WP 4: Database on commercialized GM foods
Task leader: Partner 1; other Partners: 7, 9, 14 and 15 (data input from all partners)

1. Objectives and expected output

The aim of this work package was to set up a database on commercialized genetically modified foodstuffs world wide, e.g. approval information, modified DNA sequences, detection methods, including primer and probe sequences and literature about these foodstuffs. The database was tested by all partners of the project and was made publicly available at project's end. The database is in particular useful to officials in food control laboratories and authorizing organizations.

2. Description of work and results

The first prototype version (1.0) of the database was distributed on a 3 ½ inch floppy disc to all partners of the project during the second meeting in Gent 1998. The development history and a manual that accompanies the database has been published in detail by partners 14 and 15 as a proceedings report [Lindenmeyer and Hemmer, 1999]. Detailed information on the entire project and in particular the WP 4 database project is available on a public DMIF-GEN website under: http://food.ethz.ch/dmif-gen/. The website contains an index of all DMIF-GEN work packages, in particular of work package 4 (database development). It also contains a link to the latest database releases, descriptions of forms, addresses of group members and further information to support database users, such as a list of known bugs and problems.

The concluding report of the database group is reflected here not shortened because of the particular meaning of the database for the future dissemination of results.

Database on commercialized genetically modified foods[*]

Jakob Lindenmeyer[†], Wolfram Hemmer[‡], Lillian Auberson[†], Martin Schrott[§], Andreas Wurz[**], Matthias Foth[††], Hermann Rüggeberg[††], Peter Singer[‡‡], Hermann Broll[§§]

1. Abstract

The goal of work package 4 (WP4) of the DMIF-GEN project was the development of a database on detection methods for genetically modified foods. The data themselves is mainly based on publicly available information compiled in a pre-existing database [Hemmer]. The structure of this database was developed in an iterative approach of 8 steps. The developed conceptual schema is independent of a specific database management system (DBMS) and globally normalized (no redundancy). Two logic structures were extracted thereof: an Internet version based on Inprise InterBase and Java (by the German Federal Institute for Health Protection of Consumers and Veterinary Medicine) and a local version based on the latest version of MS Access (by the Agency BATS and the Swiss Federal Office of Public Health). Using these two approaches as a basis, this document shows the detailed process of the structure and schema development of the database as well as the characterization of data types and domains.

[*] This is an abridged version for the final report of the DMIF-GEN project. The full version is available under: http://food.ethz.ch/dmif-gen/final/
[†] Agency for Biosafety Research and Assessment of Technology Impacts of the Priority Program Biotechnology of the Swiss National Science Foundation (BATS), Clarastr. 13, CH-4058 Basel, Switzerland. www.bats.ch, E-mail: jakob@lindenmeyer.ch
[‡] GATC GmbH, R & D, Fritz Arnold Str. 23, D-78467 Konstanz, Germany. www.gatc.de, E-mail: w.hemmer@gatc.de
[§] Swiss Federal Office of Public Health (SFOPH), Division of Food Science, CH-3003 Bern, Switzerland. www.admin.ch/bag/, E-mail: martin.schrott@bag.admin.ch
[**] Gene-Scan GmbH, D-79111 Freiburg, Germany. www.genescan.com, E-mail: wurz@genescan.com
[††] Hanse Analytik, Bremen, Germany. www.hanse-analytik.de, E-mail: rueggeberg@hanse-analytik.de
[‡‡] Joker Concept GmbH, Berlin, Germany. E-mail: singer@joker-concept.b.uunet.de
[§§] German Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV), D-14195 Berlin, Germany. www.bgvv.de, E-mail: h.broll@bgvv.de

The information for every data set includes the country in which the GMO has been approved, contact persons, literature about the GMO, the genetic construct and the DNA sequence as far as available and detection methods to identify the genetic alteration in foodstuffs. By choosing the query forms in the independent windows it will be possible to search for specific items like organisms, genetic elements, etc. The online version will be accessible from everywhere, independent which hard- and software is used. Moreover different access levels could be set up to control data distribution.

2. Introduction

Within the last 7 years US authorities have approved 50 genetically modified (GM) plants. Half of the approvals concerned corn (13) and tomato (11), followed by soybean and cotton (each 5), potato and rapeseed (each 4). Two approvals concerned beet and squash and only a single approval has been given for chicory, flax, papaya and rice, respectively (table 1). For almost all of these products, the US authorities were the first to approve them for food use. Whereas none of the products approved so far needs to be labeled as GMO product in the US, regulations of other countries require labeling. Switzerland in contrast, requires labeling of GMO products on the basis of detectability using DNA-based detection methods if the product contains more than 1 per cent GMO. The limit of one per cent is based on experiences in the detection of GMOs and was the result of a consensus between the industry, farmers, authorities and consumers [Schrott]. Within the EU, labeling is also required for certain products as regulated by the Novel Food Directive [Regulation (EC) No 258/97] and related directives. A detection limit of probably one per cent is likely to be introduced in the near future. Under these circumstances, the increasing globalization of trade represents an extensive challenge to our food control authorities with respect to the control of compliance with national requirements (labeling and compulsory premarket approval).

The goal of work package 4 of the DMIF-GEN project SMT4-CT96-2072 was to set up a database, which would comprise important information for the development of detection methods for commercialized genetically modified foods. This includes information on the genetic elements introduced, modified DNA sequences, detection methods (including DNA-extraction procedures, primer and probe sequences) and literature about the foodstuffs. During the first phase of the project a concept had been established defining the format of the data, sources of data, filters and controls for data entry, concepts about maintenance, access and distribution (according to the technical annex of the DMIF-GEN project in [Schreiber]). The following chapters focus on the data format, the types and domains, as well as the structure of the database.

3. A systematic development approach in 8 steps

The major goal of a database is the correct, safe, efficient and permanent management of the data. The data themselves stay much longer in use, than all the application programs that use them. Therefore it is more important for a good database to have a long-term structure for the data, than to have a good-looking implementation in a specific database. This long-term structure is called a conceptual schema. With the upcoming of new DBMS it can be taken as a long-term basis to develop new DBMS-specific logical schemata (figure 1).
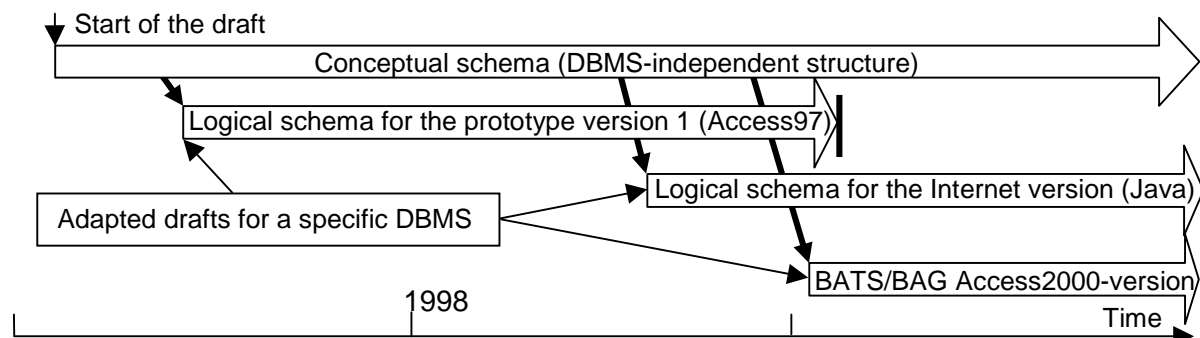

Figure 1: The conceptual schema is the basis for a long-term data structure. Thereof different logical schemata can be extracted. (Adapted according to [Zehnder]).

Table 1: Genetically modified crops approved for food use in the USA. Sorted by crop and date of approval (Date according to ISO 8601:1988: "Representation of dates and times". Form: CCYY-MM-DD/hh:mm:ss with leading zeros. In case of reduced precision, parts are omitted starting from the right-hand side [ISO]). Status: September 20th, 1999. Two examples (Maximizer® Maize Bt176 and Flavr Savr™ Tomato) are further discussed in figures 5, 9 and 12 and table 5.

Crop	Company	Altered Trait	Approval date
Beet	AgrEvo	Herbicide tolerant (Phosphinothricin)	1998-04-28
	Novartis Seeds	Herbicide tolerant (Glyphosate)	1998-12-23
Chicory	Bejo	Male sterile	1997-11-07
Corn	Ciba Seeds (Novartis)	Insect resistant (Lepidopteran) **Example Maximizer® Maize Bt 176*****	1995-05-17
	AgrEvo	Herbicide tolerant (Phosphinothricin)	1995-06-22
	Monsanto	Insect resistant (Lepidopteran)	1995-08-22
	DeKalb	Herbicide tolerant (Phosphinothricin)	1995-12-19
	Northrup King	Insect resistant (European Corn Borer)	1996-01-18
	Plant Genetic Systems	Male sterile	1996-02-22
	Monsanto	Insect resistant (European Corn Borer)	1996-03-15
	DeKalb	Insect resistant (European Corn Borer)	1997-03-28
	Monsanto	Insect resistant (European Corn Borer) and Herbicide tolerant (Glyphosate)	1997-05-27
	Monsanto	Herbicide tolerant (Glyphosate)	1997-11-18
	AgrEvo	Insect resistant (Lepidopteran) and Herbicide tolerant (Phosphinothricin)	1998-05-08
	Pioneer	Herbicide tolerant (Phosphinothricin) and Male sterile	1998-05-14
	AgrEvo	Herbicide tolerant (Phosphinothricin) and Male sterile	1999-04-22
Cotton	Calgene	Herbicide tolerant (Bromoxynil)	1994-02-15
	Monsanto	Insect resistant (Lepidopteran)	1995-06-22
	Monsanto	Herbicide tolerant (Glyphosate)	1995-07-11
	DuPont	Herbicide tolerant (Sulfonylurea)	1996-01-25
	Calgene	Herbicide tolerant (Bromoxynil) and Insect resistant (Lepidopteran)	1997-04-30
Flax	U of Saskatchewan	Tolerant to soil residues of sulfonylurea	1999-05-19
Papaya	Cornell Univ.	Virus resistant (PRSV)	1996-07-31
Potato	Monsanto	Insect resistant (Coleopteran)	1995-03-02
	Monsanto	Insect resistant (Colorado potato beetle)	1996-05-03
	Monsanto	Insect resistant (Colorado potato beetle) and Virus resistant (PLRV)	1998-01-29
	Monsanto	Insect resistant (Colorado potato beetle) and Virus resistant (PVY)	1999-02-25
Rapeseed	Calgene	Oil profile altered	1994-10-31
	AgrEvo	Herbicide tolerant (Phosphinothricin)	1998-01-29
	Monsanto	Herbicide tolerant (Glyphosate)	1999-01-27
	AgrEvo	Herbicide tolerant (Phosphinothricin) and Male sterile	1999-03-22
Rice	AgrEvo	Herbicide tolerant (Phosphinothricin)	1999-04-15
Soybean	Monsanto	Herbicide tolerant (Glyphosate)	1994-05-19
	AgrEvo	Herbicide tolerant (Phosphinothricin)	1996-07-31, 1998-04-30 and 1998-10-14
	DuPont	Oil profile altered	1997-05-07
Squash	Upjohn)	Virus resistant (WMV2 and ZYMV)	1994-12-07
	Asgrow	Virus resistant (CMV, WMV2 and ZYMV)	1996-06-14
Tomato	Calgene	Fruit ripening altered **Example Flavr Savr™ Tomato†††**	1992-10-19, 1994-10-03, 1994-11-18, 1995-03-23, 1995-07-28, 1996-10-09
	DNA Plant Techn.	Fruit ripening altered	1995-01-17
	Zeneca & Petoseed	Fruit Polygalacturonase level	1995-06-06
	Monsanto	Fruit ripening altered	1995-09-27
	Agritope	Fruit ripening altered	1996-03-27
	Monsanto	Insect resistant (Lepidopteran)	1998-03-26

Data source: The Federal Register, product approvals by USDA/APHIS (US regulatory agency), current status of petitions under: http://www.aphis.usda.gov/biotech/petday.html

---

*** Example of Maximizer® Maize Bt176 from Novartis: data set example in figure 9.
††† Example Flavr Savr™ Tomato (MacGregor's) from Calgene: see table 5 and figures 5 and 12.

The development of the DMIF-GEN database followed the 8 iterative steps of database development according to [Zehnder] (figure 2). Some adaptations to that process have been made, when information from a number of existing databases from partners has been integrated into the new prototype. The advantage gained, was to have a conceptual schema, that could be implemented in different DBMS (e.g. Inprise InterBase or MS Access, see figure 1) with all redundancies eliminated: every fact should only be integrated once. The disadvantage of such a systematic development was that the structure of the database became very complex and contained many more tables than all the individual databases combined.

The development of the structure is not a unique process, because there is often more than one way to model the real world. To improve a resulted model, it is often necessary to go back to the first two steps and iterate through the following processes. The first 5 steps of the systematic development method are independent of a specific DBMS and result in the conceptual schema. The last 3 steps depend on a specific DBMS and result in a logical schema necessary for the implementation of a ready-to-use database.
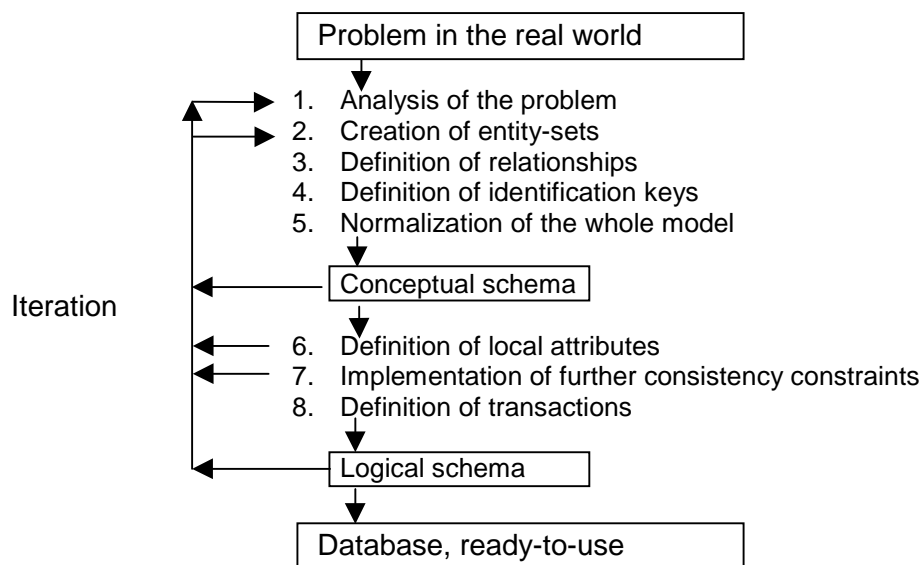
Problem in the real world

1. Analysis of the problem
2. Creation of entity-sets
3. Definition of relationships
4. Definition of identification keys
5. Normalization of the whole model

Conceptual schema

Iteration

6. Definition of local attributes
7. Implementation of further consistency constraints
8. Definition of transactions

Logical schema

Database, ready-to-use

Figure 2: The 8 iterative steps of database development adapted according to [Zehnder]

4. The conceptual schema

The conceptual schema is an independent data description. It is not designed according to a specific computer-infrastructure but to the real world. It is independent from a specific DBMS or computer-system. The conceptual schema is an orientation tool for coordinating the collaborative interaction between computer experts and other scientists, to support for changes in the system logic and to improve the long-term data planning. It is a master plan for all applications based on the schema. As happened in the DMIF-GEN database development, it's possible to extract several logical schemata for different DBMS (e.g. InterBase 5.5 or Access 2000) from the same conceptual schema (see figure 1).
The conceptual schema can either be developed manually or supported by tools. A manual development means more detail work but gives a better overview. On the other side tools overtake the work of all the details, but they seduce to involve too many details so that the overview gets lost. The conceptual schema of the DMIF-GEN database was partly developed with the help of ERWin, an entity relationship modeling software for Windows operating systems.

5. Step 1: **Analysis of the problem**

The goal of work package 4 of the DMIF-GEN project SMT4-CT96-2072 was to set up a database, which would comprise important information for the development of detection methods for commercialized genetically modified foodstuffs. This included information on the genetic elements introduced, modified DNA sequences, detection methods and literature (for details see project description in [Schreiber]). The basis for the database-development was the existing BATS-database of Wolfram Hemmed as described in the BATS-report 2/97 [Hemmer].

Through questionnaires the project partners were involved in the evaluation of the goals. In the second DMIF-GEN questionnaire all project partners were asked to analyze the importance of 13 topics of the database. This resulted in a ranking of the main goals and showed, that information about the genetic elements comprised the most important data (3,89 out of 4). This was followed by detection and PCR-Information (3,78) and approval details (3,67). The full results are available on the web under http://food.ethz.ch/dmif-gen/results/.
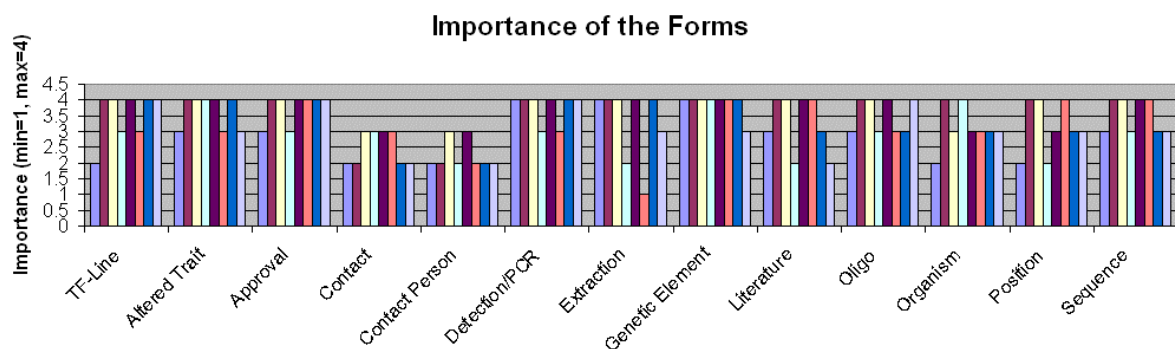

Figure 3: The importance of 13 topics of the database according to a questionnaire to all DMIF-GEN project partners [Lindenmeyer]. For a detailed description of the topics see table 2.

6. Step 2: **Creation of entity sets**

An entity is an individual element in the real world (e.g. an author name). Entities with similar properties (e.g. authors, title, date and journal of a publication) are summarized in entity sets (e.g. "Literature"). An entity set is a group of entities with the same or similar properties but different values. The attributes of the entity sets are not explored in detail until in step 6. Table 2 shows the entity sets used to describe the data of the detection methods for genetically engineered foods and the surrounding information:

Table 2: DMIF-GEN entity sets and description.

Entity set	Description of the entity
Altered trait	Describes the altered characteristic in the phenotype of the GMO as compared to the parent organism.
Contact	Specifies organizations, institutions, companies or authorities where information about the GMO is available.
Contact person	Specifies the coordinates of individual persons who can provide information concerning the GMO.
Country	Defines the country where an approval or a request for an approval has been made.
Detection	Properties of the methods to identify genetically modified foods.
Extraction	Describes the method, the costs and the quality of the procedure for the extraction of DNA (RNA) during the analysis of GMO products.
Genetic Element	Information about the type of the introduced genetic element (promoter etc.), the altered characteristic after expression of the genetic element in the GMO as compared to the parent organism, and where and when this trait is expressed.
Legal decision	Describes the regulation and the status of any legal decision (e.g. Approval, Petition, Notification) for the GMO in the countries concerned. Historically the relation "Legal decision" was first named "Approval". In order to include more general entries it was then changed into "Legal decision". Therefore, some attribute names still start with the name approval.
Literature	Provides bibliographical information of various publication types (literature references, databases, software, Internet addresses) concerning the GMO, giving information about relevant data such as production, quality, performance, safety issues etc.
Oligo	Describes DNA/RNA- or amino acid-sequences typical for a particular GMO; only primers and oligos; mostly synthetic. Whole genomes or genes or longer parts of genes, see entity set "Sequence".
Organism	Classifies the GMO according to taxonomic nomenclature.
PCR	Specifies materials and methods for the analysis of DNA or RNA of the GMO product.
Position	Describes the position relative to known genes of the parent organism and the copy number of the introduced gene.
Product	Information about the product type, the state, shape or form, as well as about the treatment and preservation of the product.
Production	Attributes about the process of development of a specific product.
Transformation line	The centerpiece of the database; according to which a GMO can be identified. This relation describes the GMO as a descendant of the parent organism, the genetic elements introduced, the production method and the resulting altered trait.
Sequence	Known DNA- or amino acid-sequences of the GMO. Mostly natural, not synthetic. Contains sequences of whole genomes or genes or longer parts of genes, but not primers and oligos. For primers and oligos, see entity set "Oligo".

7. Step 3: **Definition of relationships between the entity-sets**

In step 3 every entity set out of step 2 was linked to other entity sets, if there was a relationship between the two. A relationship is the combination of an association with its counterpart between two entity sets. An association defines how many entities out of an entity set A can be matched to a specific entity in another entity set B. Roughly the associations can be grouped into 4 types:

- Single association ("1"): One entity out of entity set A matches to exactly one entity out of entity set B.
- Conditional association ("c"): One entity out of A matches to one or zero entities out of B.
- Multiple association ("m"): One entity out of A matches to one or several entities out of B.
- Multiple-conditional association ("mc"): One entity out of A matches to zero, one or several entities out of B.

The overview over the relationships between the entity sets can be improved, if the relationships are characterized with an describing expression and the associations of the relationship. Figure 3 shows a simplified version of the entity-relationship model with the relationships between the most important entity sets of the DMIF-GEN database.
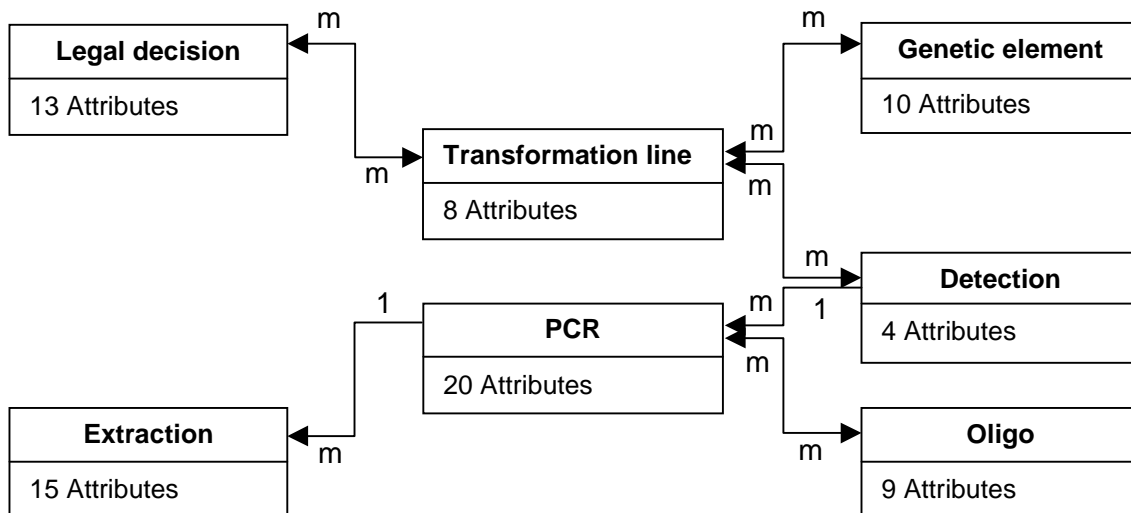
Figure 4: A simplified structure of the relationships between the main entity sets. Details and attributes in figure 6.

8. Step 4: **Definition of the identification keys**

In step 4 every entity set out of step 2 acquires an identification key. The identification key can either be naturally (existing, unique attribute) or artificially designed (special designed attribute, mainly a continuous number). The DMIF-GEN database used artificially designed identification keys. Their names are built of the entity sets name combined with the extension "_id" for identification key. E.g.: contact_id, contact-person_id, decision_id, detection_id, element_id, extraction_id, line_id, literature_id, oligo_id, organism_id, pcr_id, position_id, sequence_id, trait_id.

9. Step 5: **Normalization the whole model**

In step 5 non-hierarchical and recursive relationships are replaced by 2 hierarchical relationships (1-1, 1-c, 1-m and 1-mc) and a relationship-entity set. The names of the new relationship-entity sets are built up from the names of the two connected entity sets divided by a hyphen "-". E.g.: TF_LINE-GENETIC_ELEMENT contains the names of the transformation line and its corresponding genetic element (see figure 5). The new relationship-entity set contains the two identification keys of the connected entity sets as foreign keys. The relationship is built by the connection of these two foreign keys to the corresponding entity sets.

Figure 5: The relationship-entity set "TF_LINE-GENETIC_ELEMENT" with its 378 data sets matching each genetic element to the corresponding transformation line(s), e.g. as shown in figure 5 the genetic elements pCGN 1436, PG antisense gene, nptII, P-35, tml 3', mas 5' (mannopine synthase), mas 3' and LacZ of the Flavr Savr™ Tomato line 501-1436-1001 (for details see table 5).

The result of step 5 was a globally normalized entity-relationship model called the conceptual schema (see figures 2 and 4). With step 5 the conceptual development of the database was finished. The conceptual schema contains the entity relationship model (entity sets and relationships) and a set of identification keys out of step 4. Until now most work was done by pencil and paper and not on the computer. After step 5 started the logical development with a focus on Access and InterBase as specific DBMS as development tools (on the computer).

10. The logical schema: 2 examples of DBMS

The logic schema describes the data in the data description language of a specific database management system (DBMS). It is not DBMS-independent as is the conceptual schema. DBMS as MS Access deliver many development-tools to design a database structure and the transactions. The result of this second part of the database-development is a logical schema for a specific application and a specific DBMS. The implementation of the database is the transformation of the logical schema into commands for the DBMS by its development-tool. The logical schema imports the entity sets, the relationships and the identification keys from the conceptual schema. The development-tool of the DBMS does especially support the development of the logical schema in describing the local attributes (6th step in figure 2) and in the consistency checks (7th step): e.g. MS Access offers detailed descriptions of most aspects of the attributes and the description of the data types (9 different ones as shown in figure 8) and limited domains for a specific attribute. To offer an optimal selection in these domains (table 3) was one of the major work of step 6 in the development of the logical schema of the MS Access version of the DMIF-GEN database.

11. Decision for a specific database management system (DBMS)

Since the development of the logical scheme is dependent on a specific DBMS, the members of the work package had to choose the best candidate. To evaluate the technical background of all project members, 2 questionnaires were sent to the partners of the DMIF-GEN project: The first one in spring 1997 (presentation at the 2nd DMIF-GEN meeting in Gent in October 1997) and the second one in spring 1998 (presentation at the 3rd DMIF-GEN meeting in Rome in October 1998). The questionnaires can be accessed under: http://food.ethz.ch/dmif-gen/questionnaire/. Some results have been published in the Internet [Lindenmeyer].

Already the first questionnaire showed that nearly all DMIF-GEN partners possessed an Intel Pentium PC. Three partners using mainly Apple Macintosh hardware also had access to a PC. On the other hand, the first questionnaire revealed that 13 of the 23 partners or inter laboratory study participants were using a 32-bit operating system such as Windows NT 4.0 or Windows 95. However, most of the partners having Windows 3.X as operating system were planning to update to one of the 32-bit operating systems within the near future. 4 out of 5 DMIF-GEN partners already had MS Access as database management system. Although most partners had older versions, the present know-how in working with MS Access will be applicable for the work with any Access version. One year later, in October 1998, already 90% had version 8 of the DBMS Access. Filemaker was available to only 2 partners.

From the beginning it was a discussion in the WP4 whether to develop a local or a web-based database. Since only few partner had Internet access in 1997, the first questionnaire didn't favor an Internet version. This changed in the second questionnaire in 1998 when 95% of the partners had Internet access and 56% wanted a web-based database (see figure 6), which has several advantages compared to a local database, especially concerning access and distribution: e.g. updated data, access controls, online submission or hard- and software independent accessibility).

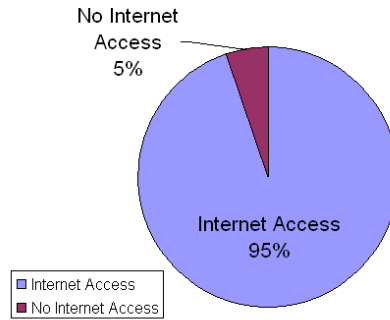**Future Versions of the DMIF-GEN database**
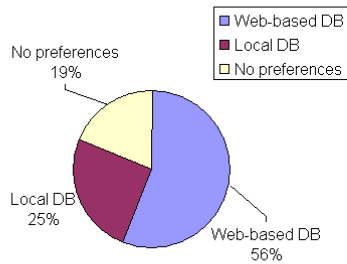
Internet Access

Figure 6: Results of the 2nd DMIF-GEN questionnaire under http://food.ethz.ch/dmif-gen/results/ concerning preferred future database and infrastructure (Internet connection).

Microsoft's Access is a flexible database-tool for personal computers with Intel processors and a Windows operating system. It contains a relational DBMS and a development tool with a shared data basis. The switch between development- and operation-mode is very easy. The main advantage of MS Access compared to its major competitor, Claris Filemaker Pro, was the wider distribution of software and knowledge among the partners of the DMIF-GEN project. Another major point were the already existing MS Access databases of some partners (BATS, SFOPH and Gene-Scan) of the DMIF-GEN project. It is also important to note, that the simplicity of Filemaker leads to several restrictions in the data structure, especially in a network structure. Moreover there are more restrictions in the design of queries in Filemaker than in Access.

The results of the first DMIF-GEN questionnaire showed that already in 1997 all DMIF-GEN partners had access to PC's with an Intel Pentium processor, more than 50% were employing a 32-bit operating system (the rest planned to upgrade) and around 80% already had MS as database management system [Lindenmeyer]. On a work package 4 (database development-group) meeting in the German Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV) in Berlin in May 1997 it was therefore decided unanimously to use MS Access as the database management system for the development of a local version of the DMIF-GEN database.

12. Step 6: **Definition of local attributes**

Before starting with step 6, the conceptual schema was imported into the chosen DBMS, Access. The import procedure contained the definition of tables and identification keys and in a second step the drawing of the relationships between the tables. The result in the "Relationship-Window" of MS Access (figure 7) looked similar to the entity relationship model, with the difference, that in the Access-model all local attributes are included as well. The import of the pencil-drawn conceptual schema (figure 4) into the DBMS Access (figure 7) was a system change and therefore contained several restrictions. E.g. the data types in MS Access are restricted to 9 types: text, number, memo, automatic value, date/time, currency, boolean, hyperlink and OLE-object. Special data types like for example the domain "A, C, G, T" for DNA-sequences are not possible. But Access offers tools to define an own domain of values in a given data type (mainly text). Dozens of such domains defined in step 6 (table 3) made it a lot easier to make transactions with the DMIF-GEN database.
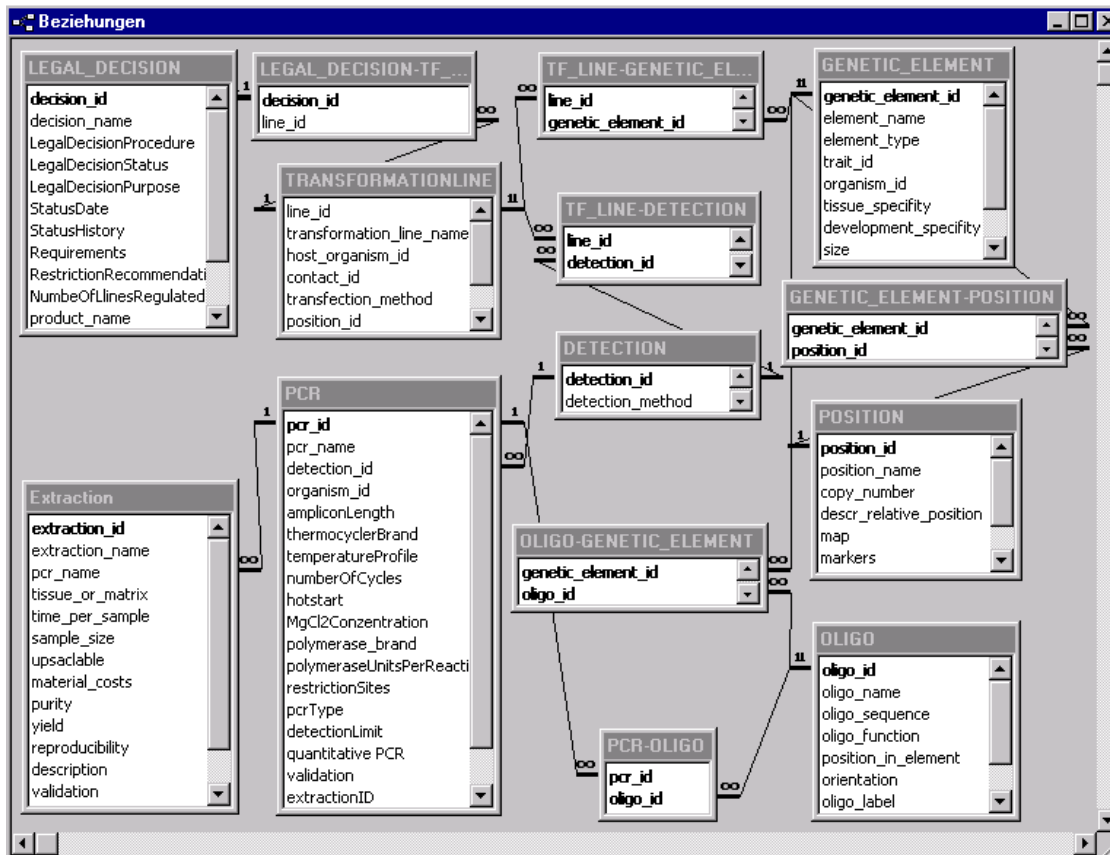
Figure 7: The "Relationship-Window" of MS Access: simplified entity relationship model from figure 4 together with all local attributes and the relationship entity-sets (e.g. figure 5).

The input of the local attributes started after the input of entity relationship model and identification keys into the new DBMS Access. Local or describing attributes are called local, because they are only used in one entity set, not like the key attributes (or global attributes).

As the development window of the table "Legal decision" shows in figure 8, each local attribute consists of a field name, one out of 9 data types and a description about the attribute. For each attribute it is possible to define if it is necessary to include data or if it is possible to leave it empty. Mostly it is necessary to include an identification key, a name and the author of the data set.

Feldname	Felddatentyp	Beschreibung
decision_id	AutoWert	data set identification number
decision_name	Text	reference name for identification of GMO approval
LegalDecisionProcedure	Text	what kind of legal decision. Selection: Petition, Notification, Deregulation, Others, Unknown
LegalDecisionStatus	Text	current status of approval. Selection: submitted, complete, pending, approved, rejected, expired, withdrawn, susper
LegalDecisionPurpose	Text	purpose of the legal decision. Selection: food, feed, seed production, field release, technical, pharmacological, resea
StatusDate	Datum/Uhr	date of beginning of actual status. Date form according to ISO 8601:1988; CCYY-MM-DD/hh:mm:ss with leading ze
StatusHistory	Text	List of all the status and dates this decision went through
NumbeOfLlinesRegulated	Zahl	number of GMO lines under the same Legal decision
comment	Memo	comment field, to collect all uncategorized information
Requirements	Text	whether labeling of the product is required
RestrictionRecommendat	Text	restrictions (geographic restriction, acreage limitation…), requirements (whether labeling of the product is required o

Figure 8: The MS Access development window of the table "Legal decision". For the corresponding view in a web-form in Internet Explorer 5 see figure 9.

To restrict the choice of possible entries, it is recommended to design special "domains" for the data to be entered. A domain is a set of different data values of the same scalar data type. Example: Petition, Notification, Deregulation are values of the attribute "Procedure" of the relation "Legal decision". They are all from the data type "text". Such domains allow the formatted presentation of attributes. This means, that the description of an attribute allows only a value out of a given domain, but no other values. Attributes with domains are similar to the 9 data types of MS Access. A domain can be tight or tolerant, if it contains only a few or many selectable values.

Table 3: 6 Examples of domains of formatted attributes of the DMIF-GEN database.

Entity set	Genetic element	Transformation_Line	Altered_Trait
Attribute	*Element_type*	*transfection_method*	*altered_trait type*
Selectable **values** of the domain	Chromosome	**Biological:**	Fungal resistance
	Expression Cassette	Agrobacterium tumefaciens	**Insect resistance:**
	Intron		Lepidoptera
	Genome	Agrobacterium rhizogenes	Coleoptera
	Integrated Sequence		Diptera
	Leader Sequence	Cell competence	Nematode resistance
	Oligo	**Physical:**	Virus resistance
	Plasmid	Microinjection	Bacterial resistance
	Promotor	Electroporation	Antibiotic resistance
	Terminator	Particle gun	**Herbicide tolerance:**
	Structural Gene (=SG):	**Chemical:**	Bromoxinyl
	altered	Calcium phosphate precipitation	Glufosinate;
	antisense		Phosphoinotricin
	altered codon usage	Polyethyleneglycol	Glyphosate
	chimeric	Liposomes	Sulfonyl urea
	truncated	Other	Altered profiles
	deletion / insertion	Unknown	Stress resistance

Entity set	Detection	Legal decision	Legal decision
Attribute	*Detection method*	*LegalDecisionPurpose*	*LegalDecisionStatus*
Selectable **values** of the domain	**Nucleotide based:**	food	received
	PCR	feed	complete
	LCR (ligase chain reaction)	seed production	pending
	RAPD(random amplified polymorphic DNA)-typing	field release	approved
		technical	rejected
	RFLP(restriction fragment length polymorphism)-analysis	pharmacological	expired
		research	withdrawn
	AFLP(amplified fragment length polymorphism)-analysis	production under containment	suspended
			extended
	probe hybridisation screening	production for extraction	End comment
	Protein based:		FRnotice
	Immunological	enzyme production	Other
	ELISA	Other	Unknown
	Protein Activity	Unknown	

To describe the entity set "Product" it would be nice to use attributes with defined domains of international standards, such as (e.g.) the internationally standardized food description language LanguaL [Hendricks], [Ireland-Ripert] under http://food.ethz.ch/langual/. LanguaL is an open international framework for food description, based on the principle of a faceted thesaurus, where each food indexed is described by a set of standard terms grouped in facets characteristic of the nutritional and/or hygienic quality of a food. Examples are the biological origin, the methods of cooking and conservation, and technological treatments. Proposed integration: 14 attributes of the relation "Product" for the 14 LanguaL-facets. Each of the attributes has to be filled in by selecting the best matching name and number out of the Windows compatible "Thesaurus Manager".

Table 4: The 14 attributes of the LanguaL-Thesaurus for the entity set product (adapted according to [Hendricks], [Ireland-Ripert]).

Attribute	Description
A. Product type	Food group having common consumption, functional or manufacturing characteristics.
B. Food source	Individual plant or animal from which the food product or its major ingredient is derived; also a chemical food source.
C. Part of plant or animal	Anatomical part of the plant or animal from which the food product or its major ingredient is derived.
E. Physical state, shape or form	The physical state of the food product. Solid food products are further subdivided by shape or form.
F. Extent of heat treatment	Heat treatment causes chemical changes and/or reduction of enzyme and of microbial activity and thus affects food safety and shelf life.
G. Cooking method	Cooking means raising the temperature of a food by heat or microwaves for a time sufficient to convert it from a raw or partially cooked state to a partially or fully cooked state. The physical and biochemical changes in the food and its components which affect the safety, palatability or nutritional characteristics of the food.
H. Treatment applied	Treatment or processes applied to the product or any indexed ingredient.
J. Preservation method	The methods contributing to the prevention or retardation of microbial or enzymatic spoilage and thus to the extension of shelf life.
K. Packing medium	The medium in which the food is packed for preservation and handling or the medium surrounding homemade foods.
M. Container or wrapping	The main container material, the container form, and the material of the liner lids or ends.
N. Food contact surface	The specific container or coating materials in direct contact with the food.
P. Consumer group, dietary use or label claim	Consumer group, human or animal, for which the food product is produced and marketed; dietary use, where the food has special characteristics, claims or uses or is intended for individuals with particular dietary needs; and label or labeling claims, used when special dietary use factor terms were derived from actual food labels.
R. Geographic places and regions	Country codes according to ISO 3166-1:1997 "Codes for the representation of names of countries and their subdivisions" [ISO] or regions and places as in the thesaurus.
Z. Adjunct characteristics of food	Additional groups of factor terms of a miscellaneous nature.

There are always unformatted attributes that cannot be restricted through a given domain, e.g. a comment field or a description field. Unfortunately long text or sequences are not easy to search by queries such as QBE- or SQL-queries. But big and powerful DBMS, such as Oracle, support the search for unformatted attributes through full-text search. Also the Internet version of the DMIF-GEN database based upon InterBase 5.5 and Java plans to provide a full-text search on its data.

13. Step 7: **Implementation of further consistency constraints**

Consistency constraints are fulfilled, if the data follow some predefined restrictions, e.g. if the data correspond to their data type, if the data are a selection of the corresponding domain as specified in step 6 and if there aren't expressions that oppose each other. Step 7 checks these consistency constraints and - if necessary - adds new ones or removes some, to give more tolerance and flexibility for the operation of the database. It is also possible to make a tighter restriction for the domains and data types than specified in a first approach in step 6. MS Access supports the consistency checks through tools like reference integrity (no connections to empty data sets) and identification keys in the design process of the relationships.

14. Step 8: **Definition of transactions**

Transactions include operations like data-mutation through forms or tables and data-extraction through queries or forms. Several members of the WP4 subgroup contributed exemplary questions concerning detection of genetically engineered foods. A collection of these questions has been provided as queries in the distributed version 1.0 (e.g.: "Which transformation-lines contain the promoter P-35S?"). By adjusting such queries new questions can easily be created. Adjusting QBE- or selection-queries is harmless to the content of the database (in contrary to the action queries). In MS Access 2000 most of the proposed questions can be implemented as QBE-queries. More complex queries are implemented with

the SQL-view in MS Access 2000. A manual for MS Access and the design of QBE- and SQL-queries provided by one of the authors is available for free on the Internet under: [Fessler et al.]. The project partners proposals of typical questions that should be implemented as queries in the DMIF-GEN database contained examples like:

- Which genetic elements are common to all approved species in a specific country?
- How can you differentiate between the different lines?
- List all lines where organism = "tomato" and promoter = "P35S".
- List the lines that do not contain the genetic elements "P35S", "NOS" and "NPT II".
- List all primers detecting "P35S", "NOS" or "NPT II".
- List the genetic elements differing between BT-11 and BT-176.
- List all legal decisions after a certain date of approval.
- List all attributes of the products which are approved, but not in the EU.
- Which approvals are covered by applying a specific PCR-system?
- Is the "P35S"-promotor in a specific line the same as the one used in another line?
- What is the copy number of a certain element in a specific transgenic line?
- Which other plant lines do have the same arrangement of genetic elements?
- Which plants do have notification/application of the EU according to the Novel Foods ordinance?
- In which countries are which GMO-plants grown at a specific point in time?
- What are the regulatory accepted methods for detection of GMOs in a particular country?

15. History of the MS Access database

In 1997 the DMIF-GEN database development group of work package 4 (WP4 subgroup) designed a conceptual schema of a comprehensive relational database providing information relevant for the development of methods for detecting genetically modified organisms (GMOs). Based on this model and based on an already existing database [Hemmer], a prototype for the DMIF-GEN database had been implemented by the Agency BATS and the Swiss Federal Office of Public Health (SFOPH), with technical support provided by the Institute of Scientific Computing of the Swiss Federal Institute of Technology (schedule see figure 1).

The developed prototype version 1.0 ran under the database management system Microsoft Access 97 (Version 8). The database consisted of 44 interconnected tables containing 210 attributes. In 1998 a limited number of data sets were entered into the database in order to allow testing of its applicability by all members of the DMIF-GEN project. The database was primarily designed to support the development of nucleotide-based detection methods. At present PCR is clearly the method of choice for detecting GMOs. However, much of the information necessary for the development of protein-based detection methods can also be deduced from the database and more specific data may be added in future versions. In the first version, the database was restricted to GMOs that have been approved for commercial use.

The first prototype was distributed at the 2nd DMIF-GEN meeting in Gent in 1997 and evaluated 6 month later in the 2nd DMIF-GEN questionnaire [Lindenmeyer]. The feedback from the questionnaire was used to develop a more user friendly version 2.0 presented at the third DMIF-GEN meeting in Rome in 1998. Based on the conceptual schema (figure 4), the German Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV) together with Joker Concept GmbH developed an Internet version 3 (figure 11) [Singer], [Jankiewicz], and the Agency BATS and the SFOPH developed an Access 2000 version with slight differences (Interface figure 9, schedule figure 1).

Authorities have (among other services) the duty of information of the public about new trends in genetically modified foods. E.g. which GM foods are imported in which amounts or what will appear on the market in the next time. Therefore the authorities (e.g. the SFOPH) need to stay up to date about GM foods. Since especially Switzerland does not receive too many petition documents for an approval procedure, it was necessary to get the needed information about new GM foods from outside, mainly from the USA. Normally the information was gathered in a background report (e.g. [Hemmer] or [Regenass-Klotz]) written through literature study, scattered publications and personal contacts. But this often takes more time and leads to more mistakes than a systematic approach with the support of general instructions and guidelines and an actualized database.
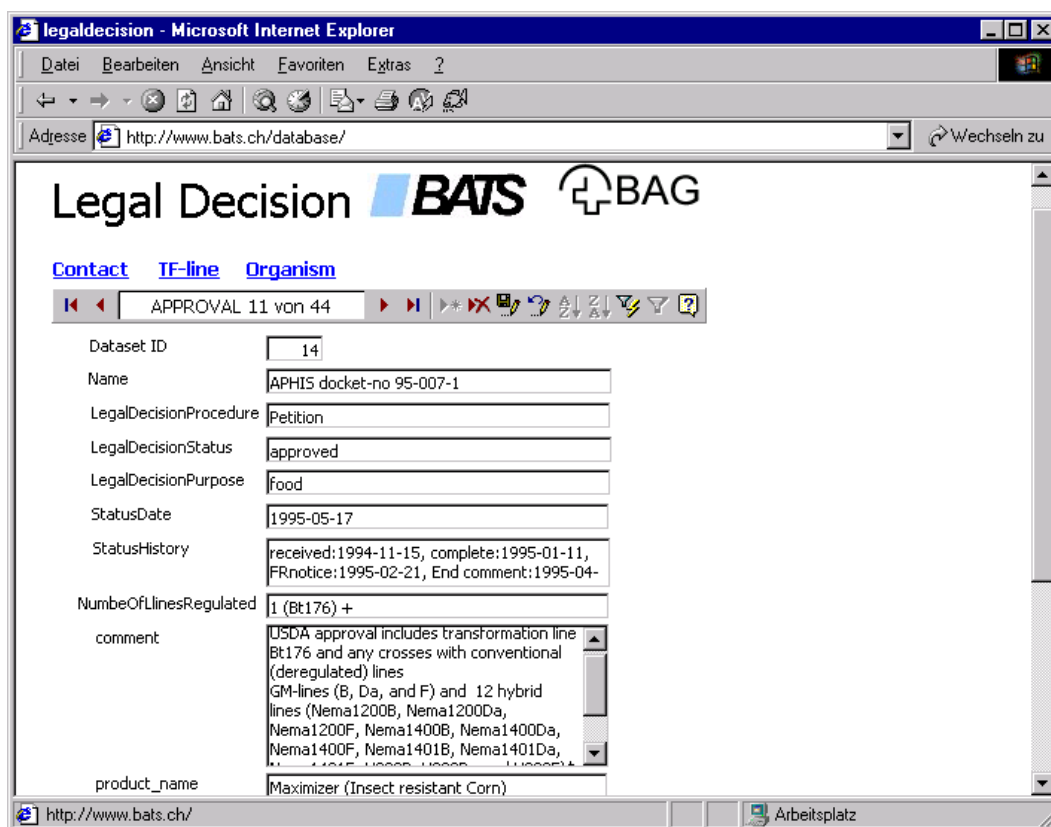
Figure 9: The Web-Interface of the BATS-BAG Access 2000 database (form "Legal decision") in a web-browser (Internet Explorer 5), presenting the corn Bt 176-data set by using the structure defined in figure 8 and the selection of the domains out of table 3.

Medium Title:	URL:	Description:
BATS-Report 2/97	http://www.bats.ch/abstr/297intro.htm	This report reviews currently approved geneticall
DMIF-GEN database online	http://www.bgw.de/GEN_DMIF_APP.html	Result of the SMT-project (EU), data by BATS He
TransGen Database	http://www.TransGen.DE/	Searchable database on world-wide releases. Cor
ISB (= Virginia Tech Information Systems for Bio	http://www.nbiap.vt.edu/cfdocs/fieldtests1.cfm	Structured and searchable database of the enviror
Biotechnology Permits Database	http://www.aphis.usda.gov/biotech/database.html	Complete database (1987-present) for movement
Federal Register Database	http://www.gpo.ucop.edu/search/fedfld.html	The database contains all Federal Register entrie
SEBIOTECH Biotechnology Database	http://www.sebiotech.org.uk/database.htm	Searchable market information reports and compa
EU field trials database	http://biotech.jrc.it/gmo.htm	Overview of all notifications circulated so far amor
OECD Database (Biotrack)	http://www.olis.oecd.org/biotrack.nsf	This searchable prototype Database for products
RKI Database on European Field Releases	http://www.rki.de/GENTEC/GENTEC.HTM	The German Robert Koch institute publishes a Sh
Arche Noah BioGene Database	http://www.biogene.org/indexx.html	Database as an instrument to evaluate the marke
Index of Public Register Entries for Notifications	http://www.environment.detr.gov.uk/acre/marketin	Index of public register entries for notifications to
RAGE Database	http://www.rage.org.nz/	The RAGE (revolt against genetic engineering) da
OURFOOD Database	http://www.56.cyberhost.net/ourfood/Genetic_moc	OUR FOOD is a database containing information
UCS (Union of Concerned Scientists) Database	http://www.ucsusa.org/search.html	Searchable website containing a database on a w
Releases of GMOs in Australia	http://www.science.gov.au/gmac/pis_book/pis_co	Archive of Public Information Sheets relating to de
FIELD TRIALS OF GMOS IN BRAZIL	http://www.fiocruz.br/cict/oquee/estrut/dect/bis/lib	Archive on releases of GMOs into the environmen
CFIA (Canadian Food Inspection Agency) Field T	http://www.cfia-acia.agr.ca/english/ppc/biotech/fie	Simple listing of field trials and registered product
The Current Status of Transgenic Crop Plants in	http://ss.s.affrc.go.jp/docs/sentan/eguide/edevelp.	One single table listing organism, main trait, struc
World Seed Statistics	http://photon-63.iprolink.ch/~assinsel/stat.htm	Tables on commercial markets for seed and plant
World Intellectual Property Organization PCT Dat	http://pctgazette.wipo.int/eng/index.html	Searchable or browsable database on internationa
esp@cenet Patent Information Service at the Eur	http://www.european-patent-office.org/espacenet/i	The European Patent Office offers the esp@cenet
FAOSTAT (Food and Agriculture Organization of	http://apps.fao.org/cgi-bin/nph-db.pl?subset=agric	FAO statistical data collection on agriculture prov
PANNA (Pesticide Action Network North Americ	http://www.panna.org/panna/whatis.html	The Pesticide Information Service (PESTIS) is a fi
NFM (Natural Foods Merchandiser) Database	http://www.nfm-online.com/	Freely searchable database on NFM (Natural Foo
National Food Safety Database	http://www.foodsafety.org/index.htm	The goal of the National Food Safety Database pr
Releases of Genetically Engineered Crops in Arg	http://siiap.sagyp.mecon.ar/http-hsi/english/conat	Simple table listing all releases of genetically eng
USPTO (US Patent and Trademark Office) Datab	http://www1.uspto.gov/patft/	USPTO offers web access to bibliographic and ful
IBM Intellectual Property Network	http://www.patents.ibm.com/home	The Intellectual Property Network (IPN) lets you s

Figure 10: Data source for background reports and detection methods. Extraction of the relation "Literature": Online databases with Internet address (URL) and description.

Because of special requirements of the Swiss Government, the Agency BATS modified together with the Swiss Federal Office of Public Health (SFOPH) the first conceptual schema according to new requirements of the authorities (e.g. information about production).

The conceptual schema of the DMIF-GEN database was slightly modified to support not only the detection of genetically modified foods, but also the acquisition of the necessary data for such background reports. This can also contain review-information from the Internet such as e.g. [Regenass-Klotz] in the case of corn Bt 176. The logical schema was thoroughly optimized to support the production of background-data for reports. The data source was mainly retrieved from the collected full-length petitions of the USA (table 1) and several data extracted from online databases as listed in figure 10.

16. Internet version of the DMIF-GEN database

Many problems concerning access and distribution of the database led to the conclusion, that an online database would be a better solution with several advantages compared to the local version: everybody has access to it from anywhere, no matter which hard- and software is used, access can be controlled through passwords for different levels and user-privileges, data-distribution is very easy because of online submission of the data and the information is always actualized.

The disadvantage of using the existing logical schema with MS Access97 as Internet DBMS was the poor performance and the low security level as an online database. Access is designed for single machines or small local networks. For the Internet version of the DMIF-GEN database it was therefore decided to design a new logical schema for a database server. InterBase 5.5 from Inprise was chosen as the new DBMS. An advantage of InterBase was the disposability of Java as fundamental part. The InterClient is a Java-driver, that can be used as applet on a client without any installations. Figure 11 shows the concept of the web-connectivity of the DMIF-GEN database using client-server architecture, Java and Inprise InterBase as DBMS.
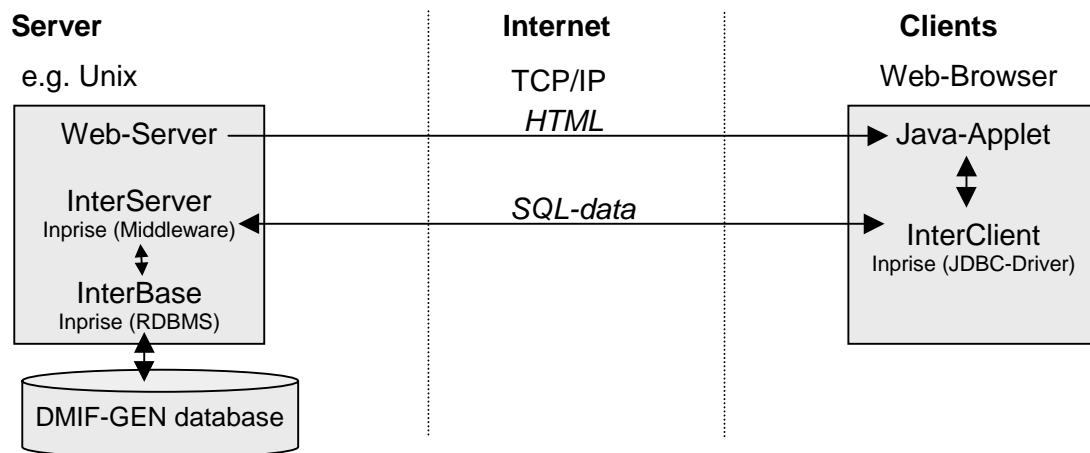

Figure 11: Web-connectivity of the DMIF-GEN database using client-server architecture, Java and Inprise InterBase as DBMS (adapted after [Singer] and [Jankiewicz]).

To use the Internet-version of the DMIF-GEN database the client needs the following hard- and software configurations:

a) Access to the Internet and at least a 56 kbit/second modem
b) JAVA enabled Web-Browser (at least Internet Explorer 4.01 or Netscape 4.5.1)
c) The latest JAVA-VM version (e.g. 1.1.6)
d) At least 64 MB RAM
e) 350 MHz Intel Pentium II CPU or G3 266 MHz Apple Macintosh

By using a computer with less performance the speed will be reduced. The Internet version can be accessed under: http://www.bgvv.de/GEN_DMIF_APP.html. The login window allows several access levels (e.g. as guest, user or administrator). The query form window in figure 12 allows the selection of one of the queries, basically corresponding to the entity sets in

table 2. It is possible to open more than one query form at the same time, but not of the same type.

The query form (e.g. Transformation line in figure 12) allows the presentation of a data set with all its connections to other data sets. The area of the direct data of the data set is in the middle on the top in figure 12, just under the title of the query form. The direct connections (1:1 and 1:c) to other tables are shown on the right and on the left (e.g. "Tomato, Lycopersicon esculentum" as organism of the presented transformation line 501-1436-1001 in figure 12). The indirect connections (1:m and 1:mc) are shown in the middle on the bottom. Because several data sets (e.g. the genetic elements polygalacturonase (PG), nptII, P-35S, tml 3' and P-mas) can be linked to the actual data set transformation line 501-1436-1001, all corresponding data sets are listed in a listbox. An empty listbox means that there are not yet any connections to other data sets of that table (e.g. PCR, Detection or Literature).

Figure 12: The query result form "Transformation line" with its connections. Data source out of table 5.

17. Data collection on DNA extraction methods

Partner 7 of Work Package 4 of the DMIF-GEN project developed an external data set on DNA extraction methods. It is a MS Access97 database, like the first version of the DMIF-GEN database. The data set on DNA extraction methods contains 36 MS Word files and also a management system for accessing the files. Each of the 36 MS Word files corresponds to one DNA extraction method. Information is given on the involved project partners, steps of the DNA extraction procedure, a list of samples on which the method was already used and the appraisal of performance, and references (if available). The data on extraction methods derive from Work Package 1.1: "Development of DNA extraction methods for raw and processed foods." Nearly all the project partners were involved in collecting the data on DNA extraction methods.

The management system for the data set on DNA extraction methods is based on MS Access 97 and is designed to facilitate the search for extraction methods suitable to a particular environment, substrate and condition. Since all the extraction data is contained in the 36 MS Word files, it is also possible to access the information without MS Access, although it would not be possible to carry out search operations such as queries or forms or other database functions.

The data set on DNA extraction methods consists of 8 tables, grouped around a primary table called "tabSubstrat" (see figure 13). The non-connected tables tabDNAModule, tabHomogenisation and tabQuantifizierung contain hyperlinks to the 36 MS Word on specific DNA extraction methods.
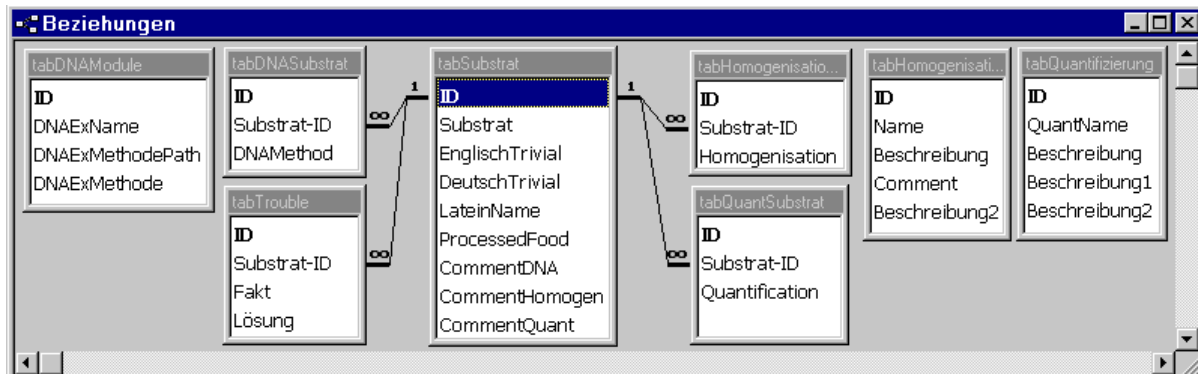

Figure 13: The relations of the Data collection on DNA extraction methods.

The graphical user interface allows access to 81 substrates and 36 possible extraction methods. The most important substrates such as corn, potato, rape, rice, soybean or tomato are divided into several subgroups that reflect the various forms in which these substrates appear in food, such as soya beans, soya granulated, soya flour, soya tofu, soya oil, soya lecithin, soya sprouts, soy sauce, soya drink, soya carob, soya chocolate, soya snack or as white bread with soya as one of the baking ingredients. For each substrate under a specific processing condition, the DNA extraction data set can find the extraction method. For example, soybean DNA from soybeans processed in chocolate cream can be extracted through "Qiagen Dneasy" or through "CTAB/Wizard method for chocolate samples" (see figure 14).

Figure 14: The Substrate-form of the data collection on DNA extraction methods.

The size of the data set on DNA extraction methods is 1.12 MB (36 MS Word sheets plus the management system); making it easy to distribute on floppy discs. In order to ensure that the correct path is taken from the management system to the word files, the data set should be copied from the floppy into a directory labeled "database" on partition C: of the personal computer. This data set is available, free of charge, from partner 7. Information on ordering and conditions is available on the Internet under: http://food.ethz.ch/dmif-gen/extraction/.

18. Selection of the data to be managed in the database

The data included in the database should be useful for the development of detection methods for genetically modified foods. A literature-review of the existing methods clearly revealed that DNA-based detection methods, most notably PCR-based ones, are the methods of choice [Köppel et al.], [Pietsch et al.], [Studer et al.], [Wurz and Willmund], [Hupfer et al.]. With respect to sensitivity, applicability for processed food stuffs, specificity and applicability in routine laboratory analysis, DNA-based methods clearly outperform protein-based methods or other possible techniques [Hemmer]. Thus, the data are specifically selected to provide genetic information on the elements that have been introduced into GMOs. The database will be restricted to GM crops approved for commercial use.

Data specifically relevant for protein-based detection methods have not yet been included into the database. However, such information can in part be deduced from the present data (e.g. use of tissue-specific promoters). Expression levels of proteins in approved genetically modified crops that have been introduced by genetic engineering have been previously described [Hemmer]. Such data may be integrated into future versions of the database. The database is also designed to include descriptions of methods that have already been employed for detecting GMOs [Köppel et al.], [Pietsch et al.], [Wurz and Willmund]. In addition, it is planned to include information on the experience of project partners (e.g. PCR-primers and probes or methods for DNA-extraction). New detection methods will also be included. [Studer et al.], [Hupfer et al.].

During the first phases of the database development, data on the various transformation lines of approved GM crops, including genetic data, were extracted from an existing database [Hemmer]. These data are exclusively derived from approving authorities, from official publications of these authorities or from petition documents of applicants, after the deletion of confidential business information. This ensures maximal reliability of the information and helps to minimize dead-end approaches caused by guessing or incomplete information. An example of such information is shown in table 5, describing the approval of the Flavr Savr™ Tomato by various US authorities. With respect to detection methods and primer systems the database was designed to receive its information from various DMIF-GEN partners. An example for this project internal information transfer is the data collection on DNA extraction methods (figures 13 and 14). Nearly all project partners contributed to this data collection.

Table 5: Data source (Flavr Savr™ Tomato) out of BATS-report 2/97 [Hemmer].

ID	1
References	USDA/APHIS; Safety assessment (Redenbaugh et al., 1992, CRC Press)
PRODUCT INFORMATION	
Plant	Tomato
Product name	**Flavr Savr™ Tomato** (MacGregor's)
Further specification	GM lines from approx. 40 different transformation events, using 2 slightly different plasmids and crosses with traditional varieties (501-1436-1001, ... , 540A-4109a-1823)
Scientific name	*Lycopersicon esculentum* Mill
Host organism	*L. esculentum*, tomato lines 501, 502, 7B, 22B, 28B, N73, 114F, 141F, 105F, 35F, 84F, 88F, 121F, 137F, 138F, 519F, 531A, 532A, 540A and 585A)
Company	Calgene Inc.
Contact	Calgene Inc.; 1920 Fifth Street; Davis, CA 95616, USA
Altered trait	Fruit ripening delayed
Classification	PQ
Purpose	Enhanced fresh market value
Safety remarks	Data on potential toxins, tomatine level, acute toxicity tests in rats
Qualities checked	Increased fungal resistance, stable inserted, taste, horticultural traits, Ca, Mg, Fe, Na; vitamins A, B1, B2, B6 and C
GENETIC INFORMATION	
Plasmid	pCGN1436 (driving nptII by mas 5' and mas 3') or pCGN4019a (driving nptII by P-35S and tml 3')
Inserted genes	Flavr Savr™ gene (= antisense polygalacturonase) (1-3 copies), nptII (1-3 copies), partial LB and RB, at a single site (haploid)
Transfection method	*Agrobacterium tumefaciens*
Transgene 1	Flavr Savr™ gene (polygalacturonase (PG) antisense gene)
Source of tg 1	Tomato
Protein product 1	None
Expression 1	No; level of native PG mRNA is >90 % reduced; residual enzyme activity of native PG is < 1 % of control lines
Mechanism 1	Antisense RNA complexes endogenous sense mRNA for PG (transcription for native mRNA might also be downregulated), thus reducing the levels of PG which normally degrades pectin, a major component of the cell wall in tomato fruit
Promoter 1	(double-) CaMV 35S (d-P-35S)
Terminator 1	tml 3'
Transgene 2	nptII (= kan^r, neo^r = neomycin phosphotransferase II gene)
Source of tg 2	Transposon Tn5 (*E. coli* K12)
Protein product 2	APH(3')II (Aminoglycoside-3'-phosphotransferase II)
Expression 2	< 0.08 % of total protein
Mechanism 2	Allows for selection during plant tissue culture. APH(3')II inactivates neomycin, kanamycin and genticin/G418) by ATP-dependent phosphorylation of the 3'-hydroxyl group of the aminohexose moiety of these aminoglycoside antibiotics. This phosphorylation interferes with uptake and binding of the aminoglycoside to the bacterial ribosome
Promoter 2	mas 5' (mannopine synthase) or CaMV35S promoter (different plasmids)
Terminator 2	mas 3' (polyA region from mannopine synthase gene of pTiA6) or tml 3' (see plasmid)
Transgene 3	parts of lacZ
Source of tg 3	*E. coli*
Protein product 3	None
Expression 3	No
Mechanism 3	-
Promoter 3	-
Terminator 3	-
Detailed sequences	pCGN1436 sequence from RB to LB
References for	pg, nptII
REGULATORY INFORMATION	
Approved by (authority)	APHIS docket-no 92-087-1, 94-096-1, 94-125-1, 95-015-1, 95-056-1, 96-080-1
Approved for (country)	USA
Restrictions	none
Requirements	-
Labeling	not required
Date	1992-10-19, 1994-10-03, 1994-11-18, 1995-03-23, 1995-07-28, 1996-10-09 (see **table 1**).
Comment	FDA approved (1994-05); EPA approval not required

19. References

1) *BAG Bulletin No. 33/1997*, Bundesamt für Gesundheit, Bern, Switzerland, pp. 7-9. http://www.admin.ch/bag/sucht/publist/d/bulletnd.htm

2) Bray, T., Paoli, J., Sperberg-McQueen, C.M. (1998) *Extensible Markup Language (XML) 1.0.* W3C Recommendation, 10-February-1998, World Wide Web Consortium. http://www.w3.org/XML/

3) Fessler, A., Kemme, B., Kuckelberg, A., Lindenmeyer, J., Ma, D., Schlotke, F., Zeller, E. (1999) *Datenverwaltung mit MS Access.* Institute of Scientific Computing, ETH Zurich, Zurich, Switzerland, WS 98/99. http://www.inf.ethz.ch/personal/lindenme/publications/access/AccessTutorial.html

4) Hemmer, W. (1997) *Foods derived from genetically modified organisms and detection methods.* BATS-Report 2/97, Basel, Switzerland, ISSN 1420-228X. http://www.bats.ch/abstr/297intro.htm

5) Hendricks, T.C. (1992) *LanguaL: An automated Method for Describing, Capturing and Retrieving Data about Foods.* In Simopoulos, A.P., Butrum, R.R. (eds): International Food Databases and Information Exchange, World Rev. Nutr. Diet., vol. 68, Basel, Switzerland, Karger, 1992, pp. 94-103.

6) Hupfer, C., Hotzel, H., Sachse, K., Engel, K.-H. (1998) *Detection of the genetic modification in heat treated products from Bt maize by polymerase chain reaction.* Z. Lebensm. Unters. Forsch.

7) Ireland-Ripert, J., Holden, J., Møller, A., Schlotke, F., Slimani, N., Varsányi, I., Volatier, J.L. (1997) *Food Description / LanguaL.* Poster presentation at the 16th IUNS International Congress of Nutrition, Montreal, Canada. http://food.ethz.ch/langual/

8) ISO-Standards: ISO 639:1988, *Code for the representation of names of languages*; ISO 2108:1992, *International standard book numbering (ISBN)*; ISO 3166-1:1997, *Country codes*; ISO 3166-2:1998, *Country subdivision code*; ISO 3297:1998, *International standard serial number (ISSN)*; ISO 8601:1988, *Representation of dates and times.* International Organisation for Standardisation (ISO), Geneva, Switzerland. http://www.iso.ch/

9) Jankiewicz, A. (1999) *Das DMIF-GEN Datenbankprojekt der Europäischen Gemeinschaft.* Primäre Schlüssel der Entwicklung von Standardmethoden zur Detektion von GMOs in Lebensmitteln. Diplomarbeit, angefertigt am BgVV, Fachbereich 2, Dr. habil. K.W. Bögl, eingereicht am Institut für Lebensmittelchemie der Technischen Universität Berlin, Arbeitskreis Prof. Dr. L.W. Kroh, Berlin, Germany.

10) Köppel, E., Stadler, M., Lüthy, J., Hübner, P. (1997) *Sensitive Nachweismethode für die gentechnisch veränderte Sojabohne 'Roundup Ready^{TM}'.* Mitt. Gebiete Lebensm. Hyg., 88: pp. 164-175.

11) Lindenmeyer, J. and Hemmer, W. (1999) *DMIF-GEN database on genetically engineered foods.* Proceedings of the European Research Project SMT4-CT96-2072: Development of Methods to Identify Foods Produced by Means of Genetic Engineering. BgVV-Hefte 05/1999: pp. 65-76. Berlin, Germany. ISBN 3-931675-44-0. ISSN 0948-0307. http://food.ethz.ch/dmif-gen/lisbon/

12) Lindenmeyer, J. and Omlin, D. (1999) *Project description of the BATS-BAG database on genetically modified foods.* Technical manual. Unpublished.

13) Lindenmeyer, J. (1997-1999) *Website of the DMIF-GEN project.* Project information especially about WP 4. http://food.ethz.ch/dmif-gen/

14) Lindenmeyer, J., Hemmer, W., Auberson, L., Omlin, D., Schrott, M., Wurz, A., Singer, P., Broll, H. (1999): *Data structure for genetically modified foods.* Structure of a

database for the DMIF-GEN project SMT4-CT96-2072; Proceeding and handout at the 4th DMIF-GEN meeting 1999 at the Institute for Biological and Experimental Technology (IBET), Lisbon, Portugal, from 4. to 5. October 1999. http://food.ethz.ch/dmif-gen/lisbon/

15) Lindenmeyer, J. (1998) *Results of the two DMIF-GEN questionnaires.* 2nd DMIF-GEN meeting Gent 1997 and 3rd DMIF-GEN meeting Rome 1998. Rome, Italy. http://food.ethz.ch/dmif-gen/results/

16) LMV (1995) *Lebensmittelverordnung (Swiss Food Ordinance) of 1 March 1995*, SR 817.02. Switzerland. http://www.admin.ch/ch/d/sr/8/817.02.pdf

17) Pietsch, K., Waiblinger, H.U., Brodmann, P., Wurz, A. (1997) *Screeningverfahren zur Identifizierung „gentechnisch veränderter" pflanzlicher Lebensmittel.* Deutsche Lebensmittelrundschau, 93 (2): pp. 35-38. Germany.

18) Regenass-Klotz, M. und Maeschli, C. (1999) *Antibiotika-Resistenzgene in Nahrungspflanzen.* Bioweb Podium, Agency BATS, Basel, Switzerland. February 1999. http://www.bioweb.ch/dossiers/antibiotikaresistenz/einfuehrung.html

19) Regulation (EC) No 258/97, (1997) *Regulation (EC) No 258/97 of the European Parliament and of the Council of 27 January 1997 concerning Novel Foods and Novel Food Ingredients.* Official Journal of the European Communities, No L 43/1-5. http://europa.eu.int/eur-lex/en/lif/dat/en_397R0258.html

20) Schlotke, F., Becker, W., Ireland, J., Møller, A., Ovaskainen, M.L., Monspart, J., Unwin, I. (2000) *EUROFOODS Recommendations for Food Composition Database Management and Data Interchange.* COST Action 99 Report, European Commission, Luxembourg, To be published. http://food.ethz.ch/schlotke/pub/2000EUROFOODSRecommendations.pdf

21) Schreiber, G. et al. (1995) *The Technical Annex. SMT4-CT96-2072. Development of Methods to identify Foods produced by means of Genetic Engineering (DMIF-GEN).* Project description. Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin (BgVV), Berlin, Germany. 13.11.1995 (Application).

22) Schrott, M., Syfrig, J. (1999) *Neue Bestimmungen zur GVO-Deklaration: Informierter Entscheid beim Lebensmittelkauf.* BioTeCH-Forum 3/99 pp. 6-7. Basel, Switzerland. http://bics.ch/g/forum/1999/3/

23) Singer, P. (1999) *Technical documentation and user manual of the DMIF-GEN database.* Joker Concept GmbH, Berlin, Germany. http://www.bgvv.de/GEN_DMIF_APP.html

24) Studer, E., Dahinden, I., Lüthy, J., Hübner, P. (1997) *Nachweis des gentechnisch veränderten 'Maximizer'-Mais mittels der Polymerase-Kettenreaktion.* Mitt. Gebiete Lebensm. Hyg., 88: pp. 515-524.

25) Wurz, A., Willmund, R. (1997) *Identification of transgenic glyphosate-resistant soybeans.* in: Foods produced by means of genetic engineering. 2nd Status Report. Schreiber, G.A., Bögl, K.W. (eds.) Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin, Berlin, Germany. BgVV-Hefte 1/1997, pp. 115-117. ISBN 3-931675-07-6. http://www.bgvv.de/publik/hefte.htm

26) Zehnder, C.A. (1998) *Informationssysteme und Datenbanken.* B.G. Teubner Verlag Stuttgart, Germany und vdf Hochschulverlag, Zürich, Switzerland. 6. Auflage. ISBN 3-7281-2019-7. http://www.vdf.ethz.ch/info/2019.html

CONCLUSION and OUTLOOK

Surpassing the goal of a stationary version to be contributed on disk, an Internet version of the DMIF-GEN database has been realized. The Internet version opens the opportunity for optimal dissemination of the project's results. Moreover, working up and integration of new data sets will be easily possible in the future. The DMIF-GEN database may present a widely used platform for data input to collect a comprehensive knowledge about GMO containing foods. The database above all addresses to laboratories confronted with method development but also to competent authorities world wide. For the latter a particular access level with confidential data would be of future interest. (The public version of the DMIF-GEN database contains no confidential data).

Actually the Internet version is upgraded and checked for bugs before it is made freely accessible. This will probably be the case in February/March 2000 with a website linked with the BgVV homepage (http://www.bgvv.de/). Detailed information about the database project in the scope of the DMIF-GEN project is to be found under http://food.ethz.ch/dmif-gen/.

The DMIF-GEN database is a relational database and used the entity relationship model as data schema. Because the relational database approach has some disadvantages (especially in data interchange [Schlotke]), it is a long-term plan to transfer the data structure of the DMIF-GEN database into an XML-application (eXtensible Markup Language [Bray]), when XML has been established as an Internet standard. XML is a meta-language for the design of markup languages such as HTML. It allows the definition of many classes of documents. With a future XML standard the data schema of the DMIF-GEN database can be transferred into a more object-oriented way.