

Matthias Frisch, Kendall R. Lamkey and Albrecht E. Melchinger

Storage of molecular marker data in databases for efficient use in plant breeding programs

With the increased use of molecular markers in plant breeding programs, appropriate storage of these data becomes an important issue. We present a standardized database design facilitating data exchange and analysis.

1 Introduction

Classical plant breeding uses phenotypic information about a plant or its relatives to improve the genotype of future generations. The phenotype of a plant is assessed by measuring, e.g., yield or resistance and superior phenotypes are assumed to be the result of superior genotypes. Breeding progress is made by mating superior genotypes to generate the next generation. This concept was extended during the last 15 years by including results from DNA analyses, so called molecular markers, in the decision process of plant breeding. Meanwhile, molecular markers have become an important tool in plant breeding (Lee 1995; Melchinger 1990; Young 1999). Areas of application include studies about genetic diversity, inheritance of quantitative characters, marker-assisted selection, and genetic fingerprints for forensic investigations as well as plant variety protection. The data underlying these various applications have the same general structure.

Molecular marker data are analyzed either with statistical software, for example Piepho and Koch (2000) used SAS (R) (SAS-Institute 1988), or data are analyzed with programs written especially for this task such as Arlequin (Schneider et al. 1997), GDA (Lewis and Zaykin 1999), G-MENDEL (Holloway and Knapp 1993), or PLABQTL (Utz and Melchinger 1996). Often experimental data are stored either as input files of such programs or with spreadsheet software. Because neither of these options are designed for data storage the following problems often arise: (1) The same data are repeatedly stored in different locations, this may provoke data inconsistency if only one copy of the dataset is changed and requires an unnecessarily large amount of storage capacity; (2) only the experimenter can reproduce the coding and structure of the stored data, which complicates reanalysis of the data; (3) considerable time is required to convert data stored in a certain format into another format that can be input in larger databases or another software and (4) combining data from several experiments for a joint analysis is difficult. To our knowledge, there exists no concept for efficient storage of molecular marker data, which focuses on applications in a plant breeding program.

Our objective was to develop a data structure for storage of molecular marker data in databases, which overcomes the shortcomings of data management in spreadsheets or input

files of analysis software. The proposed data structure avoids redundant storage of experimental data and provides a standardized storage format, which facilitates retrieval, reanalysis, and exchange of the data.

2 Database structure

The basic entity of data storage is the observation of presence or absence of an allele at a marker locus for a DNA sample in a study. In the present context, we used allele to describe one of several possible outcomes of an experiment, which generates distinguishable results when applied to different DNA samples. The term marker was used to describe the combination of all methods used to generate these results. In a more abstract terminology: Marker was used as a synonym for 'polymorphism generating experiment' and allele as synonym for 'result of a polymorphism generating experiment'. In order to trace the origin of a DNA sample to the individual from which it was generated, a unique identifier is assigned to each sample. A study comprises the application of a set of experiments to a set of DNA samples.

Molecular marker data are obtained by scoring the banding pattern on an electrophoresis gel (Fig. 1) or by a DNA sequencer. Each band on a gel or each peak detected by the sequencer results from DNA of a certain fragment length. The presence or absence of bands, certain banding patterns, or peaks for a certain DNA sample are assessed, and the presence or absence of the corresponding alleles is determined. A marker can have one or several alleles, for example a single sequence repeat (SSR) marker generates fragments of varying length, each of which is regarded as an allele. In contrast, an amplified fragment length polymorphism (AFLP) marker generates only one single band, which is either scored qualitatively or quantitatively (Piepho and Koch 2000) by measuring the intensity of the band with a DNA sequencer. An allele can also consist of a certain banding pattern, e.g., for a restriction fragment length polymorphism (RFLP) marker.

Considering the observation of an allele in a genotype in a study, there are three possibilities: (1) The allele was observed, (2) the allele was not observed, or (3) the result of the experiment is unknown. The information stored in the database represents which of these three events occurred for

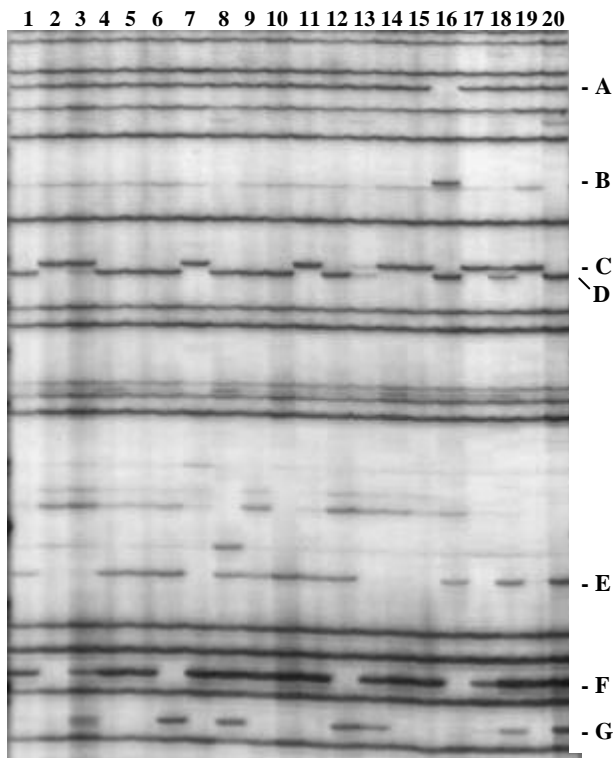


Fig.1: Analysis of genetic diversity in corn salad (*Valerianella locusta* L.) using AFLP markers with the primer combination *Eco*-AGT x *Mse*-CG. The numbers 1 to 20 denote the DNA samples from 20 inbred lines, A to G are polymorphic markers (*Jasmina Muminovic, unpublished data*).

each combination of DNA sample and allele. In the case of qualitative scoring of AFLPs, also the intensity of the band is stored. For a study with i individuals and k markers of which each has a_k alleles, $n = i \cdot \sum_k a_k$ database entries are required. (Obviously only those alleles occurring at least once in the study are relevant.)

In studies, where at least one allele of each marker was observed for each DNA sample, the complete information can be stored by generating a database entry with only observed and missing alleles. Each combination of DNA sample and allele, for which no database entry is present, indicated that the respective allele was not observed. In such a study with m missing values, the maximum of $n = 2ik + m$ database entries are required. This storage mode requires considerably less database entries than storing also unobserved alleles (the amount of saving depends on the average number of alleles per marker), it usually can be applied with high quality data sets.

The core of the database structure is a table named 'observed marker data', which holds the information about the results of the experiments. For each database entry one row in this table is generated, consisting of columns for study, identifier, allele, and state.

An example for entries in the 'observed marker data' table is shown in Table 1. The first data block shows results from an RFLP study with maize inbred lines named 'tigs'. (The

name of the studies can be chosen freely.) In this study only observed alleles (state = 1) and missing alleles (state = 9) were entered into the database. For example, line CML117 was heterozygous at marker BNL5.62, carrying alleles 10.975 and 394, while line CML118 carries allele 11.884 homozygous. At line CML118, the observation for the allele 9.979 at the marker UMC164 is missing. The second block of data shows results from an AFLP study with wheat inbred lines. The data shows markers M004 to M010, resulting from applying the primer P6061M49 to the inbred line D01. In this study, the complete information for each combination of line and marker was entered into the database (state = 0 for absent alleles in addition to the 1's and 9's). (With qualitative scoring the 1's for observed alleles would be replaced with a number coding for the scored intensity of the band.) In the third data block, results from a SSR study with a maize population are shown. For example, the first individual of Population 22 (coded by POP22-1) is heterozygous at marker phi08, it carries the alleles 360 and 366.

Additional tables contain information about (1) the assignment of the analyzed individuals to taxonomic units (table 'list of identifiers'), (2) alleles and markers (tables 'list of markers' and 'list of alleles'), and (3) studies and used lab protocols ('list of studies' and 'list of lab protocols'). A diagram with all tables, the corresponding column definitions, and the relations between the tables is shown in Fig. 2.

The table 'list of studies' describes the studies for which data are stored in the database, it contains information about the mode of entering the data (either complete or only observed and missing alleles), a title for the study, the year when it was carried out, the lab where it was carried out, the name of the responsible person(s), and the name of the lab protocol used. Information about lab protocols are stored in a table named 'list of lab protocols'. The lab protocol should be comprehensive and describe each step in such a detail that it could be reproduced by others. Because each lab protocol can be applied to several studies, the table 'list of lab protocols' is related 1:n with the table 'list of studies' via an identifier for each lab protocol. The table 'list of studies' is related 1:n with the table 'observed marker data' via the name of the study (each study consists of many entries in table 'observed marker data').

The table 'list of identifiers' contains information about the taxonomic units to which an identifier belongs and a corresponding name. At least one entry for each identifier is required, which is generated during the data entry. In addition to this primary identification, an identifier can be assigned to an arbitrary number of further taxonomic units. Consequently the tables 'list of identifiers' and 'observed marker data' are related n:n via the description of the identifier. The following example illustrates this concept (Table 2): Consider that during data entry an identifier was assigned to the taxonomic unit 'maize-individual' with the name 'POP22-1', this plant can subsequently be assigned to the taxonomic unit 'maize-population' with the name 'POP22' and to the taxonomic unit 'maize-heterotic-pool' with the name 'A'.

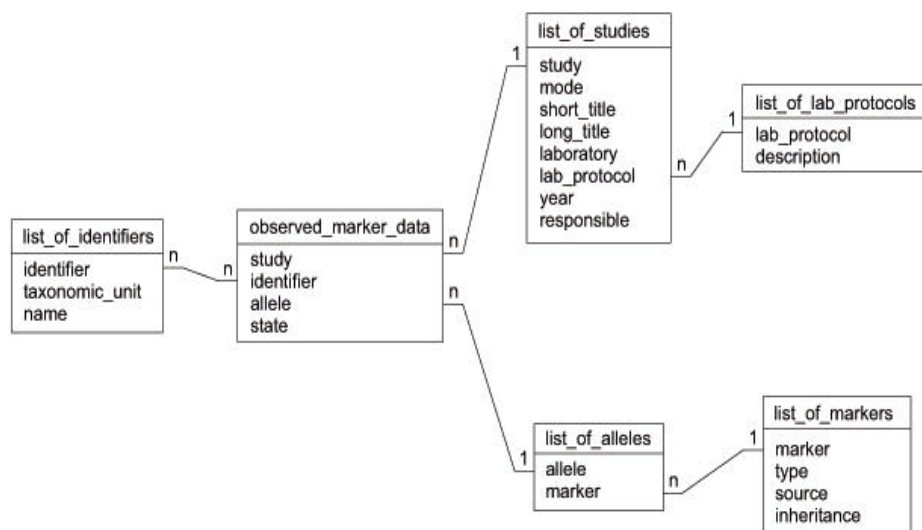


Fig.2: A diagrammatic representation of the structure of the database described. Each rectangle symbolizes a table in the database, the first line of text in the rectangle refers to the name of the table and the subsequent lines of text refer to the column headings. The lines symbolize the logical structure between the contents of the tables by joining rows with identical content. At the ends of each lines the symbols 1 or n describe the relationship: A 1 next to a table indicates that the respective entry occurs only once in this table, while an n means that it may occur an arbitrary number of times in a table.

Information about markers is stored in the 'list of markers', it holds for example information about the type, source, and inheritance. A list of all alleles that were observed by applying a certain marker is stored in the table 'list of alleles'. The table 'list of markers' is related 1:n with the table 'list of alleles' (each marker can have several alleles), and the table 'list of alleles' is related 1:n with the table 'observed marker data' (each allele can be observed at several DNA samples).

3 Discussion

Implementation of the above database structure on a personal computer proved to be in the scope of experimenters who are responsible for the respective statistical data analysis. We provide a sample implementation, which uses the database software SQL-Server (TM) as back end for data storage and the statistical software SAS (R) as front end for data import, data processing, and printing reports along with a detailed description and the complete source code. The sample implementation comprises routines to create tables and indexes, import data from a wide range of raw formats, export data to a format that can be used for data exchange, and update data in order to bring studies using different nomenclature into line. Furthermore, routines to print reports, retrieve data such that they can be analyzed with SAS, and download datasets from an internet server (which uses the proposed data model) are included.

Application of the relational database model together with a medium degree of normalization of the underlying tables allows the integration of data from these small personal databases into large-scale production databases. While the internal format of a large-scale implementation of the proposed data structure requires further structural components (e.g., linking the tables with automatically generated numerical primary keys or using triggers to assure

referential integrity more conveniently), the presented tables can be used as templates for views used for data retrieval and data input by plant breeders. This ensures a common interface for data exchange, standardized analyses, and joint analyses of data from different sources.

The flexibility of the proposed data structure as a working tool for the experimenter and the possibility of integration in larger solutions is one of its main advantages. In contrast, the integration of data stored in spreadsheet format or in input format for analysis software in large-scale database systems is difficult, time consuming, and prone to errors.

Oftentimes, molecular marker data are generated from individuals or populations for which other data are also stored in databases. Examples are pedigree databases or yield trial databases. The taxonomic unit for which these data were collected may vary and differ from the taxonomic unit used for DNA extraction and marker analyses. By assigning an identifier in the marker database to the taxonomic unit for which external data is stored, marker data and external data can be joined and queried together. If there are existing databases for marker or allele information, the 'list of alleles' table and the 'list of markers table' can be replaced by the existing tables (or in the case of a different organization with corresponding views).

The possibility to link the marker data to other existing data sources enables the experimenter to perform a wide range of data analyses. Performing such analyses with separately stored data would require extensive data editing, which is time consuming and a source of errors.

We applied the proposed data model to store the data of several fingerprinting projects: (1) An RFLP study comparing the genetic diversity in progenitor and derived lines of the

Table 1: Example for entries in the 'observed marker data' table, showing results from a RFLP study with maize inbred lines, an AFLP study with wheat lines, and a SSR study with maize populations.

Study	Identifier	Allele	State
Tigs	CML117	BNL5.62/10.975	1
Tigs	CML117	BNL5.62/3.949	1
Tigs	CML117	UMC164/11.479	1
Tigs	CML117	UMC164/37.445	1
Tigs	CML118	BNL5.62/3.295	1
Tigs	CML118	UMC164/9.979	9
Tigs	CML120	BNL5.62/11.884	1
...
wheat01	D01	P6061M49-M004	0
wheat01	D01	P6061M49-M005	0
wheat01	D01	P6061M49-M006	1
wheat01	D01	P6061M49-M007	1
wheat01	D01	P6061M49-M008	0
wheat01	D01	P6061M49-M009	0
wheat01	D01	P6061M49-M010	0
wheat01	D01	P6061M49-M011	0
wheat01	D01	P6061M49-M012	1
...
cimmyt03	POP22-1	phi08-360	1
cimmyt03	POP22-1	phi08-366	1
cimmyt03	POP22-17	phi08-366	1
cimmyt03	POP22-18	phi08-366	1
cimmyt03	POP22-19	phi08-360	1
cimmyt03	POP22-3	phi08-370	1
cimmyt03	POP22-21	phi08-363	1
cimmyt03	POP22-5	phi08-366	1
cimmyt03	POP22-2	phi08-363	1
cimmyt03	POP22-22	phi08-372	1
...

Table 2: Example for entries in the 'list of identifiers' table, showing individuals used in a RFLP study with maize inbred lines, an AFLP study with wheat lines, and a SSR study with maize populations.

Identifier	Taxonomic unit	Name
...
CML117	maize-line	CML117
CML117	maize-line	CML117
CML118	maize-line	CML118
CML120	maize-line	CML120
...
CML117	maize-heterotic-poolA	...
CML118	maize-heterotic-poolA	...
CML118	maize-heterotic-poolB	...
...
D01	wheat-line	D01
...
POP22-1	maize-individual	POP22-1
POP22-17	maize-individual	POP22-17
POP22-18	maize-individual	POP22-18
POP23-1	maize-individual	POP23-1
...
POP22-1	maize-population	POP22
POP22-17	maize-population	POP22
POP22-18	maize-population	POP22
POP23-1	maize-population	POP23
...
POP22-1	maize-heterotic-poolA	...
POP22-17	maize-heterotic-poolA	...
POP22-18	maize-heterotic-poolA	...
POP23-1	maize-heterotic-poolB	...
...

reciprocal recurrent selection program with Iowa Stiff Stalk Synthetic and Iowa Corn Borer Synthetic No. 1 maize populations (Hagdorn et al. 2002); (2) a SSR study about genetic diversity in tropical maize populations and the relation between genetic diversity and heterosis (Warburton et al. 2002, Reif et al. 2002); (3) a study assessing the quality of SSR data from maize inbreds originating from different sources of maintenance breeding (Heckenberger et al. 2002); and (4) a comparison of different marker systems in wheat cultivars (Bohn et al. 1999).

Using the proposed data structure created large synergy effects: (1) The need for editing data files, which is considerably error-prone, was eliminated. (2) Using the database together with a standard statistical software system assured a high quality of statistical analysis. (3) Due to the standardized data structure, it was possible to reuse statistical analysis routines programmed for one project in other projects with related subjects. (4) The data base assured the quick availability of data from earlier studies for reanalysis.

Acknowledgments

We thank Drs. Martin Bohn, Joanne Labate, and Marilyn Warburton for comments and suggestions on the database structure and the sample implementation. We also thank the anonymous reviewers for comments which helped to improve this paper.

Note

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

4 References

BOHN, M., Utz, H. F., Melchinger, A. E. (1999): Genetic diversity among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. *Crop Science* 39: 228-237.

HAGDORN, S., Lamkey, K. R., Frisch, M., Guimaraes, P.E.O., Melchinger A. E. (2002): Molecular Genetic Diversity among Progenitors and Derived Elite Lines of BSSS and BSCB1 Maize Populations. *Crop Science*. In press.

HECKENBERGER, M., Melchinger A. E., Ziegler J. S., Joe, L. K., Hauser, J. D., Bohn, M. (2002): Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data. *Molecular Breeding*. In press.

HOLLOWAY, J. L., Knapp, S. J. (1993): G-MENDEL 3.0 software for the analysis of genetic markers and maps. Oregon State University, Corvallis, Oregon.

LEE, M. (1995): DNA markers in plant breeding programs. *Advances in Agronomy* 55: 265-344.

LEWIS, P. O., Zaykin, D. (1999): Genetic Data Analysis: Computer program for the analysis of allelic data. Dept. of Statistics, North Carolina State University, Raleigh, North Carolina.

MELCHINGER, A. E. (1990): Use of molecular markers in breeding for oligogenic disease resistance. *Plant Breeding* 104: 1-19.

PIEPHO, H.-P., Koch, K. (2000): Codominant analysis of banding data from a dominant marker system by normal mixtures. *Genetics* 155: 1459-1468.

REIF, J. C., Melchinger, A.E., Xia, X.C., Warburton, M.L., Hoisington, D.A., Vasal, S.K., Srinivasan, G., Bohn, M., Frisch, M. (2002): Genetic diversity within and between seven tropical maize populations investigated with SSR markers and relation to the heterosis of their crosses. In review.

SAS-INSTITUTE (1988): SAS/STAT User's guide, Release 6 Edition. SAS-Institute Inc., Cary, North Carolina.

SCHNEIDER, S., Kueffer, J.-M., Roessli, D., Excoffier, L. (1997): Arlequin: A software for population genetic data analysis. *Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.*

UTZ, H. F., Melchinger, A. E. (1996): PLABQTL: A program for composite interval mapping of quantitative trait loci. *J. Quant. Trait Loci* 2 (1).

WARBURTON, M., Xia, X., Crossa, J., Franco, J., Melchinger, A. E., Frisch, M., Bohn, M., Hoisington, D. (2002): Large scale fingerprinting methods for the analysis of genetic diversity of CIMMYT maize germplasm. *Crop Science*. In press.

YOUNG, N. D. (1999): A cautiously optimistic vision for marker-assisted breeding. *Molecular Breeding* 5, 505-510.

Storage of molecular marker data in databases for efficient use in plant breeding programs (M. Frisch, Kendall R. Lamkey and Albrecht E. Melchinger)

Summary

With the increased use of molecular markers in plant breeding programs, appropriate storage of these data becomes an important issue. The concept for storage of molecular marker data in databases proposed in this paper is simple and generic so that it can be implemented on a personal computer and integrated in large-scale database systems. Application of the proposed data structure simplifies standardized statistical analyses and reanalyses of experimental data as well as data exchange and reuse of programmed analysis routines.

Key words: *Computer database, molecular marker data, plant breeding*

Speicherung molekularer Marker Daten in Datenbanksystemen ermöglicht deren effiziente Anwendung in der Pflanzenzüchtung (M. Frisch, Kendall R. Lamkey and Albrecht E. Melchinger)

Zusammenfassung

Mit dem zunehmenden Einsatz molekularer Marker in der Pflanzenzüchtung wird eine sinnvolle Archivierung von Markerdaten immer wichtiger. Das hier vorgestellte Konzept zur Speicherung von Markerdaten in einer Datenbank ist einfach und allgemein gestaltet, so daß es zum einen auf PCs

aber auch als Teil von größeren Datenbanksystemen implementiert werden kann. Der Einsatz der vorgeschlagenen Datenstruktur vereinfacht eine standardisierte Datenauswertung, die Reanalyse von Experimenten sowie den Datenaustausch und die Wiederverwendung programmierter Analyseroutinen.

Stichworte: *Datenbanksystem, Molekulare Marker Daten, Pflanzenzüchtung*

Dr. Matthias Frisch is research associate at the Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany.

Prof. Dr. Kendall R. Lamkey is professor of corn breeding and research geneticist at the USDA-ARS, Department of Agronomy, Iowa State University, Ames, Iowa 50011-1010, USA.

Prof. Dr. Albrecht E. Melchinger is professor of applied genetics and plant breeding at the Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany.

Gehörtes und Gelesenes – Seen Elsewhere

Ein ganzes Dorf ist online

In Deutschland gibt es das erste Internet-Dorf, die knapp 300 Einwohner zählende Hunsrück-Gemeinde Oberhambach im Landkreis Birkenfeld. Die Computer mit Internet-Anschluss sollen die fehlenden Läden im Ort ersetzen. Die Idee von der umfassenden PC-Versorgung, die vor rund einem Jahr von Landesregierung und örtlicher Initiative BIR inform e. V. geboren wurde, war, dass jeder Haushalt einen von Sponsoren gespendeten Computer mit ISDN-Anschluss erhielt. Die Initiative, die auch wissenschaftlich begleitet wurde, hat zumindest das Interesse an der neuen Technik gefördert. Allerdings ist die Anfangs-Euphorie der Oberhambacher beim Online-Einkauf schnell verfliegen, da die bestellten Waren oft teurer waren und teilweise nicht pünktlich geliefert wurden.

(gekürzte dpa-Meldung von Udo Lorenz vom 16.04.2002)

Buchbesprechungen

Doluschitz R., Spilke J. (Hrsg): Agrarinformatik
Verlag Eugen Ulmer, Stuttgart 2002, ISBN 3-8252-2230-6, 24.90 Euro

Mit Gründung der GIL wurde im Jahr 1980 in Deutschland eine Plattform geschaffen, welche die beginnenden Forschungsaktivitäten auf dem Gebiet der Agrarinformatik - damals wagte man allerdings noch nicht, diesen Begriff zu verwenden - unterstützen und die Präsentation sowie den Austausch von Ergebnissen ermöglichen sollte. Die Schaffung der Schriftenreihe 'Informationsverarbeitung Agrarwissenschaft' (später umbenannt in 'Agrarinformatik') bis zur Herausgabe der 'Zeitschrift für Agrarinformatik' waren weitere wichtige Schritte in Richtung der Ausformung der Agrarinformatik als eigenständiges Forschungs- und Lehrge-

biet. Bis heute sind mehrere Buchtitel erschienen, die im wesentlichen nur Teilaspekte der Agrarinformatik behandeln. Das Erscheinen eines umfassenden Lehrbuches über Agrarinformatik ist aus dieser Sicht geradezu überfällig: ein Glücksfall, daß sich ein Team kompetenter Autoren zur Bearbeitung des ehrgeizigen Zieles gefunden hat, in einem Buch den weiten Bogen von den Grundlagen bis zu den verschiedenartigen Einsatzmöglichkeiten der Informatik im Agrarbereich zu spannen und damit einen weiteren Meilenstein bei der Entwicklung der Agrarinformatik zu setzen.

Besonders wertvoll ist das Buch dort, wo nicht nur bestimmte Informatik-Methoden beschrieben, sondern parallel deren Anwendung auf agrarische Problemstellungen erläutert wird:

- Beispielsweise möchte ich allen, die in verfügbaren EDV-Lehrbüchern schwierige Themen wie 'relationales Datenmodell und Normalformen' überblättert haben, den Abschnitt über Datenbank-Modelle zur Lektüre empfehlen; denn am Beispiel 'Herdenmanagement von Milchschafen' wird die Modellierung von Daten gut verständlich auch in den Details der zu berücksichtigenden Informatik-Aspekte aufbereitet.
- Für sehr gut gelungen halte ich auch den Abschnitt, in dem eine Einordnung der Agrarinformatik z.B. mit einer Darstellung der Bedeutung von Information vorgenommen wird und viele ökonomische Aspekte diskutiert werden. Dies gilt in gleicher Weise für die Kapitel über Prozesssteuerung in der Pflanzen- und Tierproduktion, die trotz einer sehr hohen Informationsdichte auch für Nichtexperten gut verständlich sein dürften. Es darf nicht überraschen, wenn es bei der ersten Auflage eines Buches auch Einiges zu bekräfteln gibt:
- Jeder Leser wird schnell ein paar Schreibfehler und Layout-Schwächen (verbesserungsfähige Grafiken u.a.) entdecken. Nach meinem Eindruck wären diese aber größtenteils vermeidbar gewesen, wenn im Verlag die Lektorats-Aufgaben ernster genommen würden.
- Es offenbaren sich dem Leser auch an manchen Stellen kleinere Abstimmungsprobleme zwischen den Autoren; so wird beispielsweise in einem Abschnitt behauptet, Google sei die effektivste Suchmaschine, im Hauptabschnitt über Suchmaschinen werden aber nur Altavista und Fireball (nicht Google) als Beispiele erwähnt.
- Auch inhaltliche Kritikpunkte sind vorzubringen: die speziellen Anforderungen in Agrarsparten wie Forstwirtschaft oder Gartenbau wurden zu wenig berücksichtigt; beispielsweise wird das wichtige Thema 'Gewächshauscomputer / Prozesssteuerung im unter-Glas-Anbau' nicht bearbeitet.

Das Buch wird in jedem Fall auf meiner Literatur-Liste zu EDV-Lehrveranstaltungen eine zentrale Position finden, weil mit ihm eine wichtige Lücke geschlossen wird. In diesem Sinn wünsche ich dem Buch eine rasche Verbreitung, was zusätzlich den positiven Nebeneffekt hätte, daß den Autoren bzw. dem Verlag baldmöglichst die Chance zur Beseitigung der angesprochenen formalen und inhaltlichen Defizite in einer 2. Auflage des Buches gegeben wird.

Georg Ohmayer, FH Weihenstephan, Freising

C. Brandstetter und M. Fries: E-Business im Vertrieb, Potenziale erkennen, Chancen nutzen – Von der Strategie zur Umsetzung, München, Wien: Carl Hanser Verlag, 2002. ISBN 3-446-21791-6; 359 S.; EURO 49,90

Mittlerweile sind eine Reihe von Büchern zum Thema „E-Business und Internet“ erschienen. Die Chancen, die für Unternehmen aus dem EBusiness erwachsen, sind enorm. Der Weg bis zur erfolgreichen Umsetzung ist jedoch meist sehr steinig, Fehler in der Umsetzung werden oft spät oder gar nicht erkannt. Als Indiz hierfür dient das Scheitern vieler Internet Start-Ups und E-Business-Vorhaben.

Die beiden Autoren C. Brandstetter und M. Fries, E-Commerce Manager bei T-Systems debis Systemhaus in Düsseldorf, sind dem Ziel des Buches, E-Business Vorhaben für den praktischen Geschäftsbetrieb zu modellieren unter Berücksichtigung der Kundenbedürfnisse, sicherlich gerecht geworden. Dabei werden B2B- und B2C-Beziehungen gleichermaßen berücksichtigt. Grundlage des Buches stellt ein idealtypischer Ablauf eines E-Business Projektes dar, anhand eines fiktiven Unternehmensbeispiels wird die Umsetzung der dargestellten Vorgehensweise praktisch beschrieben.

Das Buch gliedert sich neben einer informativen Einleitung mit einer Kurzdefinition des Begriffs „E-Business“ in zwei Hauptteile. Im ersten Hauptteil erfolgt eine umfassende Darstellung der Grundlagen zum Vertrieb über das Internet. Dabei werden Aspekte der Kundenorientierung, die Erschließung neuer Geschäftsfelder über die IT-Technologie, die Bedeutung des Vertriebs, die durch die Nutzung der IT-Technologie erwachsenden Chancen und Potenziale und die Herausforderungen und Hürden, die durch das E-Business für die Unternehmen erwachsen, anschaulich dargestellt. Aufbauend auf diesen Grundlagen widmet sich der zweite Hauptteil der praktischen Umsetzung von E-Business Vorhaben im Unternehmen, mit den Unterkapiteln Strategie, Konzeption, Implementierung und Betrieb.

Das Buch wendet sich in erster Linie an Praktiker, zu deren Aufgaben die Umsetzung und Überwachung von IT-Projekten zählen. Eine Vielzahl von Vorgehensmodellen und Checklisten, teilweise mit grafischer Unterstützung, helfen bei der praktischen Umsetzung im eigenen Unternehmen. Als zusätzlicher Service besteht die Möglichkeit, Vorgehensmodelle und Tools zur Projektrealisierung kostenlos auf der Website zum Buch herunterzuladen.

Insgesamt ist das Buch „E-Business im Vertrieb“ ein gelungenes Werk, das anschaulich und praxisnah diverse Aspekte des Vertriebs und Support über das Internet darlegt. Die klare Beschreibung der verschiedenen Schritte von der Strategie bis zur Implementierung und dem laufenden Betrieb zeichnen dieses Buch aus und die Autoren liefern interessante Ansatzpunkte, wie E-Business- Strategien auch nach der abgeflauten Interneteuphorie erfolgreich sein können.

M. Emmel, Universität Hohenheim im Mai 2002