
STATUS	V	1		GNS_STATUS	N	Status of the seed sample

<b>GNS_GERMINATION_VW</b>		Recent germination results per GNR per RNR on received and accessed material Based on table GNS_GERMINATION				
Column	Type	Len	Key	Decoded in	Null	Description
ADDR	V	6			N	Address code of the institute that performed the tests
ANR	N				Y	Accession number in the CGN collection
CNR	N				N	Code for the crop
DORMANCY	N				Y	Percentage seeds with dormancy
E_DATE	D				N	Date of entry in the table
GNR	V	5			Y	Generation number of the accession
METH	N				N	Code of the used method
RNR	V	6			N	Receipt number
SCORE	N				N	Germination result (percentage)
STATUS	V	1		GNS_STATUS	N	Status of the seed sample

<b>GNS_METHOD</b>		Decode table for method codes				
Column	Type	Len	Key	Decoded in	Null	Description
M_DESCR	V	120			N	Description of the method
METH	N		P		N	Code of the method
UNIT	V	6			N	Unit of the score

<b>GNS_O_TYPE</b>		Decode table for origin types				
Column	Type	Len	Key	Decoded in	Null	Description
O_DESCR	V	80			N	Description of the origin type
O_TYPE	V	1	P		N	Code for the origin type

<b>GNS_PARALLEL</b>		Numbers that the seed samples have or had in other collections				
Column	Type	Len	Key	Decoded in	Null	Description
ADDR	V	6	PR	GNS_ADDR	N	Address code of the institute that gave the number
ANR	N				Y	Accession number in the CGN collection
CNR	N		R	GNS_CROP	N	Code for the crop
PNR	V	10			N	Number given by the institute (ADDR) to the seed sample
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number
STATUS	V	1		GNS_STATUS	N	Status of the seed sample

<b>GNS_PASSPORT</b>		Information about identity, history and storage of seed samples (see 2.1)				
Column	Type	Len	Key	Decoded in	Null	Description
ADDI_PAS	V	240			Y	Additional passport information
ALTI	N				Y	Altitude of the collection site (in meters)
ANCEST	V	255			Y	Ancestral information
ANR	N				Y	Accession number in the CGN collection
APPROVAL	V	1			Y	Approval of Plant Protection Service to fulfil the requirements for an EU plant passport (European Commission, 1997): Y = yes, N = No
B_GNR	V	5			N	Generation number of base material
BATCH_NUM	V	10		GNS_BATCH_STORAGE	Y	Batch number of archived material
CHROMNR	V	2			Y	Number of chromosomes
CNR	N		R	GNS_CROP	N	Code for the crop
COLD_LOC	V	10			Y	Location of the distribution -bags in the cold storage room (+4 °C)
COLLNR	V	20			Y	Collection number, given by the collector
D_ADDR	V	6	R	GNS_ADDRESS	N	Code for the seed donor
DISTRICT	V	80			Y	District in which the collection site is located
E_DATE	D				N	Date of entry in the table
EST_REST_BAG	N				Y	Estimated number of seeds in the rest bag
EXTERN_ANR	V	10			Y	External accession number of external material (status 'E')
FREE_LOC	V	10			Y	Location of the rest-, regeneration-, and germination-bags in the freeze storage (-20 °C)
G_AMOUNT	N				Y	Number of germination-bags
G_CONT	N				Y	Number of seeds in the germination -bags
GENUS	V	25			Y	Genus name
GRAIN_WEIGHT	N				Y	1000- grain weight (in gram)
HIGHLIGHT	V	255			Y	Most important feature of the accession
IRRIG	V	1			Y	Code for irrigation of the collection site: += yes, - = no
LATI	V	5			Y	Geographical latitude of the collection site in degrees and minutes
LOCATION	V	255			Y	Exact location of the collection site
LONGI	V	6			Y	Geographical longitude of the collection site
LOW_GRM	V	20			Y	Reason to expect low germinability: No, Dormancy, Short day material, Wild

<b>GNS_PASSPORT</b>		Information about identity, history and storage of seed samples (see 2.1)				
Column	Type	Len	Key	Decoded in	Null	Description
LR_YEAR	V	4			Y	Year of the last regeneration of base material
LRU_YEAR	N				Y	Year of the last regeneration of user material
M_AMOUNT	N				Y	Number of regeneration –bags
M_CONT	N				Y	Number of seeds in the regeneration-bags
NA_REMARK	V	120			Y	Remarks in case of not accessed
NAME	V	80			Y	Principal name or breeding line number
O_ADDR	V	6	R	GNS_ADDRESS	Y	Code of the origin (breeder or collector)
O_CNT	V	3	R	GNS_COUNTRY	Y	Code for the origin country
O_TYPE	V	1	R	GNS_O_TYPE	Y	Code for the origin type
O_YEAR	V	4			Y	Year of first reference
POP_TYPE	V	1			Y	Code for the population type
R_CONTENT	N				Y	Content of rest-bag (in gram)
R_REASON	N		R	GNS_R_REASON	Y	Code for reason not to access
RNR	V	6	P		N	Receipt number
SAFETY_DUP_ADDR	V	6	R	GNS_ADDR	Y	Code of institute maintaining the safety duplicate
SAFETY_DUP_DATE	D				Y	Date that the CGN safety duplicate was sent
SCNR	V	4	R	GNS_SUB_CROP	Y	Code for the subcrop
SEC_NAME	V	80			Y	Secondary name of the accession
SPECIES	V	80			Y	Species name
ST_DATE	D				Y	Date that the status has been changed
STATUS	V	1		GNS_STATUS	N	Status of the seed sample
SUB_SPEC	V	80			Y	Taxonomic sub specific name
TREATY	V	10		GNS_TREATY	Y	Status of the accession in terms of international regulations
U_AMOUNT_COLD	N				Y	Number of distribution-bags in the +4 °C storage
U_AMOUNT_FREE	N				Y	Number of distribution-bags in the –20 °C storage
U_CONT	N				Y	Number of seeds in the distribution-bags
U_FREE_LOC	V	10			Y	Location of the distribution -bags in the –20 °C storage
U_GNR	V	5			N	Generation number of user material

<b>GNS_PATHS</b>		Definition of paths for structuring the collections				
Column	Type	Len	Key	Decoded in	Null	Description
PATH	V	10	P		Y	Path string
P_DESCR	V	40			Y	Description of the path
WEIGHT	N				Y	Weight of the path, i.e. relative importance

**GNS\_PROC\_ACCESS** Temporarily information on seed samples that are going to be accessed (module: accession, menu item: process, access)

Column	Type	Len	Key	Decoded in	Null	Description
ANR	N				N	Accession number in the CGN collection
E_DATE	D				Y	Date of entry in the table
GNR	V	5			N	Generation number of the stored material
G_AMOUNT	N				N	Number of germination-bags
LR_YEAR	V	4			N	Year of the last regeneration
M_AMOUNT	N				N	Number of regeneration-bags
NAME	V	15			Y	Principal name or breeding line number
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number
U_AMOUNT_COLD	N				N	Number of distribution-bags to be stored in +4 °C storage
U_AMOUNT_FREE	N				N	Number of distribution-bags to be stored in -20 °C storage

**GNS\_PROC\_ARCHIVE** Temporarily information on seed samples that are going to be archived (module: archive, menu item: process, archive)

Column	Type	Len	Key	Decoded in	Null	Description
ANR	N				N	Accession number in the CGN collection
BATCH_NUM	V	10		GNS_BATCH_STORAGE		Batch number
E_DATE	D				Y	Date of entry in the table
NA_REMARK	V	120				Remark on archivation
R_REASON	N		R	GNS_R_REASON		Reason code for archivation
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number

**GNS\_PROC\_COLS** Temporarily stored columns of an experiment for easier entrance of experiment data in GNS\_PROC\_SCORE (module: f\_book 2, menu item: fieldbook, characteristics)

Column	Type	Len	Key	Decoded in	Null	Description
ENR	N		PR	GNS_EXPERIMENT	N	Code for the experiment
METH	N		PR	GNS_METHOD	N	Code of the method
TRT	N		PR	GNS_TRAIT	N	Code for the trait

**GNS\_PROC\_DISTRIBUTION** Temporarily information about material that is going to be distributed (module: seeddist, menu item: process, order seed and handle seed)

Column	Type	Len	Key	Decoded in	Null	Description
ADDR	V	6	PR	GNS_ADDRESS	N	Address code of a person, institute, company, etc.
AMOUNT	N				N	Number of distribution-bags that are left after this distribution
ANR	N				N	Accession number in the CGN collection
CNR	N		R	GNS_CROP	N	Code for the crop
D_REMARK	V	40			Y	Remark about the distribution of the accession
E_DATE	D		P		N	Date of entry in the table
GNR	V	5			Y	Shows the generation number of the accession
NUMB_BAGS	N					Number of bags to be distributed
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number
STATUS	V	1		GNS_STATUS	N	Status of the seed sample

Column	Type	Len	Key	Decoded in	Null	Description
<b>GNS_PROC_LABELS</b> Temporarily information for generation of labels (module: labels, menu item: process, label)						
ANR	N					Accession number in the CGN collection
CNR	N		R	GNS_CROP		Code for the crop
G_LAB	N					Number of labels for germination bags
M_LAB	N					Number of labels for regeneration bags
RNR	V	6	PR	GNS_PASSPORT		Receipt number
R_LAB	N					Number of labels for rest bags
STATUS	V	1		GNS_STATUS		Status of the seed sample
U_LAB	N					Number of labels for distribution bags
<b>GNS_PROC_OBJ_SEQ</b> Temporarily stored objects (=accessions) of an experiment for easier entrance of experiment data in GNS_PROC_SCORE (module: f_book 1, menu item: fieldbook,objects)						
Column	Type	Len	Key	Decoded in	Null	Description
ANR	N				Y	Accession number in the CGN collection
CNR	V	3			N	Code for the crop
ENR	N		PR	GNS_EXPERIMENT	N	Code for the experiment
GNR	N				Y	Generation number
OBJ_SEQ	N		P		N	Sequence number of an object in an experiment
RNR	V	6	R	GNS_PASSPORT	N	Receipt number
STATUS	V	1		GNS_STATUS	Y	Status of the seed sample
<b>GNS_PROC_REJECT</b> Temporarily information on material that will get the status 'not accessed' (module: rejection, menu item: process, reject)						
Column	Type	Len	Key	Decoded in	Null	Description
ANR	N				Y	Accession number in the CGN collection
E_DATE	D				N	Date of entry in the table
R_REASON	N		R	GNS_R_REASON	N	Reason code for not accessing
R_REMARK	V	120			Y	Remark on not accessing
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number
<b>GNS_PROC_SCORE</b> Temporarily stored characterization data, to be moved to GNS_SCORE (module cpyexp, menu item: fieldbook, move)						
Column	Type	Len	Key	Decoded in	Null	Description
ANR	N				Y	Accession number in the CGN collection
CNR	N		R	GNS_CROP	N	Code for the crop
E_DATE	D				N	Entry date
ENR	N		PR	GNS_EXPERIMENT	N	Code for the experiment
GNR	V	5			Y	Generation number
METH	N		PR	GNS_METHOD	N	Code of the method
OBJ_SEQ	N		P		N	Sequence number of an object in an experiment
RNR	V	6	R	GNS_PASSPORT	N	Receipt number
S_REMARK	V	40			Y	Remark on the score
SCORE	V	8			N	Score of characterization or evaluation
STATUS	V	1		GNS_STATUS	N	Status of the seed sample
TRT	N		PR	GNS_TRAIT	N	Code for the trait

<b>GNS_PSP_REMARKS</b>		Decode information for codes used in field ADDI_PAS, table GNS_PASSPORT				
Column	Type	Len	Key	Decoded in	Null	Description
PSP_CD_DESCR	V	120			N	Remark code description
PSP_CODE	V	3	P		N	Code for the remark

<b>GNS_R_REASON</b>		Decode table for codes of reasons to archive or not to access				
Column	Type	Len	Key	Decoded in	Null	Description
R_REASON	N		P		N	Code for the reason why a seed sample has been archived or has not been accessed
REASON	V	80			N	Description of the reason to archive or not to access

<b>GNS_RECENT_GERMINATION_VW</b>		Recent germination results per RNR on received and accessed material Based on table GNS_GERMINATION				
Column	Type	Len	Key	Decoded in	Null	Description
ADDR	V	6			N	Address code of the institute that performed the tests
ANR	N				Y	Accession number in the CGN collection
CNR	N				N	Code for the crop
DORMANCY	N				Y	Percentage seeds with dormancy
E_DATE	D				N	Date of entry in the table
GNR	V	5			Y	Generation number of the accession
METH	N				N	Code of the used method
RNR	V	6			N	Receipt number
SCORE	N				N	Germination result (percentage)
STATUS	V	1		GNS_STATUS	N	Status of the seed sample

<b>GNS_RNR_PATH</b>		Path numbers and priority per accession to define the structure of the collection				
Column	Type	Len	Key	Decoded in	Null	Description
PATH	V	10		GNS_PATHS	Y	Path string
PRIORITY	N				Y	Priority of the path
RNR	V	6	P		Y	Receipt number

<b>GNS_SCORE</b>		Evaluation and characterization results				
Column	Type	Len	Key	Decoded in	Null	Description
ANR	N				Y	Accession number in the CGN collection
CNR	N		R	GNS_CROP	N	Code for the crop
E_DATE	D				N	Date of entry in the table
ENR	N		PR	GNS_EXPERIMENT	N	Code for the experiment
GNR	V	5			Y	Generation number
METH	N		PR	GNS_METHOD	N	Code for the method
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number
S_REMARK	V	40			Y	Remark on the score
SCORE	V	6			N	Score of the trait
STATUS	V	1		GNS_STATUS	N	Status of the seed sample
TRT	N		PR	GNS_TRAIT	N	Code for the trait

Column	Type	Len	Key	Decoded in	Null	Description
<b>GNS_SCORE_CROP_VW</b> Number of scores, experiments and methods per crop/trait-combination, based on table GNS_SCORE. Includes accessed and received samples						
CNR	N				N	Code for the crop
EXP	N				Y	Code for the experiment
METH	N				Y	Code for the method
SCORES	N				Y	Number of characterization or evaluation data
TRT	N				N	Code for the trait
<b>GNS_SCORE_EXP_VW</b> Overview of data per experiment, with description of TRT and METH, based on tables GNS_SCORE, GNS_TRAIT, GNS_METHOD						
Column	Type	Len	Key	Decoded in	Null	Description
CNR	N				N	Code for the crop
ENR	N				N	Code for the experiment
MDESC	V	120			N	Description of the method
METH	N				N	Code for the method
TRAIT	V	80			N	Name of the trait
TRT	N				N	Code for the trait
UNIT	V	6			N	Unit of the score
<b>GNS_SCORE_VW2</b> Overview of crop/trait/method combinations in GNS_SCORE. Based on table GNS_SCORE						
Column	Type	Len	Key	Decoded in	Null	Description
CNR	N				N	Code for the crop
METH	N				N	Code for the method
TRT	N				N	Code for the trait
<b>GNS_STATUS</b> Overview of crop/trait/method combinations in GNS_SCORE. Based on table GNS_SCORE						
Column	Type	Len	Key	Decoded in	Null	Description
DESCR	V	40			N	Description of the status
STATUS	V	1	P		N	Status of the seed sample: R=Received, A=Accessed, N=Not accessed, C=Archive, E=External
<b>GNS_SUB_CROP</b> Decode table for sub crop codes						
Column	Type	Len	Key	Decoded in	Null	Description
CNR	N		R	GNS_CROP	N	Code for the crop
SCNR	V	4	P		N	Code for the subcrop
SUB_CROP	V	80			N	Name of the subcrop

---

<b>GNS_SYS_PARAMETERS</b>		Parameter table to facilitate the use of other applications and hardware from GENIS				
Column	Type	Len	Key	Decoded in	Null	Description
DATATYPE	V	1			N	Datatype of the parameter
ID	N		P		N	Id of the parameter
NAME	V	40			N	Name of the parameter
DESCRIPTION	V	1000			Y	Description of the parameter
VALUE	V	200			N	Value of the parameter

---

<b>GNS_TRAIT</b>		Decode table for trait codes				
Column	Type	Len	Key	Decoded in	Null	Description
HEADER1	V	6			N	First line of header string
HEADER2	V	6			Y	Second line of header string
TRAIT	V	80			N	Name of the trait
TRT	N		P		N	Code for the trait

---

<b>GNS_TREATY</b>		Decode table for treaty codes				
Column	Type	Len	Key	Decoded in	Null	Description
DESCR	V	100			N	Description of the treaty
TREATY	V	20			N	Treaty name
TTY	N		P		N	Code for the treaty

---

## 3.2 Menu and modules

To gain access to the screens a windows drop down menu is used. Some menu items have restricted access, depending on the Oracle User name. The menu items to which a user has no access have a grey background; the accessible menu items are bright black and white.

Several screens (windows) can be opened simultaneously. Some caution is needed here since it can lead to apparent inconsistency of the data. Changes in one screen are not automatically changed in other screens. For performing several specified actions it is necessary to close all other windows, except the one to work with. These actions are changing the status of accessions, ordering seed, and handling seed orders.

When a screen is opened, a toolbar appears, representing some frequently used functions. In the following table the full menu with their sub menus is given; the modules that can be accessed via the menu; the name, purpose and type of the module; and the access to the menu for the different user types. Submenus within menus are indented.

*Legend: Type gives the type of the module: m= menu, s = screen, u = utility; Access gives the access for the different users to the module: y = yes.*

Menu Item	Module name	Purpose	Type	Access			
				dba	smr	adm	usr
Main menu	Main	Access to all screens	m	y	y	y	y
Description	Description	Access to all description screens	m	y	y	y	y
Passport	Passport	Query, enter and update passport data	s	y	y	y	y
Parallel	Parallel	Query, enter and update parallel data	s	y	y	y	y
Characterization	Characterization	Access all screens about characterization and evaluation	m	y	y	y	y
Accession_Receipt	Charinf	Query characterization and evaluation data per receipt number or CGN number	s	y	y	y	y
Crop_overview	Charcrop	Query overview of characterization and evaluation data per crop	s	y	y	y	y
Experiment_overview	Charexp	Query overview of characterization and evaluation data per experiment	s	y	y	y	y
Crop/trait/method_comb	Sct	Query trait/method combinations per crop	s	y	y	y	y
Fieldbook	Fieldbook	Access the fieldbook screens	m	y	y	y	
Objects	f_book1	Data entry of receipt numbers and CGN numbers of an experiment	s	y	y	y	
Characteristics	f_book2	Data entry of traits and methods of an experiment	s	y	y	y	
Update	Scores	Update fieldbook records of an experiment	s	y	y	y	
Print	Prexprep	Run a detailed report of one experiment for purpose of checking	u	y	y	y	

Menu Item	Module name	Purpose	Type	Access			
				dba	smr	adm	usr
Move	Cpyexp	Move all records from an experiment from GNS_PROC_SCORE to GNS_SCORE	s	y	y	y	
Rejection and Archive	Not_accessed	Access all screens with information on not accessed or archived seed samples	m	y	y	y	y
Passport	Rejpassp	Query, enter and update passport data of not accessed seed samples	s	y	y	y	y
Parallel	Rejparal	Query, enter and update parallel data of not accessed seed samples	s	y	y	y	y
Germination	Rejgerm	Query, enter and update germination data of not accessed seed samples	s	y	y	y	y
Distribution	Rejdist	Query, enter and update distribution data of not accessed seed samples	s	y	y	y	
Characterization	Rejscore	Query, enter and update characterization data of not accessed seed samples	s	y	y	y	y
External	Exteral-collection	Access all screens with information on external collections	m	y	y	y	y
Passport	Extpassp	Query, enter and update passport data of external collections	s	y	y	y	y
Parallel	Extparal	Query, enter and update parallel data of external collections	s	y	y	y	y
Distribution	Extdist	Query, enter and update distribution data of external collections	s	y	y	y	
Characterization	Extscore	Query, enter and update characterization data of external collections	s	y	y	y	y
Seed management	Seed_ management	Access all seed management screens	m	y	y	y	
Distribution	Seeddist	Query distribution data	s	y	y	y	
Storage	Stogrm	Query and update storage data.	s	y	y	y	
Germination	Allgrm	Query, enter and update all germination results.	s	y	y	y	
Batch storage	Safety	Query, enter and update information on received safety duplicates and other batch storage	s	y	y	y	
Process	Orders_ procedures	Access to all orders and processes screens	m	y	y	y	
Order_ seed	Odist	Enter CGN numbers for distribution and run a report of passport data	s	y	y	y	
Handle_ seed	Prdist	Query CGN numbers for distribution and handle them	s	y	y		

Menu Item	Module name	Purpose	Type	Access			
				dba	smr	adm	usr
Access	Accession	Assign accession numbers and change status from 'R' to 'A'	s	y	y		
Reject	Rejection	Changing status from 'R' to 'N'	s	y	y	y	
Archive	Archive	Changing status from 'R' or 'A' to 'C'	s	y	y		
Label	Labels	Query, enter and update CGN numbers for printing labels	s	y	y		
Decode	Decinf	Access all decode screens	m	y	y	y	y
Address	Addrinfo	Query, enter and update address codes	s	y	y	y	y
Country	Cntinfo	Query, enter and update country codes	s	y	y	y	y
Crop	Cropinfo	Query, enter and update crops and their codes	s	y	y	y	y
Subcrop	Scrpinfo	Query, enter and update sub_crops	s	y	y	y	y
Origin type	Origininfo	Query, enter and update origin types	s	y		y	y
Experiment	Expinfo	Query, enter and update experiments and their codes	s	y	y	y	y
Trait	Trtinfo	Query, enter and update traits	s	y	y	y	y
Method	Methinfo	Query, enter and update methods and their codes	s	y	y	y	y
Reason	Rsninfo	Query, enter and update reasons for not accessing	s	y	y	y	y
International_address	Adrint	Query, enter and update international (FAO)address codes	s	y	y	y	y
Passport_rem_co	Remcode	Query, enter and update remark codes used in passport screen	s	y	y	y	y
Status	Status	Query, enter and update status codes	s	y	y	y	y
Treaty	Treaty	Query, enter and update treaty codes	s	y	y	y	y
Reports	Make_reports	Access to all report menus	m	y	y	y	y
Excel Reports	Excel Reports	Starts an Excel reports workbook (see Appendix VI)	u	y	y	y	y



## 4. Detailed description of some tables

Apart from the formal rules considering field length, the case, etc., the tables follow many other conventions about how they should be used. These will be described in detail for the following tables containing descriptive information:

- The GNS\_PASSPORT table (4.1).
- The GNS\_PARALLEL table (4.2).
- The GNS\_SCORE table (4.3).

After a brief introduction of the table, the fields will be discussed: first the field name will be given, followed by the colloquial name between brackets and a description.

The fields are ordered following the Genis screen order.

### 4.1 The GNS\_PASSPORT table

The GNS\_PASSPORT table is used for storing passport information of the accessions in the CGN germplasm collection. This table tries to give detailed information about the accessions' original finding place, ancestors, donors, history and taxonomy. Also information on the current storing place, regeneration and amount of seeds are part of the GNS\_PASSPORT table. For management purposes these fields of the passport table appear under a different menu together with recent germination data. Since the fields are long one should enter (long) text as readable as possible. Not: 'c. 3km s. of Lutjebroek', but: 'about 3 kilometer south of Lutjebroek'.

For clarity reasons, abbreviations should be avoided when entering data. If they can not be avoided, or when they are very well interpretable (e.g. S.W.-France) the rules given in Appendix II should be followed.

The fields of GNS\_PASSPORT

ANR (CGN number)

The unique accession number of each accession. These numbers are sequentially given to each accession that enters one of the CGN collections.

RNR (Receipt number)

A unique number given when a seed sample is nominated for inclusion in the collection (see 2.3.1 Accession / no accession). The receipt number has 6 digits. The first two digits represent the year; the last four are a sequential number.

EXTERN\_ANR (External accession number)

The accession number of material that is part of an external collection.

BATCH\_NUM (Batch number of archived material)

A number that gives information of the storage location of a seed sample that has been archived. The number is entered in the procedure that changes the status of the material from 'A' or 'R' to 'C'. With this number one can find, in GNS\_BATCH\_STORAGE, the location of the box in which the material is stored.

**STATUS (Status)**

A one letter code. 'A' for an 'Accessed' seed sample, 'R' for a 'Received' seed sample, 'N' for a 'Not Accessed' seed sample, 'C' for an 'Archived' seed sample and 'E' for a sample that is part of an external collection.

**CNR (Crop number)**

The code number of the crop. Each crop in the CGN collections has a crop number. A number will usually correspond to a genus. A complete list of current crop codes is decoded in table GNS\_CROP. The database administrator can create new codes.

**SCNR (Subcrop number)**

The subcrop code, a sub division of the crop in user oriented groups, e.g. the crop 'Lettuce' is divided in butterhead lettuce, crisp lettuce, etc. All current subcrop codes are decoded in table GNS\_SUB\_CROP. The database administrator can create new sub crop codes.

**GENUS (Genus)**

The Latin name for the genus of the accession. The first letter should always be a capital. The name can be followed by an author name if available. This author name may contain spaces. E.g.: 'Triticum L.'

**SPECIES (Species)**

The name of the species. This name should always be written in lower case, followed by the authors' name if available. E.g.: 'turgidum (L.) Thell.'

**SUB\_SPEC (Subspecific name)**

A subspecific name, the name of the subspecies, botanical variety, forma or cultivar group to which the accession belongs, preceded by 'subsp.', 'var.', 'f.' or 'group' (cultivated variety names are not entered in this field but in the next: Principal name). The subspecies (subsp.), botanical variety (var.) and forma (f.) name always starts with a lower case letter, a cultivar group (group) always with a capital. If this subdivision is followed by an author name, this may contain spaces.

**NAME (Principal name)**

The principal name of the accession, this can be the cultivated variety name, the local name of a land race or the line number of research material. It might even be a PI-number or collection number, if the accession is known under that number. Variety- and land race names should always start with a capital.

**SEC\_NAME (Secondary name)**

The accession's other names besides the principal name. This can be a synonym of the cultivated variety name or land race name. If there is more than one name to be entered in this field they should be separated by a ';'. E.g.: 'Matar; Ethiopia 31'

**ANCEST (Ancestor)**

Ancestral information about the accession. The first position is used for a letter code, indicating what kind of information follows. Following codes can be used:

Code	Description
C	Interspecific cross
I	Inbred line from
M	Mutant from
P	Pedigree
S	Selection from
X	Other

This code is immediately followed by a ':', after which the information follows. It is also possible to give only a code with a '?'. E.g. if it is only known that the accession is a mutant; 'M:' is used.

Pedigrees are given using the system of Purdy et al (1968):

not: (((Isaria x Haisa) x WMRII) x Donaria) x Union

but: Isaria/Haisa//WMRII/3/Donaria/4/Union

not: (Hanna x Atlas7) x (Turk x Atlas8)

but: Hanna/7\*Atlas//Turk/8\*Atlas

If a number is not known, a '?' can be used instead.

E.g.: 'S: Irene'

E.g.: 'P: C300/?\*Rondo//Furore'

#### O\_TYPE (Origin type)

The type of the collection site. The following codes (decoded in GNS\_O\_TYPE) can be used:

Code	Description
1	Wild habitat
2	Ruderal
3	Farm field
4	Farm store, threshing place
5	Backyard
6	Local market
7	Commercial market, seed trade
8	Institute, university, genebank, breeding company
9	Other

#### O\_CNT (Origin country)

The code of the origin country, the country where the accession was collected or bred. O\_CNT is decoded in GNS\_COUNTRY. The DBA can assign new country codes.

#### TREATY (Information on access and benefit sharing of the accession)

The status of the accession in terms of international regulations concerning the access and ownership of the material: 'Treaty' indicating that the accession comes under the 'International Treaty on PGRFA', 'Pre CBD' indicating that the accession doesn't come under the Treaty and is from before the Convention on Biodiversity, and 'Post CBD' if it is from after the CBD.

## DISTRICT (District)

The name of the district (province, county, state, mountain range) where the collection site was located.

E.g.: 'Punjab'

To clarify the district and/or province names it may be preceded by 'Province: ', 'District: ' or 'Estate: '. E.g. 'District: Potosi, Province: Frias'.

## LOCATION (Location)

Exact description of the location of the collection site.

E.g.: 'Jullundur'

E.g.: '3 km west of Lutjebroek'

E.g.: 'Road Uzgen to Mirzaaki, c. 4 km after Mirzaaki, Salaam Alik'

## ADDI\_PAS (Passport remark)

Important additional passport data, preceded by a three-letter code, immediately followed by a ':' and the information. The codes are decoded in GNS\_PSP\_REMARKS. Until now following codes have been used:

Code	Description
ASP	Aspect at the collection site (S=South, W=West, E=East, N=North, and combination e.g. SSW)
BGR	BGRC accession numbers and other numbers
CDA	Collection day/month
CHI	Coded habitat information (see Appendix I)
CHR	Possible chromosome number when unknown
CUL	Cultivation method
EBS	Erwin Bauer Sortiment number
EPO	Effective population size at the collection site (followed by the number of flowering plants)
HAB	Habitat at collection site (followed by a description of the habitat)
MIA	More information is available (followed by an information source or a clear reference)
MON	Number in the IVT Monster book
NIE	NAME in English
OBN	Original botanical name (the botanical name used by the seed donor)
PLA	Number of collected plants
RCO	Rest collection number
REM	Remark of any kind
SDO	Secondary donor (followed by the name or code of the institute who gave the material to the one who gave it to the CGN)
SLO	Slope at collection site (in degrees)
SOI	Soil type at collection site
SPH	Soil pH-value at collection site (E.g. 7.6 or 10.1)
TYP	Type of variety or selection (groups within sub crops)
WAC	Wageningen Potato Collection number

If there is more than one item in the Passport remark field, they are separated by a ‘;’.

E.g.: ‘TYP:Agilo; CUL:under glass’

The code CHI (Coded habitat information is declared in Appendix I)

#### HIGHLIGHT (Highlight)

The HIGHLIGHT field used for storing highly relevant information on the accession for which there is no suitable place somewhere else. The field is entered as free text. Each highlight is supposed to be interpretable as such, so no codes or relative scales should be used.

#### E\_DATE (Entry date)

Automatically generated date when entering a new record.

#### D\_ADDR (Donor address)

The address code of the donor of the accession. The institute, genebank, university, company etc., that gave the accession to the CGN. This field always has a value and is decoded in table GNS\_ADDRESS. CGNDBA, CGNSMR and CGNADM can create new codes.

#### POP\_TYPE (Population type)

A code for the population type. Following population types are available:

Code	Description
W	Wild
L	Land variety
B	Breeder’s variety
R	Research material
H	Hybrid

A land variety is defined as a population that is (has been) grown, but that is not a product of plant breeding, as opposed to a breeder’s variety. A land variety, in which some superficial selection has been done, like separation in several morpho-types, is still considered to be a land variety.

#### CHROMNR (Chromnumber)

Number of chromosomes in a somatic cell.

#### O\_ADDR (Origin address)

The address code of the institute where the accession originated. This can be a breeding company, but also the expedition who collected the accession. Like all addresses O\_ADDR is decoded in GNS\_ADDRESS. CGNDBA, CGNSMR and CGNADM can create new codes.

#### O\_YEAR (Origin year)

The year in which the accession was first documented. This can be the year it was introduced as a cultivar, or the year it was collected. For Dutch cultivars we use the year that breeders right was granted.

#### COLLNR (Collection number)

The number that was attributed to the accession when it was collected. This is the most original number of the accession.

E.g.: ‘Rm 8761/5’

**LONGI (Longitude)**

The longitude of the collection site. The first three positions are for the degrees, the next two for the minutes, and the last for the hemisphere (E or W). If the minutes are not known, these positions remain blank.

E.g.: '03842E', '103 W'

**LATI (Latitude)**

The latitude of the collection site. The first two positions are for the degrees, the next two for the minutes, and the last for the hemisphere (N or S). If the minutes are not known, these positions remain blank.

E.g.: '4133N'

**ALTI (Altitude)**

The altitude of the collection site in meters above sea level.

E.g.: '2500'

E.g.: '-3'

**IRRIG (Irrigation)**

Code indicating whether the collection site was irrigated ('+') or not ('-').

**APPROVAL (Approval)**

Approval of Plant Protection Service, Wageningen, The Netherlands, after checking for seed-borne diseases to fulfil the requirements for an EU plant passport (European Commission, 1997): Y = yes, N = no.

**COLD\_LOC (Storage location +4 °C room)**

Location of seed samples for distribution in the cold storage room (+4 °C).

COLD\_LOC consists of a three digit shelf number and a four digit box number separated by a '-'. E.g. '034-2056'.

**FREE\_LOC (Storage location -20 °C room)**

Location of the seeds for germination, multiplication and rest seed in the freeze storage (-20 °C).

FREE\_LOC has the same format as COLD\_LOC.

**U\_FREE\_LOC (Storage location -20 °C room)**

Location of the seeds for distribution in the freeze storage (-20 °C). U\_FREE\_LOC has the same format as COLD\_LOC.

**U\_AMOUNT\_COLD (Number of bags with seeds for distribution in the +4 °C storage)**

Number of bags with seeds for distribution stored in the +4 °C room.

**U\_AMOUNT\_FREE (Number of bags with seeds for distribution and/or user regeneration in the -20 °C storage)**

Number of bags with seeds for distribution and/or user regeneration stored at -20 °C freezers.

**G\_AMOUNT (Number of bags for germination tests)**

Number of bags with seeds for germination tests. They are stored at -20 °C.

**M\_AMOUNT (Number of bags for base regeneration)**

Number of bags with seeds for base regeneration. When accessing a seed sample the number of bags reserved for regeneration is two. They are stored at -20 °C.

U\_CONT (Number of seeds per user bag)

Number of seeds in the distribution bags. This number is dependent on crop and/or total amount of seeds.

G\_CONT (Number of seeds per germination bag)

Number of seeds in the germination bags. 100 or 200, crop and total amount dependent.

M\_CONT (Number of seeds per regeneration bag)

Number of seeds in the regeneration bags, crop dependent.

R\_CONTENT (Amount of seeds in per rest bag (gram))

The amount of seed not used for previous purposes. Stored in one or two bags the number of seeds is crop and total-amount dependent.

EST\_REST\_BAG (Estimated number of seeds in the rest bag)

The amount of seeds in the restbag. Calculated if the 1000-grain weight is known and entered in GRAIN\_WEIGHT

GRAIN\_WEIGHT (1000 grain weight of the seeds)

The weight of 1000 seeds in gram.

ST\_DATE (Status changed date)

Date that the status of the seed sample has changed. From 'R' to 'A' in case of new accessed material. From 'R' or 'A' to 'N' in case of not accessed. From 'R' or 'A' to 'C' in case of archive (see also: STATUS).

B\_GNR (Storage generation number of base material)

Number of regenerations of the base material (multiplication and germination bags) since arrival at CGN. When seeds arrive the default B\_GNR 0 (zero) is assigned.

U\_GNR (Storage generation number of user material)

Number of regenerations of the user material (user bags and rest bag). U\_GNR changes when a so-called user regeneration takes place. When the user bags and rest seeds are finished and the germination rate of the base material has not dropped new user material can be generated out of the last user bags. When user seed is generated from material with U\_GNR 1, the U\_GNR becomes 1a. If it is generated from material with U\_GNR 1a it becomes 1aa. If no user regeneration has taken place or directly after a base multiplication, B\_GNR is equal to U\_GNR.

LR\_YEAR (Last base regeneration year)

Last regeneration year of base material.

LRU\_YEAR (Last user regeneration year)

Last regeneration year of user material.

LOW\_GRM (Reason low germinability)

Short indication of the possible reason to expect low germination results: No, Dormancy, Short day material, Dormancy and wild, Wild.

SAFETY\_DUP\_ADDR (Safety duplicate)

Address where the safety duplicate of the accession is kept.

SAFETY\_DUP\_DATE (Safety dup. sent)

Date that the safety duplicate was sent to SAFETY\_DUP\_ADDR.

## 4.2 The GNS\_PARALLEL table

The GNS\_PARALLEL table is used for storing all numbers that accessions of the CGN collections have in other collections. For instance, if a CGN accession was requested or taken over from another collection holder (or the other way around), or if a collected sample was split shortly after collection and divided between two genebanks. Sometimes material is identified by an accession number or introduction number like PI- and CI-numbers, these numbers are also stored in the GNS\_PARALLEL table. It is also possible that specific numbers of accessions from other collections are stored in the GNS\_PARALLEL table. This can occur as the accession is found to be a duplicate of a CGN accession.

One has to realise that accessions with parallel numbers are probably not genetic identical to the CGN material. However, the origin population might be the same, genetic composition can differ due to several causes like splitting or bulking of accessions, or different regeneration methods used by genebanks.

The table consists of the following fields: ADDR, ANR, CNR, PNR, RNR and STATUS. All fields of GNS\_PARALLEL have to have a value except ANR.

ANR (CGN number)

The unique accession number of each accession. These numbers are sequentially given to each accession that enters one of the CGN collections.

RNR (Receipt number)

A unique number given when a seed sample is nominated for inclusion in the collection (see 2.3.1 Accession / no accession). The receipt number has 6 digits. The first two digits represent the year; the last four are a sequential number.

CNR (Crop number)

The code number of the crop of the accession. This number is duplicated in the GNS\_PASSPORT table, but included for integrity reasons. A complete list of current crop codes can be decoded from GNS\_CROP.

ADDR (Address)

The address code of the institute where the accession has, or had, a number. These address codes are decoded in table GNS\_ADDRESS.

PNR (Parallel number)

The number of the corresponding accession in the collection of the institute identified in the previous field.

## 4.3 The GNS\_SCORE table

The GNS\_SCORE table is used for storing all observations made on the accessions. The table consists of the following fields: ANR, RNR, STATUS, CNR, GNR, ENR, TRT, METH, E\_DATE, SCORE and S\_REMARK.

All fields of GNS\_SCORE, except ANR, GNR and S\_REMARK, have to have a value. The fields are:

ANR (CGN number)

The unique accession number of each accession. These numbers are sequentially given to each accession that enters one of the CGN collections.

**RNR (Receipt number)**

A unique number given when a seed sample is nominated for inclusion in the collection (see 2.3.1 Accession / no accession). The receipt number has 6 digits. The first two digits represent the year; the last four are a sequential number.

**CNR (Crop number)**

The code number of the crop of the accession.

**GNR (Generation number)**

The generation number of the scored population. In the GNS\_SCORE table GNR can refer to an U\_GNR as well as a B\_GNR of the GNS\_PASSPORT table dependent on which material is used. It indicates from which regeneration the seed resulted, and can be used to trace possible mistakes in seed handling. Be aware that new-formed seeds are next generation.

**ENR (Experiment number)**

A code number of an experiment, regeneration, or some other event, that resulted in the score. The experiment numbers can be decoded from GNS\_EXPERIMENT, which gives a description of the experiment. The database administrator (DBA) can create new codes.

**E\_DATE (Entry date)**

The date the record entered the database. It is automatically given to each new record.

**TRT (Trait number)**

A code number for the trait that was scored. Traits are decoded in GNS\_TRAIT. The database administrator can create new codes.

**METH (Method number)**

A code number for the method that was used in scoring. This code numbers are decoded in GNS\_METHOD. The database administrator can create new codes.

**SCORE (Score)**

The actual score for the trait coded by TRT, using method METH during experiment ENR. In the score variation within the accession may be indicated using a standardised system (see Appendix IV).

**S\_REMARK (Remark)**

Any remark about the score.

## 4.4 The GNS\_PATHS and GNS\_RNR\_PATH tables

The GNS\_PATHS and GNS\_RNR\_PATH tables are used for storing information about the structure of the collections. GNS\_PATHS contains path-strings that describe the hierarchical structure of the collection. For example: '06' is the path string of the lettuce collection since it has crop number 6. In this collection a distinction can be made in the cultivated and the wild material, resulting in three new path-strings: '061' for the cultivated material, '062' for the wild material and '060' for the material of which it is unknown if it is cultivated or not. The cultivated material could be split up into '0611' being butterhead lettuce, '0612' being crisp lettuce, etc. GNS\_RNR\_PATH stores for each RNR in GNS\_PASSPORT the most detailed path-string. For example, an accession could have the path-string '0711331' indicating that it is genus Brassica ('07'), species oleracea ('071'), cultivated ('0711'), originating in Europe ('07113'), etc.

The fields of GNS\_PATHS:

PATH (Path-string)

The string identifying one group in the hierarchy of a crop. There is a two-character path-string for each crop corresponding to the crop number, e.g. '06' for crop 6. All 'parent strings of path-strings should exist; if '061553' is a path-string, '06155' should be a path-string too.

P\_DESC (Path description)

A short description of the path, relative to its parent. For example, if '06' is the crop 'lettuce' then the description of '061' as 'cultivated' is sufficient, it is not necessary to say 'cultivated lettuce'.

WEIGHT (Weight)

The relative importance of a path relative to the other paths below its parent. This value is used when constructing a core selection. If there are three paths below '06': '060', '061' and '062' with weights 0.0, 0.6 and 0.4 respectively, 60% of the accessions will be selected from material with path-string '061' and 40% of '062'.

The fields of GNS\_RNR\_PATH:

PATH (Path string)

See above.

PRIORITY (Priority)

Priority within all material with the same path-string when selecting a core selection, the higher the priority the sooner it will be selected.

RNR (Receipt number)

A unique number given when a seed sample is nominated for inclusion in the collection (see 2.3.1).

## 5. References

- European Commission, 1997.  
Commission Directive 97/46/EC. Official Journal of the European Communities No L 204: 43-46.
- Hintum, Th.J.L van, 1993.  
A computer compatible system for scoring heterogeneous populations. *Genetic Resources and Crop Evolution* 40: 133-136.
- Hintum, Th.J.L. van, 1999.  
The core selector, a system to generate representative selections of germplasm accessions. *Plant Genetic Resources Newsletter* 118: 64-67.
- Hintum, Th.J.L. van & F. Menting, 2001.  
Manual. Procedures for updating Genis data on the Internet. Centre for Genetic Resources (CGN), Wageningen, The Netherlands.
- Hintum, Th.J.L. van & Th. Hazekamp, 1992.  
Genis Data Dictionary, July 1992. Centre for Plant Breeding and Reproduction Research (CPRO-DLO), Centre for Genetic Resources, The Netherlands (CGN), Wageningen, The Netherlands.
- IPGRI, 1997.  
Multicrop passport descriptors (final version). In: E. Lipman, M.W.M. Jongen, Th.J.L. van Hintum, T. Gass and L. Maggioni (comp.). *Central Crop Databases: Tools for Genetic Resources Management*. International Plant Genetic Resources Institute, Rome, Italy / CGN, Wageningen, The Netherlands.
- Purdy, L.H, W. Q. Loegering, C. F. Konzak, C. J. Peterson & R. E. Allen, 1968.  
A proposed standard method for illustrating pedigrees of small grain varieties. *Crop Science* 8: 405-406.



## Appendix I.

### Codes used in CHI

Format CHI (GNS\_PASSPORT, Passport remark column (ADDI\_PAS): 'CHI:abcdefghi' where 'abcdefghi' are nine codes for the following items:

Position	Item	Codes	
a	landform	1 swamp	6 hilly dissected
		2 flood plain	7 steeply dissected
		3 plain level	8 mountainous
		4 undulating	9 other (specify using 'HAB:')
		5 hilly	
b	site	0 level	5 mid slope
		1 crest, summits	6 terrace
		2 escarpment	7 lower slope
		3 round summit	8 open depression
		4 upper slope	9 closed depression
c	soil texture	1 sand,sandy loam	3 clay loam, clay and silt
		2 loam, silty loam	4 highly organic
d	stoniness	0 none	3 tillage difficult
		1 tillage unaffected	4 tillage impossible
		2 tillage affected	5 essentially paved
e	depth of soil	1 < plough depth	3 > plough depth
		2 = plough depth	4 very deep
f	salinity	0 none	2 medium
		1 low	3 high
g	drainage	1 imperfect	3 well drained
		2 moderate	4 excessive
h	soil color	1 black	4 orange
		2 brown	5 yellow
		3 red	6 other (specified on collection form)
i	Soil nutrient content	1 poor	3 high
		2 medium	

The '-' sign is used if there is no information.

Example : CHI:5420--1--



# Appendix II.

## Symbols and abbreviations

### Introduction

Because the information in the GNS\_PASSPORT table will also be used by people not familiar with this document, the first rule should be, to keep everything as understandable as possible. Since the fields are large enough to store many characters, abbreviations should be avoided. The following lists are presented under the credo 'if we (have to) do it, let's all do it the same way'.

### Symbols

#	count / number
-	from ... to, continuum between the two
/	both appear
>	the available space was too small, the rest of the information is in the P_REMARK column

### Cardinal points

N.	North
S.	South
W.	West
E.	East
C.	Central

Always with a capital and followed by a dot, without space and with hyphen.  
E.g.: S.-Limburg, N.W.-France, C.-Poland.

### Abbreviations

Abbreviations can not be used because they are in this list, but because they are clear in the context.

a.o.	and others / among others
cont.	contamination / contaminated
cv.	cultivar
fr.	from
gr.	group
lv.	landvariety
mnt.	mountain(s)
pl.	plant
rd.	road
var.	botanical variety
km.	kilometre
c.	circa, approximately



## Appendix III.

# Privileges on tables/views for Oracle User Names

S = select, I = insert, U = update, D = delete

Table / View	CGNDBA	CGNSMR	CGNADM	CGNUSR
GNS_ADDRESS	SIUD	SIU	SIU	S
GNS_BATCH_STORAGE	SIUD	SIU	S	S
GNS_COUNTRY	SIUD	S	S	S
GNS_CROP	SIUD	S	S	S
GNS_DISTRIBUTION	SIUD	SI	S	S
GNS_DIST_CUR_VW	S	S	S	
GNS_DUPLIC_RCV	SIUD	SIUD	S	S
GNS_EXPERIMENT	SIUD	S	S	S
GNS_FBSCORE_VW	S		S	
GNS_GERMINATION	SIUD	SIU	S	S
GNS_GERMINATION_VW	S	S	S	S
GNS_METHOD	SIUD	S	S	S
GNS_O_TYPE	SIUD	S	S	S
GNS_PARALLEL	SIUD	S	SIU	S
GNS_PASSPORT	SIUD	SU	SIU	S
GNS_PATHS	SIUD		S	
GNS_PROC_ACCESS	SIUD	SI		
GNS_PROC_ARCHIVE	SIUD	SIU		
GNS_PROC_COLS	SIUD		SIUD	
GNS_PROC_DISTRIBUTION	SIUD	SIUD	SIUD	
GNS_PROC_LABELS	SIUD	SIUD		
GNS_PROC_OBJ_SEQ	SIUD		SIUD	
GNS_PROC_REJECT	SIUD	SIU	SIU	
GNS_PROC_SCORE	SIUD		SIUD	
GNS_PSP_REMARKS	SIUD	S	S	S
GNS_RECENT_GERMINATION_VW	S	S	S	S
GNS_RNR_PATH	SIUD		S	
GNS_R_REASON	SIUD	S	S	S
GNS_SCORE	SIUD	S	SIU	S
GNS_SCORE_CROP_VW	S	S	S	S
GNS_SCORE_EXP_VW	S	S	S	S
GNS_SCORE_VW2	S	S	S	S
GNS_STATUS	SIUD	S	S	S
GNS_SYS_PARAMETERS	SIUD		SIUD	
GNS_SUB_CROP	SIUD	S	S	S
GNS_TRAIT	SIUD	S	S	S
GNS_TREATY	SIUD	S	S	S



## Appendix IV.

# System for scoring heterogeneous populations

### Introduction

In genebanks the importance of within population variation is very high, the small deviating fractions within accessions often contain interesting genotypes. The importance of that variation conflicts with the large numbers of populations to be scored, and thus the limited time that can be spend per observation. To solve this problem a system has been developed that allows for quick scoring of heterogeneous populations (Hintum, 1993). After an experience of several years and many thousands of observations, it proved to be very practical.

### The system

The proposed system for scoring heterogeneous populations can only be used if the states of the trait are represented by one digit, and the symbols 'x' and '=' are not yet defined. It follows these formal rules:

- Put the scores of the fractions observed in the population behind each other in order of decreasing fraction size.
- If there is only one fraction, put an '=' -sign behind the score.
- If the ratio of two subsequent fractions is between 1.5 and 5.0, put one 'x' -sign between the two fractions, if the ratio is higher than 5.0, put two 'x' -signs in between.

This means that scores of homogeneous populations are followed by an '=' -sign. Scores of fractions are put behind each other with the largest in front, followed by the second largest etc., while between fractions with a large difference in size a 'x' and with a very large difference in size a 'xx' is scored.

For traits with distinct states this system can easily be learned and applied. When the states form a continuous scale the system is less suitable, but still useful if the differences between the states are not too small.

An algorithm for decoding the scores back to frequencies is available from the author.

### Examples

Distinct states:

Consider the trait 'flower color' with the distinct states '1: white, 2: red, 3: yellow'.

- A population with only yellow flowers (100%) will get the score '3='.
- If there are only very few red flowers (5%) amongst the yellow (95%) the score will be '3xx2'.
- If a population consists of mainly yellow (70%), some red (28%) and very few white flowers (2%) the score will be '3x2xx1'.
- If the number of white and red flowers is nearly equal, only a few more white (55%) the score will be '12'.

Continuous scale:

Consider the continuous trait 'plant length' on the scale '1=very short, 3=short, 5=medium, 7=long, 9=very long'.

#### IV - 2

- A homogeneous population with only long plants gets the score '7=1'.
- If there are some (20%) distinct very short plants amongst the long ones the score will be '7x1'.
- If there is a continuous variation between medium and very long plants, skewed to the medium length the score will be '75x9', since there are mainly long plants, many medium and a few very long.

# Appendix V.

## Genis on the Internet

### Introduction

The Internet has become an important tool for genebanks to make information on their collections available for the user community. Nowadays almost all of the information in Genis is online searchable and downloadable via Internet. Besides the information in Genis, also general information about CGN and its staff, protocols and activities is provided on the CGN web site.

### Online searchable

The user-relevant passport information in Genis is online searchable. By using a shopping cart, the user can make his/her own selection. For some crops it is also possible to make use of a core selector, which generates a (genetically) broad selection of accessions matching the criteria the user enters (Hintum, 1999). Selected material can be ordered online.

For some crops also evaluation data are online searchable. To make evaluation data searchable for users they often have to be rearranged in a well-interpretable form, providing one score per accession and trait.

### Downloadable

Complete sets of passport data are downloadable per crop. These downloadable passport data follow also the IPGRI/FAO multi-crop passport descriptor list (IPGRI 1997).

Raw evaluation data are downloadable per crop-trait combination. Unlike the online searchable evaluation data these data are exactly the same as they are stored in Genis.

All Internet files are updated every two months.

### Procedures

To make data online searchable an Access database is used. HTML and ASP files make it possible to communicate with this database via Internet in a user-friendly way.

The passport and decode tables in this Access database are imported from the Oracle server where they are created with SQL programs. Only accessions with an 'A-status' are online searchable. The online searchable evaluation data are also part of the Access database involving three tables per crop. These data are the result from manual combination of the raw observations into one score per accession per trait.

The downloadable passport and evaluation data are available as ZIPped Excel-files. These ZIP-files are created by a procedure using both SQL and Visual Basic programs. These programs not only produce the ZIP-files but also the HTML-files that link to the downloadable files on the CGN Internet site.

The procedures to generate all Internet-files are described in detail in a manual (Van Hintum & Menting, 2001).



## **Appendix VI.**

### **Major changes since previous version**

#### **Seed management**

In table GNS\_PASSPORT, the column U\_AMOUNT has been replaced with the columns U\_AMOUNT\_COLD and U\_AMOUNT\_FREE. From now on user bags can be stored and administrated under different storage regimes. A new column for the location of U\_AMOUNT\_FREE was added (U\_FREE\_LOC). Storage locations of U\_AMOUNT\_COLD are still administrated in column COLD\_LOC.

Table GNS\_DUPLIC\_RCV has been replaced by GNS\_BATCH\_STORAGE. Not only safety duplicates from other genebanks, but also boxes with archive material can be administrated in this table.

#### **User multiplication**

Column GNR has been replaced with the columns B\_GNR and U\_GNR in table GNS\_PASSPORT. With the introduction of user multiplication the generation numbers between the different types of seed bags within one accession can differ. Also a column for the year of regeneration of the user material has been added (LRU\_YEAR).

#### **Status of material**

Two new material statuses are introduced. The 'C' status for archived material and the 'E' status for external collections. A process to archive seed samples was added. Screens to access the archived and external material have been added.

#### **Reports**

The report section of Genis has been converted to an MS-Excel environment. The reports are generated with VBA code, which sends queries to the genis tables. The output of the reports is user friendly; adaptations and maintenance have become easier.

