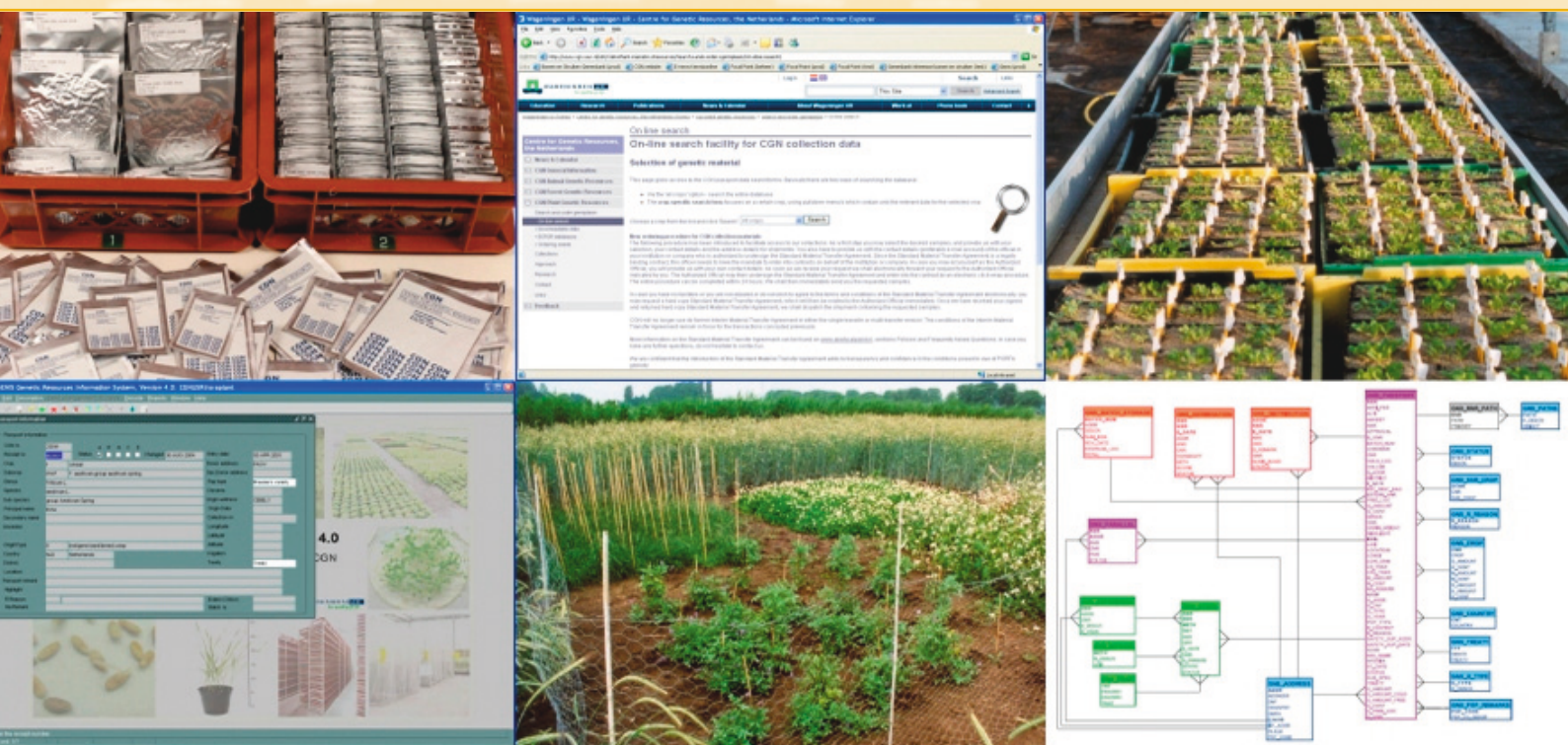


Genis Data Dictionary

F.B.J. Menting, M.G.P. van Veller & Th.J.L. van Hintum



CGN Report 2007/08

Genis Data Dictionary

F.B.J. Menting, M.G.P. van Veller & Th.J.L. van Hintum

Centre for Genetic Resources, the Netherlands (CGN)
Wageningen University and Research Centre
September 2007

Report 2007/08

© 2007 Wageningen, CGN/DLO Foundation

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of CGN/DLO Foundation.

Copies of this report can be ordered by sending an e-mail to the address below.

CGN, Centre for Genetic Resources, the Netherlands

Address : Bornsesteeg 65, 6708 PD Wageningen, The Netherlands
: P.O. Box 16, 6700 AA Wageningen, The Netherlands
Tel. : +31 317 47 70 45
Fax : +31 317 42 31 10
E-mail : cg@wur.nl
Internet : www.cgn.wur.nl

Table of contents

	page
1. Introduction	1
2. Structure	3
2.1 Information	3
2.1.1 User information	3
2.1.1.1 Descriptive information	4
2.1.1.2 Seed management information	4
2.1.2 Decode information	4
2.1.3 Process information	5
2.2 Users	5
2.2.1 Internal users	5
2.2.2 External users	6
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 Menus and modules	18
3.3 Reports section	20
4. Detailed description of some tables	21
4.1 The GNS_PASSPORT table	21
4.2 The GNS_PARALLEL table	28
4.3 The GNS_SCORE table	29
4.4 The GNS_PATHS and GNS_RNR_PATH tables	29
5. Database management	31
5.1 Content management	31
5.1.1 Descriptive information	31
5.1.2 Seed management information	32
5.2 Technical management	32
6. References	35
Appendix A Codes used in CHI	
Appendix B Symbols and abbreviations	
Appendix C Privileges on tables and views for Oracle user names	
Appendix D System for scoring heterogeneous populations	
Appendix E Genis on the internet	
Appendix F Relationship diagram	

1. Introduction

Since 1985 the Plant Genetic Resources (PGR) cluster of the Centre for Genetic Resources, the Netherlands (CGN) uses an information system for the documentation of its collections of several agricultural and horticultural crops. The information system is centrally managed and used by the staff of CGN and is referred to as Genis (GENetic resources Information System). Genis contains different kinds of data on the CGN collections that are important for the complete process of acquisition, *ex situ* conservation and distribution of material. Data important for this process and stored into Genis are:

- Passport data on the identity and collection history of accessions (i.e. the units of the genebank collection that are conserved, characterized and distributed to users).
- Characterization and evaluation data on traits that are of importance to users of the collections as well as to curators for the crops.
- Seed (storage) location and stock management data.

Information on these data can be accessed by genebank personnel from their offices and the seed storage facility. The latest version of Genis, version 4.0, is based on Oracle Database version 8.1.7.4 and was build using Oracle Designer 10g and Oracle Developer 10g. Besides a production database there are also a test database and a development database available on three database servers.

Genis version 4.0 is a web based application. This means that a client PC connects via internet to an internet application server which is connected to a database server via intranet. Thereby, the client PC only needs an internet browser. For the connection to internet there are two internet application servers (for testing and production respectively). All development, testing and maintenance takes place at the three different database servers and the two internet application servers via an application development PC which is connected to the servers. The general architecture of Genis is illustrated in Figure 1.

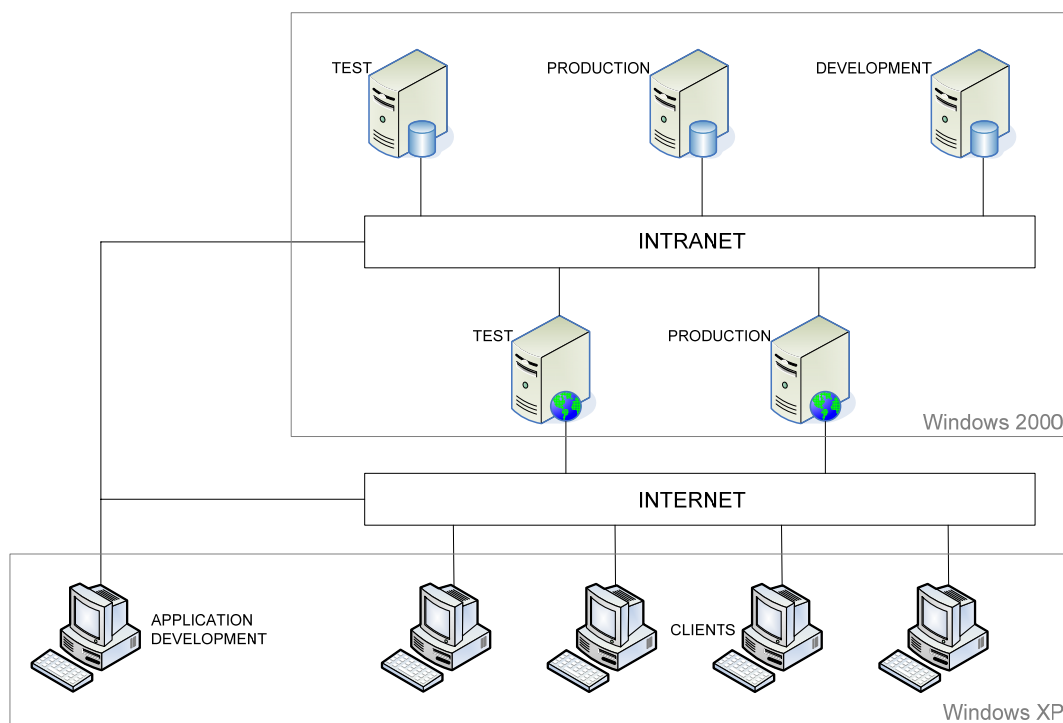


Figure 1. General hardware architecture of the GENetic resources Information System of CGN.

Genis is a relational database: the information is stored in two-dimensional tables. The tables are linked using key-fields. In 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. Genis consists of about 25 tables (including decode tables), some views, about 30 screens and a report section. Genis works on a Windows 2000 platform at the server site. On the client side PC's with Windows XP are used.

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 4.0, took place in 2007 and was a transformation from a client server application to a web based version.

In order to enable end users also to obtain information on the CGN collections, most of the information present in Genis has been made publicly available. This has been achieved by extending Genis with Genis-web. The latter database application contains most of the data of Genis and can be approached via the internet.

This report forms an update of the earlier published Genis Data Dictionary (Menting & Hintum, 2003). It gives an overview of all aspects of the information system important for understanding how Genis works. First, in chapter 2 the logical structure of Genis has been described: the different types of information in relation to the tables, the users with their privileges, and the processes. Next, in chapter 3 a complete list of all elements in Genis is given including the tables with their fields, the menus, the screens, the reports 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 in chapter 4. Chapter 5 deals with the management of contents as well as technical aspects of Genis. An overview of Genis-web, as an information system for end users of CGN collections, and the procedure for ordering material from CGN's collections via the implementation of a clickwrap procedure for the authorization of a Standard Material Transfer Agreement is given in Appendix E of this document.

2. Structure

In Genis three different 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 seeds (location, quantity, distribution, germinability, etc.) in the genebank collections.
- Decode information, giving an explanation of the codes used in the information system.
- Process information, temporary stored information necessary for processes like change of status for accessions, or generation of labels, etc.

Genis is internally used as an information system by the CGN staff (i.e. curators, seed managers and documentation staff). Four different roles can be distinguished in order to let internal users approach the information stored in Genis depending on how they use the system (see 2.2):

- End-users of the data, can only view most of the information.
- The data typists (usually curators and their assistants) can enter and update information on their crops.
- The seed manager can book off seed samples, assign accession numbers, enter and update germination data as well as information on seed storage locations.
- The database administrator owns the database and can do anything.

Several basic processes can be provided by Genis in order to support genebank management (see 2.3):

- Change of status of samples.
- Distribution of samples to users.
- Entering characterisation and evaluation data into the system.
- Generating labels for the seed bags.

These processes can be started from Genis by internal users depending on the role they have. Besides internal users, however, information on the collections of CGN, stored in Genis, is also of particular interest to (external) end users (e.g. plant breeders or researchers). In order to enable them to obtain information on the CGN collections as well, part of the information present in Genis has been made public available via Genis-web (see 2.2 and Appendix E).

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 seeds: 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 numbers that accessions of the CGN collections have in other genebank collections.
- Characterization and evaluation data, in the GNS_SCORE table, describing the phenotypes of the accessions. Moreover, these data include useful qualitative and quantitative information on traits of accessions for the user-community as well as for curators.

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 facilities, 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, by storing material in other genebanks, is documented per accession in the GNS_PASSPORT table. Safety duplication information received from other genebanks and information about batches of archived CGN-material and source material (see 2.3.1) is stored in the GNS_BATCH_STORAGE table. Only information on the number and location of received safety duplicates is 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). The following decode tables are part of Genis:

- 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 the GNS_SUB_CROP table.
- Traits scored during evaluation or characterization are stored in the GNS_TRAIT table.
- Descriptions of the methods that have been used are stored in the GNS_METHOD table.
- Information about the experiments that are applied to evaluate or characterize accessions can be found in the GNS_EXPERIMENT table.
- The country codes are explained in the decode-table GNS_COUNTRY.
- Origin types, population types, batch types, germination rate types, dormancy types, reasons for deciding not to include material in the CGN collections, the different statuses of the material and codes for information on access and benefit sharing of the seed samples are described in the GNS_REF_CODES table.
- Codes used in the field ADDI_PAS (of the table GNS_PASSPORT) are embedded in text, not in a separate column; they are decoded in the table GNS_PSP_REMARKS.

2.1.3 Process information

Process information is the information needed to perform processes such as the generation of labels for seed bags (see 2.3) or the dispatching of seed samples. This type of information is usually temporal and as soon as the process is finished the information often 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 the distribution of seeds.
- GNS_PROC_REJECT: information on seed samples that are not going to be included in the collections, 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 and evaluation data into the table 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 for seed samples.

2.2 Users

In general, two groups of users of the information on the collections of CGN can be distinguished:

- Internal users, i.e. personnel of the Plant Genetic Resources (PGR) cluster of the Centre for Genetic Resources, the Netherlands (see 2.2.1).
- External users, i.e. end users of seed material, mostly outside CGN (see 2.2.2).

2.2.1 Internal users

Only internal users apply Genis for approaching information on the collections of CGN. Because the number of these internal users (typically personnel of the PGR cluster of CGN) is limited, all new personnel of CGN receive a personal training by a member of the documentation staff of CGN, on how Genis works.

Curators use the information in Genis to monitor and improve the quality of their collections, thereby looking at the genetic diversity, overlap and gaps within and between accessions. Besides this application of Genis, the curators also use Genis to improve the documentation of their accessions and the planning of regenerations for accessions of their collections. The seed manager uses Genis for an effective seed management (information on seed location and stock management), as well as for the storage of germination data (necessary to monitor the seed quality of the accessions).

In order to enable the internal users to apply Genis for their specific demands, four different roles with different functionalities (and rights) have been defined in Genis.

- End-users of the documentation information system (CGNUSR): are able to view all data in the system and can generate some output-listings in a user-friendly way. They are not able to change or delete anything, not even by mistake.
- Data typists (CGNADM): have the same rights as end-users, but moreover are able to add and correct all accession specific and certain process and decode information.
- The seed manager (CGNSMR): has the same rights as data typists, but moreover is able to add and correct seed management data, and to place orders for labels.
- 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 C.

2.2.2 External users

The group of external users, mostly plant breeders, use the information provided to make selections of material to include in their breeding or research programs. Besides plant breeders also researchers and organizations that use material from the collection for educational as well demonstration purposes are interested in information on the collections of CGN. More information on the application area of information on the collections of CGN can be found in Roelofsen (1985), Hintum (1988) and on the CGN website (www.cgn.wur.nl).

The information on the collections of CGN is provided to the external users by extending Genis with Genis-web. The latter database application contains most of the data of Genis and can be approached via the website of CGN. The information in Genis-web is updated from Genis every two months. Details on the update procedures can be found in Appendix E and in the CGN Quality Manual.

Users can make searches in the database application (i.e. Genis-web) and can make selections of material which can be ordered online via a shopping cart manager. In the procedures that follow ordering the users have to indicate what intended use they have for the material they want to obtain from CGN's collections. A clickwrap procedure is provided for the ordering of material that falls under the Standard Material Transfer Agreement of the International Treaty for Plant Genetic Resources for Food and Agriculture. (ITPGRFA).

The steps needed for ordering of material from CGN's collections are presented in Appendix E.

2.3 Processes

As mentioned before (see 2.1.3), 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 distribution of material is administrated and the way characterization and evaluation data are entered into the system.

The various processes for the management of CGN's collections are described in more detail in the CGN Quality Manual.

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 tables to improve performance and facilitate obtaining information from Genis.

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

Some tests (e.g. germination tests) can be performed on received seed samples, to determine if they should be included in the collections. Also the passport data may be studied to determine if a sample would improve the composition of the collection. When it is decided to include a received sample in the collections, 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 include a received sample in the collections the sample gets the 'N' status ('Not Accessed'). Codes for the reasons for not including are given in the R_REASON column of the GNS_PASSPORT table. These codes are decoded in the GNS_REASON table. An additional remark about not including can be made in the field NA_REMARK in the GNS_PASSPORT table. 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 included seed sample (with status 'Not Accessed') 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 but 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 include 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 table 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 stored in the CGN genebank. An example is the apple collection that is maintained as trees (field genebank) outside the CGN storage facilities. These types of accessions will get an external accession number (EXTERN_ANR) instead of an accession number (ANR).

2.3.2 Requests

Users can request material with an 'A' or 'E' status:

- For 'A' status material, the system checks whether there is still enough seed for distribution after the requested accession numbers have been entered in the information system. If there is enough seed the requested sample is written off in the relevant table and registered as requested. The seed manager can generate 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.
- For 'E' status material a request is forwarded to the curator of the collection (which is stored outside the CGN genebank) for distribution of the material.

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 MS Excel (more information on this commercial software is available via the website: www.oracle.com). It will depend on ones privileges and experience, and the type and amount of data which option of data entering is preferred.

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

For entering characterization and 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 and 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 GNS_SCORE table.

2.3.4 Labels

A last process to be mentioned is the generation of labels for storing bags of seed samples 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 from several tables is combined.

3. Constituting elements

Besides an Oracle database management system, the information system of CGN (i.e. Genis) consists of many different elements. In this chapter the most important of these elements will be listed and described briefly:

- Tables and views with their fields (see 3.1).
- Menus and modules (see 3.2).
- Reports section (see 3.3).

3.1 Tables and views with their fields

All tables and views of Genis are listed below with their corresponding 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 each 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, R = Rowid; 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.

GENIS_REF_CODES

Table for referential codes used in Genis

Column	Type	Len	Key	Decoded in	Null	Description
RV_LOW_VALUE	V	240	P		N	Code of the referential value
RV_HIGH_VALUE	V	240			Y	Upper value of referential code in case of range of values
RV_ABBREVIATION	V	240			Y	Abbreviation of referential code
RV_DOMAIN	V	100	P		N	Domain of referential code
RV_MEANING	V	240			Y	Description of referential code
RV_TYPE	V	10			N	Type of referential code

GNS_ADDRESS

Decode table for address codes

Column	Type	Len	Key	Decoded in	Null	Description
ADDI_ADD	V	100			Y	Additional address information
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

GNS_BATCH_STORAGE		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 sent the safety duplicates (archived material = 'CGN')
BATCH_NUM	V	10	P		N Batch number
BATCH_TYPE	V	25	R	GNS_REF_CODES	N Type of batch
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 Approximate number of seed samples in the batch

GNS_COUNTRY		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
PREF	V	1	R	GNS_REF_CODES	Y 'Y' if the code is preferred, 'N' if not

GNS_CROP		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

GNS_DIST_CUR_VW		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

GNS_DIST_SM_VW		Distinct crop numbers for module prdist based on table GNS_DISTRIBUTION			
Column	Type	Len	Key	Decoded in	Null Description
CNR	N				N Code for the crop

GNS_DISTRIBUTION

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	R	GNS_REF_CODES	N	Status of the seed sample

GNS_EXPERIMENT

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

GNS_FBSCORE_VW

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

GNS_GERMINATION

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
GERM_TYPE	V	1	R	GNS_REF_CODES	Y	Type of germination test
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	R	GNS_REF_CODES	N	Status of the seed sample

GNS_GERMINATION_VW

Recent germination results per GNR and RNR on material with status 'Received' or 'Accessed', 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			N	Status of the seed sample

GNS_METHOD

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

GNS_MUTATION_LOG

Information on history of data

Column	Type	Len	Key	Decoded in	Null	Description
ACTION	V	1			N	Type of mutation, Insert, Update or Delete
CATEGORY	V	20			N	Field category of the mutation (Seed management or passport data)
COLUMN_NAME	V	30			N	Mutated column
DATA_TYPE	V	1			N	Data type of the mutated column
ID	N		P		N	Unique identifier
MUTATION_DATE	D				N	Date of the mutation
MUTATION_USER	V	20			N	User who performed the mutation
NEW_VALUE	V	240			Y	Value after the mutation
OLD_VALUE	V	240			Y	Value before the mutation
TABLE_NAME	V	30			N	Table where the mutation was performed
TABLE_ROWID	R				N	Rowid of the mutated row in the original table (TABLE_NAME)

GNS_PARALLEL

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	P		N	Number given by the institute (ADDR) to the seed sample
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number
STATUS	V	1	R	GNS_REF_CODES	N	Status of the seed sample

GNS_PASSPORT

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	R	GNS_REF_CODES	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	R	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
COLLNR	V	30			Y	Collecting 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
DORMANCY	N		R	GNS_REF_CODES	Y	Code for dormancy breaking method
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 -20°C storage
G_AMOUNT	N				Y	Number of germination-bags
G_CONT	N				Y	Number of seeds in the germination-bags
GENAUTHOR	V	30			Y	Authority for the genus name
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
LRB_YEAR	V	4			Y	Year of the last regeneration of base material
LRU_YEAR	V	4			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 status '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_DATE	V	8			Y	Date of first reference
O_TYPE	V	1	R	GNS_REF_CODES	Y	Code for the origin type
POP_TYPE	V	1	R	GNS_REF_CODES	Y	Code for the population type
R_CONTENT	N				Y	Content of rest-bag (in grams)
R_REASON	N		R	GNS_REF_CODES	Y	Code for reason not to include material in collection
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 sub crop
SEC_D_ADDR	V	6	R	GNS_ADDRESS	Y	Code for the secondary donor

SEC_NAME	V	80			Y	Secondary name of the accession
SPAUTHOR	V	30			Y	Authority for the species name
SPECIES	V	80			Y	Species name
ST_DATE	D				Y	Date that the status has been changed
STATUS	V	1	R	GNS_REF_CODES	N	Status of the seed sample
SUBAUTHOR	V	30			Y	Authority for the sub specific name
SUB_SPEC	V	80			Y	Taxonomic sub specific name
TREATY	V	10	R	GNS_REF_CODES	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_COLD_LOC	V	10			Y	Location of the distribution-bags in the cold storage room (+4 °C)
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
U_PACKAGE_DATE	D				Y	User-bag package date

GNS_PATHS 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 included in the collection (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	Y	Batch number
E_DATE	D				Y	Date of entry in the table
NA_REMARK	V	120			Y	Remark on archivation
R_REASON	N		R	GNS_REF_CODES	N	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 table GNS_PROC_SCORE (module: f_book2, 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				N	Number of bags to be distributed
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number
STATUS	V	1	R	GNS_REF_CODES	N	Status of the seed sample

GNS_PROC_LABELS Temporarily information for generation of labels (module: labels, menu item: process, label)

Column	Type	Len	Key	Decoded in	Null	Description
ANR	N				N	Accession number in the CGN collection
CNR	N		R	GNS_CROP	N	Code for the crop
G_LAB	N				N	Number of labels for germination-bags
M_LAB	N				N	Number of labels for regeneration-bags
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number
R_LAB	N				N	Number of labels for rest-bags
STATUS	V	1	R	GNS_REF_CODES	N	Status of the seed sample
U_LAB	N				N	Number of labels for distribution-bags

GNS_PROC_OBJ_SEQ Temporarily stored objects (=accessions) of an experiment for easier entrance of experiment data in table GNS_PROC_SCORE (module: f_book1, 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	R	GNS_REF_CODES	Y	Status of the seed sample

GNS_PROC_REJECT

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_REF_CODES	N	Reason code for not including material in collection
R_REMARK	V	120			Y	Remark on not including material in collection
RNR	V	6	PR	GNS_PASSPORT	N	Receipt number

GNS_PROC_SCORE

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	R	GNS_REF_CODES	N	Status of the seed sample
TRT	N		PR	GNS_TRAIT	N	Code for the trait

GNS_PSP_REMARKS

Decode information for codes used in field ADDI_PAS of 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

GNS_RECENT_GERMINATION_VW

Recent germination results per RNR on material with status 'Received' or 'Accessed'
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			N	Status of the seed sample

GNS_RNR_PATH		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

GNS_SCORE		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	R	GNS_REF_CODES	N Status of the seed sample
TRT	N		PR	GNS_TRAIT	N Code for the trait

GNS_SCORE_CROP_VW		Number of scores, experiments and methods per crop/trait-combination, based on table GNS_SCORE Includes samples with status 'Received' or 'Accessed'			
--------------------------	--	---	--	--	--

Column	Type	Len	Key	Decoded in	Null Description
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

GNS_SCORE_EXP_VW		Overview of data per experiment, with description of TRT and METH, based on tables GNS_SCORE, GNS_TRAIT and 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

GNS_SCORE_VW2 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

GNS_SUB_CROP 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 sub crop
SUB_CROP	V	80			N	Name of the sub crop

GNS_SYS_PARAMETERS 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

GNS_TRAIT Decode table for trait codes

Column	Type	Len	Key	Decoded in	Null	Description
TRAIT	V	80			N	Name of the trait
TRT	N		P		N	Code for the trait

3.2 Menus and modules

To gain access to the screens of Genis a drop down menu is used. Some menu items have restricted access, depending on the user role (see 2.2.1) that is used to approach Genis. In the application, the menu items to which a user role has no access are grey; the accessible menu items are bright black.

A toolbar is available, 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 roles. Sub menus 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	
Move	Cpyexp	Move all records from an experiment from GNS_PROC_SCORE to GNS_SCORE	s	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		
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		

Menu Item	Module name	Purpose	Type	Access			
				dba	smr	adm	usr
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
Sub crop	Scrpinfo	Query, enter and update sub -crops	s	y	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
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
Reference codes	Cgrefcodes	Query, enter and update reference codes	s	y	y	y	y
Reports	Make_reports	Access to all report menus	m	y	y	y	y
Excel Reports	Excel Reports	Starts a MS Excel reports workbook (see 3.3)	u	y	y	y	y

3.3 Reports section

Reports is a way to put together all facilities that belong to Genis in a user friendly way. Moreover, reports can function as a facilitation for the input of data by means of MS Excel templates.

Besides direct use of Genis via its menu and modules, also reports can be used to approach information that has been stored in Genis. Moreover, reports can function as a facilitation for the input of data by means of MS Excel templates.

Genis reports (version 1.1) have been generated with VBA (i.e. Visual Basic for Applications) code, which sends queries to the Genis tables using an ODBC (for Oracle) connection. The output is presented in a MS Excel workbook in a user friendly way which makes presentation of data easy. While browsing Genis reports the user might be prompted to enter additional information.

Via the reports section of Genis information can be obtained of almost all data that are stored in Genis. Not only overviews of passport, distribution and characterization and evaluation data but also all decode tables can be listed in MS Excel format. Further, via the reports section of Genis, templates are available that can be used for the request of decoding experiments, traits and methods. Examples of complex queries and links to relevant documents (e.g. descriptor lists) are also part of Genis reports.

It is also possible to integrate user applications in reports. Depending on (internal) user demands, the contents of Genis reports, and information that can be approached via it, may vary.

4. Detailed description of some tables

Apart from the formal rules considering data type, field length and format, etc., the tables of Genis 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 (see 4.1).
- The GNS_PARALLEL table (see 4.2).
- The GNS_SCORE table (see 4.3).
- The GNS_PATHS and GNS_RNR_PATH tables (see 4.4).

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.

4.1 The GNS_PASSPORT table

The GNS_PASSPORT table is used for storing passport information of the accessions in the CGN germplasm collections. 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 most fields allow long text strings, one should (if applicable) enter text as clear as possible. So, not: 'c. 3 km S. of Lutjebroek', but: 'about 3 kilometer South of Lutjebroek'.

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

The fields of table 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 crop collections.
- RNR (Receipt number)
An unique number given when a seed sample is nominated to become part of the collection (see 2.3.1). 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 the collections of CGN but not stored in CGN's storage facilities.
- BATCH_NUM (Batch number of archived material)
A number that gives information on 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 the table GNS_BATCH_STORAGE, the location of the box in which the material has been 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 under CGN's responsibility but stored outside the storage facilities of CGN.
- 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 crop codes.

- **SCNR (Sub crop number)**
The sub crop 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 sub crop 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.
e.g.: 'Triticum.'
- **GENAUTHOR (Authority name for genus name)**
The authority name for the genus of the accession.
e.g.: 'L.'
- **SPECIES (Species)**
The name of the species. This name should always be written in lower case.
e.g.: 'turgidum'.
- **SPAUTHOR (Authority name for species name)**
The authority name for the species of the accession.
e.g.: '(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.
- **SUBAUTHOR (Authority name for subspecies name)**
The authority name for the subspecies of the accession.
e.g.: '(L.) Thell.'
- **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 collecting 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. The 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_REF_CODES) 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 table GNS_COUNTRY. The database administrator 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 to 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: ', 'Department: ' or 'Estate: '.

e.g. 'Department: 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 table GNS_PSP_REMARKS. Until now the 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
CHI	Coded habitat information (see Appendix A)
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)
EFP	Former political entity (followed by the old country code and the year it was updated)
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
REM	Remark of any kind
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)
TSY	Taxonomic synonym
TYP	Type of variety or selection (groups within sub crops)
WAC	Wageningen Potato Collection number
WRF	Number for PI hybrid from NRSP-6

If there is more than one item in the Passport remark field, they are separated by a ';' .
e.g.: 'TYP:Agilo; CUL:under glass'

- 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.
- R_REASON (Rejection remark code)
Remark code in case received material (status 'R') is not entered in the genebank collections but is rejected (status 'N') or archived (status 'C'). The following codes are available (decoded in table GNS_REF_CODES).

Code	Description
1	Not enough viable material
2	Passport data don't match
3	Sample too heterogeneous
4	Already in the collection
5	Sample is a hybrid
6	Too difficult to maintain
7	Became part of a group
8	Sample is not free available
9	Too little information
10	CGN did not take responsibility
11	Responsibility transferred to other genebank

- **NA_REMARK** (Rejection remark)
Additional remark made when received material (status 'R') is not entered in the genebank collections but is rejected (status 'N') or archived (status 'C').
- **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. The database administrator, seed manager and data typists can create new codes.
- **SEC_D_ADDR** (Secondary donor address)
The address code of the secondary donor of the accession. The institute, genebank, university, company etc., that gave the accession to the donor of the CGN material. This field is decoded in table GNS_ADDRESS. The database administrator, seed manager and data typists can create new codes.
- **POP_TYPE** (Population type)
A code for the population type, decoded in table GNS_REF_CODES. The following population types are available:

Code	Description
B	Breeder's variety
H	Hybrid
L	Land variety
O	Weedy
R	Research material
W	Wild

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 morphotypes, 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 table GNS_ADDRESS. The database administrator, seed manager and data typists can create new codes.
- **O_DATE (Origin date)**
The date the accession was first documented. This can be the time it was introduced as a cultivar, or the moment it was collected. For Dutch cultivars the date that breeders right was granted or the date it appeared on the variety list is used. Format of this field: 'YYYYMMDD' with '-' for missing values.
e.g.: '1986—', '20010312'.
- **COLLNR (Collecting 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 are replaced by hyphens ('-').
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 are replaced by hyphens ('-').
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.
- **U_COLD_LOC (Storage location +4 °C room)**
Location of seed samples for distribution in the cold storage room (+4 °C).
U_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 facilities (-20 °C).
FREE_LOC has the same format as U_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 U_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 cold 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 storage facilities.
- **G_AMOUNT (Number of bags for germination tests)**
Number of bags with seeds for germination tests stored at -20 °C.

- **M_AMOUNT** (Number of bags for base regeneration)
Number of bags with seeds for base regeneration stored at -20 °C. When including a seed sample in the collection the number of bags reserved for regeneration is two.
- **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 per rest-bag (in grams))
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 number of seeds in the rest-bag. 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 grams.
- **DORMANCY** (a code for dormancy breaking method of the seeds)
The following codes are available (decoded in **GNS_REF_CODES**):

Code	Short description	Description
1	Cold treatment	10 days cold treatment at 5-10 °C
2	GA3 treatment	Soaking for 48 hours in 1000 ppm GA3
3	Cut seed top	Cut the seed top after the seeds are swollen
4	Remove seed coat	Removal of the fruitwall and seed coat after the seeds are swollen
5	Prolong period	Prolong the germination period
6	Scarification	Mechanical scarification immediately above the tips of the cotyledons in case of hardseededness

The short description will appear on the seed list that will be sent with seed samples. It is explained by the (complete) description in the 'decode part' of the seed list.

- **ST_DATE** (Status changed date)
Date that the status of the seed sample has changed. From 'R' to 'A' in case of new included material. From 'R' or 'A' to 'N' in case of not included material. From 'R' or 'A' to 'C' in case of archive (see also: **STATUS**).
- **U_PACKAGE_DATE** (user-bag package date)
Date that the user-bags are filled.
- **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**.

- LRB_YEAR (Last base regeneration year)
Last regeneration year of base material.
- LRU_YEAR (Last user regeneration year)
Last regeneration year of user material.
- SAFETY_DUP_ADDR (Safety duplicate)
Address where the safety duplicate of the accession is kept.
- SAFETY_DUP_DATE (Safety duplicate 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 genebank 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 when an accession is found to be a duplicate of a CGN accession.

One has to realise that accessions with parallel numbers are probably not genetically identical to the CGN material. Although 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 the genebanks.

The table GNS_PARALLEL consists of the following fields: ADDR, ANR, CNR, PNR, RNR and STATUS.

All fields of GNS_PARALLEL need to have a value except the field 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). The receipt number has six 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 from the GNS_PASSPORT table, but included for user friendliness. The current crop codes are decoded in table 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 ADDR 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 the table GNS_SCORE, except ANR, GNR and S_REMARK, need 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 including it in the collections of CGN (see 2.3.1). The receipt number has six 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.
- **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 the table GNS_EXPERIMENT, which gives a description of the experiment. The database administrator 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 table 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 table 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. Variation in the score within the accession may be indicated using a standardized system (see Appendix D).
- **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 this collection 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. The table GNS_RNR_PATH stores for each RNR in the GNS_PASSPORT table 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 the table GNS_PATHS are:

- **PATH (Path-string)**
A 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 also be a path-string.
- **P_DESC (Path description)**
A short description of the path, relative to it's parent. For example, if '06' is the crop 'lettuce' then the description of '061' as 'cultivated' is sufficient, it is not necessary to define the path as '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 the table GNS_RNR_PATH are:

- **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.
- **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 including it in the collections of CGN (see 2.3.1).

5. Database management

Apart from the technical aspects of Genis there are also the management aspects in terms of the management of the content and the management of the technical system as such. All relevant procedures for these aspects are part of the CGN Quality Manual and are described shortly below, the content management in paragraph 5.1, and the technical management in paragraph 5.2.

5.1 Content management

The content of Genis should be as complete and reliable as possible.

The quantity as well as the quality of passport data is largely determined in the past. However, sometimes for passport data the quantity and quality can be upgraded to a higher level by efforts of curators or other crop specialists when performing an update on (part of) the collections for which information is included in the documentation system. A discussion on the improvement of passport data for the fruit vegetable collection of CGN is given in Dooijeweert & Menting (2007).

Germination data and characterization data are of a different nature, information is added continuously and the method with which they were determined is recorded, allowing a better assessment of the reliability.

For data on the amount of seeds as well as data on storage locations of seeds, the reliability is dependent on the accuracy of employees during the processes of counting and storing of seeds. These and all other genebank management activities that are important for the conservation of the material in the collections of CGN have been described in the CGN Quality Manual to ensure the quality.

The responsibility for the quality and quantity of the data in Genis differs per datatype: the different kinds of data can be entered or updated by different Genis users, as defined in their roles and corresponding privileges. Four different roles are distinguished in Genis, as described in paragraph 2.2 of this report and in the CGN Quality Manual.

Depending on their role (i.e. end-user, data typist, seed manager or database administrator) users have different rights for making data modifications in Genis. Modifications on the data in Genis can be made by data typists, the seed manager or the database administrator. In general, codes can only be added or modified by the database administrator. For address codes, however, also users with the roles of data typist or seed manager can make additions or modifications. Suggestions for new codes must be made to the documentation project leader. For experiment codes, trait codes and method codes special request forms are available as MS Excel templates. A more detailed description of the procedures for adding and modifying data in Genis will be given below for descriptive information (see 5.1.1) and seed management information (see 5.1.2).

5.1.1 Descriptive information

Passport data can be collected under the responsibility of a curator on a MS Excel template which is available via the reports section of Genis. This spreadsheet template is flexible with respect to the format and coding of the information that is collected. However, before the information in the MS Excel file can be entered in Genis, strict requirements with respect to the format and coding need to be followed. These requirements with respect to format and coding are described in chapters 3 and 4 of this document.

Once obtained, documentation staff checks the format, coding and plausibility of the data. The improvements deemed necessary are discussed with the data supplier prior to entering the data in Genis. Only when the data

follow the requirements with respect to format, coding and plausibility they can be accepted and entered by the documentation staff.

Occasionally, passport data are also entered in Genis via its screens. In these cases, the format, coding and plausibility of the data is verified at entering the data in the database.

Parallel numbers (i.e. numbers for the same accession in another genebank collection) are supplied for entering in Genis via the same MS Excel template that is used for passport data. Normally, the documentation staff enters the parallel numbers in Genis but occasionally also other users can add or modify parallel numbers.

Characterization and evaluation data (and information on (new) methods and traits) are collected under the responsibility of a curator. After consultation with a documentation staff member data are sometimes corrected, completed or, when necessary, reduced to a single score per accession and per trait and method combination. For the coding of traits, methods and experiments request forms are available as MS Excel templates. When the data follow the requirements with respect to format, coding and plausibility, they are entered into Genis by the documentation staff.

5.1.2 Seed management information

Germination data are collected under the responsibility of the seed manager in a MS Excel file. Documentation staff performs checks on the format of these data and enters them in Genis.

In order to distribute material from CGN's collections, curators, the seed manager or documentation staff have to enter information on the accessions, that are going to be distributed, into Genis. Information on the distribution of material can also be entered automatically into Genis by the seed manager. Hereby, a special MS Excel template needs to be used. As soon as material is booked off, information on the transaction (material, user, data, etc.) and the amount of seeds will be saved automatically. This information can only be modified by the database administrator.

When material has obtained the status 'Accessed', information on the amount of seed bags and the location of the stored material is entered into Genis via the screens. Storage data can be modified via the Genis screens by the seed manager or the database administrator.

5.2 Technical management

The technical management (including backup procedures and implementations of new functionalities in Genis) is largely outsourced and laid down in separate (annual) agreements between CGN and external partners that are responsible for particular aspects of the technical maintenance. Hereby, the partners are responsible for the development, management and application of their own procedures for the services they provide to CGN. Their performance is evaluated during annual management assessments or internal audits that make part of the quality management of CGN, and are documented in the CGN Quality Manual.

Backups of Genis are included in the backup procedure for all Oracle databases of the host institute of CGN. In this backup procedure, every day an image of the Oracle servers is produced for fast recovery in case of a system breakdown. These daily images are overwritten every time a new image has been produced. For backups on the long term, every day a total export of the Oracle databases is saved. All these backups are saved for 20 working days. One of the backups is selected and saved for 12 months. Of these selected backups yet another backup is saved for another year (i.e. the backup is saved for two years). In this way, with the backups it is possible to go back 20 working days (with periods of a day), 12 months (with periods of a month) and 2 years (with periods of a year). Once every 6 months a backup of Genis is put through a recovery test. Procedures for the management of backups for Genis are described in detail in the CGN Quality Manual.

The implementation of new functionalities in Genis is done by external partners on the basis of feedback, problems and comments on Genis from the Genis users that have been collected and documented centrally in a feedback document by CGN documentation staff. Once every half year this document is evaluated by the documentation staff. During this evaluation the points of feedback are prioritised in three categories:

- Feedback that is going to be implemented. With the items in this category another document is generated that contains all information needed by the (external) Oracle programmers to implement the new features in Genis.
- Postponed feedback that either is considered not to be very important (at the moment) or that has radical consequences meaning that a major revision of Genis will be needed for the implementation.
- Rejected feedback that is neither implemented nor postponed.

For the latter two categories (Postponed and Rejected feedback) the reason for postponement or rejection is entered in the feedback document. After evaluation of the points of feedback the results are corresponded with the users that provided the feedback.

As noted, sometimes the feedback of users refers to restrictions of Genis that require more drastic changes in the approaches. The implementations to accommodate such changes often have a direct consequence for the genebank management procedures and might require major revisions of Genis (i.e. typically the above described category of Postponed feedback). An example is the separation of the concepts of 'seed sample' and 'accession', where the first is a physical representation of the second. Also the introduction of bar coding would require such drastic changes in the structure of Genis and the genebank protocols. Other, smaller examples of solutions to restrictions that need a more sophisticated approach are:

- Dealing with instability of political entities in a genebank information system. A discussion on this topic is presented in Veller & Menting (2005).
- Updating of generation numbers in tables GNS_GERMINATION, GNS_DISTRIBUTION and GNS_SCORE of Genis. A strategy for this updating is described in Menting & Veller (2005).

New implementations on the database and screens of Genis are tested by the documentation staff on the test database server on the basis of documents made available to the documentation staff by the external Oracle programmers after they have implemented new functionalities or features in Genis. Results of the tests are corresponded with the Oracle programmers and test results are archived on paper. If the test results are positive, the implementations are carried through on the production database server.

Besides several restrictions on input on the level of screens, Genis has a lot of internal integrity checks on the database level; for instance to avoid the input of non existing codes (referential keys) or to avoid the duplication of key fields (primary keys). Triggers check if the right conditions are met before firing; e.g. it is not possible to book off of seed samples when there is no seed left.

Every two months a verification test is performed on the data included in the tables of Genis. The verification program checks the format of data in fields, such as seed location data and remarks codes in the remarks field of the GNS_PASSPORT table. In cases where it is not coded in database triggers, also the integrity between tables is checked. Due to performance reasons and for easier generation of user-friendly output, some fields are duplicated between tables (denormalization). For example the fields ANR, CNR, STATUS are duplicated in several tables. The results of the verification tests are submitted to the relevant persons to check and update the incorrect data.

New implementations in Genis reports are made by the documentation staff on request of the CGN personnel. Implementations are tested on a copy of the application by the documentation staff. If the results are positive the reports application is updated with the new version. The older version of the Genis reports application is archived.

6. References

- CGN Quality Manual.
Centre for Genetic Resources, the Netherlands. Wageningen, the Netherlands.
- Dooijeweert, W. van & Menting, F. (2007).
Improving the quality of passport data of a genebank collection: approaches at CGN. Plant Genetic Resources Newsletter, accepted.
- European Commission (1997).
Commission Directive 97/46/EC. Official Journal of the European Communities No L 204: 43-46.
- Hintum, Th.J.L. van (1988).
GENIS: a fourth generation information system for the database management of genebanks. Plant Genetic Resources Newsletter 75/76:13-15.
- Hintum, Th.J.L. van (1993).
A computer compatible system for scoring heterogeneous populations. Genetic Resources and Crop Evaluation 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 & Th. Hazekamp (1992).
Genis Data Dictionary, July 1992. Centre for Plant Breeding and Reproduction Research (CPRO-DLO), Centre for Genetic Resources, the Netherlands. Wageningen, the Netherlands.
- IPGRI (1997).
Multicrop passport descriptors (final version). In: E. Lipman, M.W.M. Jongen, Hintum, Th.J.L. van, Gass, T. & Maggioni, L. (comp.). Central Crop Databases: Tools for Plant Genetic Resources Management. International Plant Genetic Resources Institute. Rome, Italy / CGN. Wageningen, the Netherlands.
- Menting, F. & Hintum, Th.J.L. van (2003).
GENIS Data Dictionary, Note 254. Centre for Genetic Resources, the Netherlands. Wageningen, the Netherlands.
- Menting, F. & Veller, M.G.P. van (2005).
Strategy for updating generation numbers in Genis. Centre for Genetic Resources, the Netherlands. Wageningen, the Netherlands.
- Roelofsen, H. (1985).
The information system of the Centre for Genetic Resources the Netherlands. CGN-Report no. 2. Centre for Genetic Resources, the Netherlands. Wageningen, the Netherlands.
- Purdy, L.H, Loegering, W.Q., Konzak, F.F., Peterson, C.J. & Allen, R.E. (1968).
A proposed standard method for illustrating pedigrees of small grain varieties. Crop Science 8: 405-406.
- Veller, M.G.P. van & Menting, F. (2005).
Dealing with instability of political entities in a genebank information system. Centre for Genetic Resources, the Netherlands. Wageningen, the Netherlands.

Appendix A Codes used in CHI

Format CHI (table 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: CHI5420-1—

Appendix B 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 can be found in the P_REMARK field

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

In general, when using abbreviations, like the ones listed below, one should be sure that they are clear in the context of their use.

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.	kilometer
c.	circa, approximately

Appendix C Privileges on tables and views for Oracle user names

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

Table / View	CGNDBA	CGNSMR	CGNADM	CGNUSR
GENIS_REF_CODES	SIUD	S	S	S
GNS_ADDRESS	SIUD	SIU	SIU	S
GNS_BATCH_STORAGE	SIUD	SIUD	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_EXPERIMENT	SIUD	S	S	S
GNS_FBSCORE_VW	S		S	
GNS_GERMINATION	SIUD	SIU	SIU	S
GNS_GERMINATION_VW	S	S	S	S
GNS_METHOD	SIUD	S	S	S
GNS_PARALLEL	SIUD	SIU	SIU	S
GNS_PASSPORT	SIUD	SIU	SIU	S
GNS_PATHS	SIUD		S	
GNS_PROC_ACCESS	SIUD	SI		
GNS_PROC_ARCHIVE	SIUD	SIU	SIU	
GNS_PROC_COLS	SIUD	SIUD	SIUD	
GNS_PROC_DISTRIBUTION	SIUD	SIUD	SIUD	
GNS_PROC_LABELS	SIUD	SIUD		
GNS_PROC_OBJ_SEQ	SIUD	SIUD	SIUD	
GNS_PROC_REJECT	SIUD	SIU	SIU	
GNS_PROC_SCORE	SIUD	SIUD	SIUD	
GNS_PSP_REMARKS	SIUD	S	S	S
GNS_RECENT_GERMINATION_VW	S	S	S	S
GNS_RNR_PATH	SIUD		S	
GNS_SCORE	SIUD	SIU	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_SYS_PARAMETERS	SIUD			
GNS_SUB_CROP	SIUD	S	S	S
GNS_TRAIT	SIUD	S	S	S

Appendix D 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 spent 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.

Information for decoding the scores back to frequencies is available in Hintum (1993).

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'.

- A homogeneous population with only long plants gets the score '7='.
- 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.

Other symbols

- 'v' is used to express that the population is variable: '3v', score = mainly 3 but variable.
- '-' is used if a fraction of the population has not been scored: '-3', half (or some more) of the plants not scored, other half has score '3'.

Appendix E Genis on the internet

Introduction

The CGN collections are used by several companies, institutes, universities and organizations for breeding activities, research activities, educational purposes and demonstrations. To enable these end users to obtain information on the CGN collections, almost all of the information present in Genis has been made public available. This has been achieved by extending Genis with Genis-web. The latter database application contains part of the data of Genis that is public available and can be approached via the website of CGN. Also, on this website general information about CGN and its staff, protocols and activities is provided.

Structure

To enable its functionality, Genis-web consists of the following parts:

- A MS Access database which is imported into a MS SQL Server database for approaching data via the CGN website. The MS Access database contains:
 - passport data (mirrored from Genis), and
 - characterization and evaluation data (processed from Genis).
- Downloadable zipped MS Excel files with passport and characterization and evaluation data, available on a server that is linked via the CGN website.
- A series of .NET modules written in C# that makes it possible to communicate with the MS SQL Server database via the Content Management System in which the website of CGN has been implemented.
- A web service, written in C#, to list pictures that are available for a given accession.
- Programs written in SQL plus and Visual Basic to generate the MS Access database files as well as the downloadable MS Excel files.

Online searchable information

The user-relevant passport information in Genis is online searchable. By using a shopping cart manager, the user can order a selection of accessions that follow the user's specific demands. 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 characterization and evaluation data are online searchable. To make these data searchable for users they often have to be rearranged in a well-interpretable form, providing one score per accession and trait.

The online searchable information as well as the shopping cart manager for placing orders from the collections of CGN is implemented in a series of .NET modules that are loaded by the Content Management System in which the website of CGN has been implemented. New functionalities in these .NET modules are implemented by an (externally hired) web programmer in a developing environment. After testing of the new functionalities, the developing environment is copied to the server from which the application is loaded by the Content Management System.

Downloadable files

Complete sets of passport data are downloadable per crop. These downloadable passport data follow the standards of 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.

The downloadable files with passport or raw evaluation data are placed on a server (outside the Content Management System) from which they can be downloaded via links on the website of CGN.

Updates of content

The online searchable information in Genis-web is updated from Genis every two months. For this update, the passport and decode tables in the MS Access database are imported from the Oracle server where they are created with SQL programs. Only information for accessions with an 'A-status' is imported and therefore online searchable via Genis-web.

The online searchable evaluation data are also part of the MS 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 MS Excel-files. The latter files are created every two months 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 website.

Further information on the update of content for Genis-web is included in a protocol which is part of the CGN Quality Manual. Also in this manual protocols are included for the update of information to international external databases (e.g. EURISCO: eurisco.ecpgr.org) and webservice (e.g. GBIF: www.gbif.org).

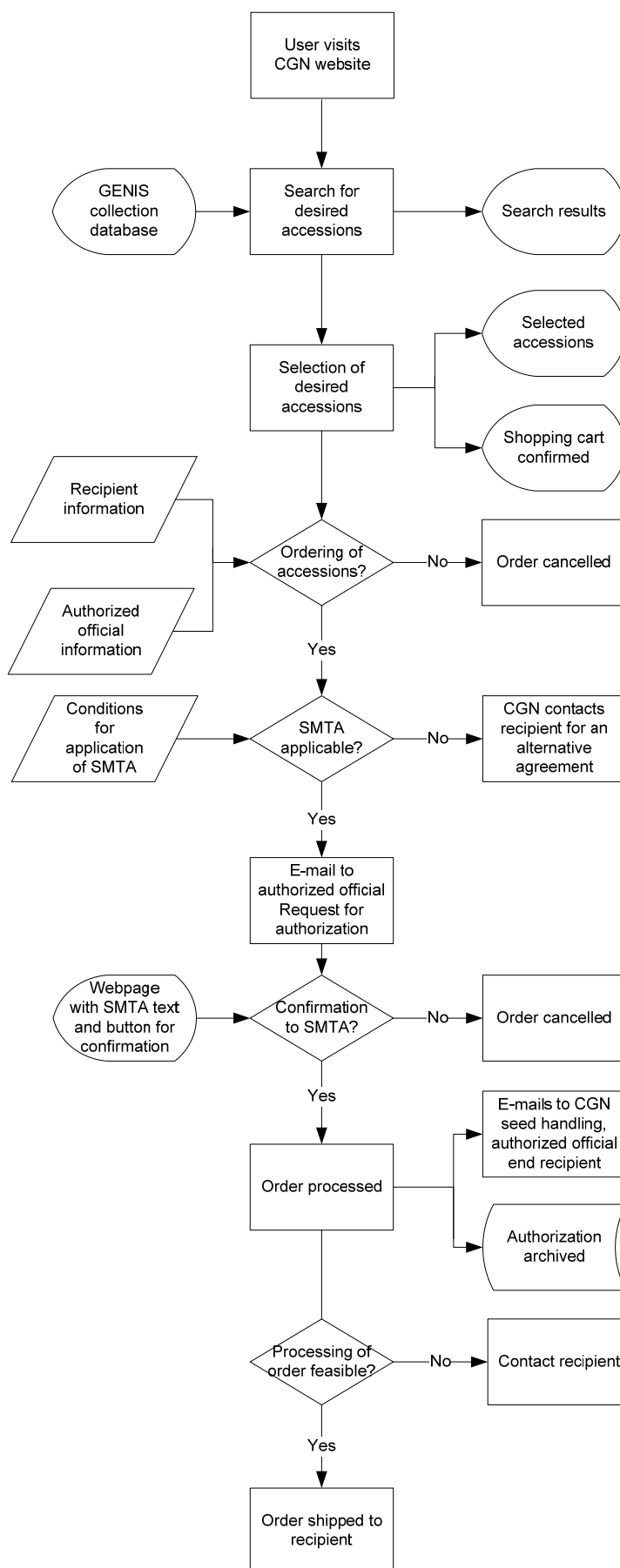
Online ordering of accessions

Online ordering of material from CGN's collections is possible via the CGN website (www.cgn.wur.nl). The user can search for desired material in Genis-web with a series of .NET modules that have been developed for communication with the MS SQL Server database via the Content Management System in which the website of CGN has been implemented. The procedure for the ordering of accessions via Genis-web is represented in the scheme below.

After making a selection of desired accessions, an order can be placed by the user with the shopping cart manager using checkboxes. To continue ordering of selected material, the user has to enter both recipient and authorized official information. This information will be used for the authorization and processing of the order.

If the Standard Material Transfer Agreement (SMTA) does not apply to the selected accession a CGN curator will contact the requestor to discuss an agreement for the transfer of the material. However, if the SMTA does apply, as in most of the cases (see the CGN Quality Manual for details), the order of the accession is finalized by confirmation to the Standard Material Transfer Agreement by the authorized official. CGN has implemented this confirmation to the SMTA in a clickwrap procedure.

The authorized official receives an e-mail with a link to a dynamic webpage with the text of the SMTA and a listing of the ordered accessions. By clicking on a button at the end of this webpage, the authorized official confirms to the SMTA. After confirmation, the authorized official, recipient and CGN seed handling are informed via e-mails on the confirmation of the authorization. The confirmation to the SMTA by the authorized official is archived. Further, it is checked if the processing of the order is feasible and in case it is, the order is shipped to the recipient. In case the order is not feasible yet, the recipient is contacted.



Appendix F Relationship diagram

