TOWARDS CO-OPERATIVE USE OF DATA FROM CROP VARIETY TRIALS

W.A. Dekkers[1] and **G. Waksman**[2]

[1] *Applied Research for Arable Farming and Field Production of Vegetables*
P.O. Box 430, 8200 AK Lelystad, TheNetherlands,
email: w.a.dekkers@pav.agro.nl

[2] *Association de Coordination Technique Agricole*
149, rue de Bercy, F - 75595 Paris cedex12, France
email: waksman@acta.asso.fr

Abstract: In a European research project, ADDA, the feasibility of exchange and mutual use of data from trials of crop varieties has been studied as an example of data exchange. Data dictionaries have been compared and numerous small differences were shown. These differences prevent the exchange and common interpretation of data. To facilitate the interchangeability of data two lines of development have been followed:

- a theoretical one the "ADDA" data dictionary;
- a technical one the data exchange format SFD.

Jointly these approaches enable data transfer between Member States of the agricultural scientific and technical data required by the Commission to meet single market needs and regulations e.g. in the field of variety testing. This paper focuses on the development of the data dictionaries.

Keywords: database; data dictionary

1. Introduction

In the European Community numerous organisations are involved in agricultural research, research on similar subjects. The exchange of information on research results occurs mainly by publication. In these publications only a restricted amount of information is presented and reuse of these data is limited. The exchange of data and communal use of experimental results leads to many problems. Differences exist in the systems of recording experimental results and there are differences in the method of observation and way of expressing the results. It is not easy to use data from different sources for an analysis. One has to ascertain that similar items are comparable, learn the structure of the database and a conversion of the data to a common dimension might be necessary.

A joint project was started in the AIR programme under contract number AIR3-CT94-1330 (PL921330). The project was co-ordinated by ACTA (France). Participants were IACR-Rothamsted (United Kingdom), INIA (Spain), DINA(Denmark), ATO-DLO and PAGV (Netherlands). The

project is aimed at developing a prototype of a method for exchanging research results of variety experiments as data. This article considers how to arrive at a common standard data dictionary with unique descriptions of the variables. These data technically could be exchanged with SFD, (Kristensen et al. 1993, Dindorp et al, 1996) and statistical analysis is possible because of the gateways with statistical programmes, like SAS, (Kristensen et al. 1993) and Genstat (Payne, R., 1996, pers. com.).

2. Possible origins of differences between variates

Variates are the properties which are measured on the experimental units of trials or are derived from primary observations. They define the response of an experimental unit to the treatment and experimental conditions. In order to be able to compare variates from different experiments variates should be well defined and measured in the same way (Escriou et al., 1996).

When comparing descriptions of variates used in different information systems, we observe a lot of common features together with a number of differences, the nature of which may be interesting to analyse and to discuss.

Firstly, there are a lot of small and useless differences. They are not very important by themselves, but their number proved to be an important obstacle, and makes the harmonisation between information systems impossible. If use was and will be made of published data dictionaries it is easier to build new information systems, and exchange of data will be possible and the difference between the various systems will be insignificant.

Secondly, analysis showed that the variates in different information systems were "broken down", in varying degrees e.g. the variate "dry matter" may exist as such, but also as the variate "dry matter of leaves at stage F1 in kg/m2" and as "dry matter of leaves at stage F1 in kg/m2 per plot". In some extreme cases, a variate aggregates a lot of information, and describes : a phenomena, at a given stage of a crop, observed on a given organ, at a given level (e.g. treatment), using a specific unit or scale ... The degree of aggregation is the origin of incompatibility between information systems.

Thirdly, there are very different data models which are more or less "spread out". A common feature of all analysed information systems is that they consist of a lot of tables. This feature seems related to the need to take into account the real structure of data and shows that exchange of data is feasible with a normalised database.

Fourthly, there is a number of inconsistencies or ambiguities which can be solved only when all methods of observation / measurement / sampling are precisely defined. For example, the scales of notation, the limits (min / max) may not correspond from one system to another one.

3. Data dictionary of variates

The first step on the way to developing a uniform data dictionary was to describe the datasets. French organisations had done work in this direction (ACTA-AGPM-CETIOM-ITB-ITCF). The intention of this work was to document actual systems in use. This work for instance gave an actual account of the state of the art in France. However not all items were well defined. Therefore an attempt was made to improve the description of the variables and to describe them in a uniform manner. This has been done for sugar beet, barley and oilseed rape. An example of such a description is given in table 1.

Table 1 An example of a description of a variate in sugar beet experiments.

ADDA_ID	BETA_NFWR
English name	Yield of roots
Dutch name	Wortelopbrengst
Danish name	Rod
French name	Poids de racine
Spanish name	Peso de raices
Eng->ADDA	1
Dan->ADDA	0(not recorded in a comparable way)
Fre->ADDA	1
Net->ADDA	1
Spa->ADDA	1
Dimension	kg/plot
Datatype	real5.1
Default values	
Range	(0 - 12) * plotsize
Definition	The Net fresh Weight of the harvested Roots without dirt and properly topped
Observation method	The surface of the plot is harvested by a trial harvester. A sample is collected and weighed in the laboratory. There the beet roots are cleaned and properly decapitated and weighed again. The dirt, head and net beet are weighed separately
Calculation method	The ratio net to gross weight of the sample and the total gross weight of the roots are used to calculate net fresh weight of roots.
Comments	

In this description information of different types is included. Data about the conversion of the value and name of a variate between countries, technical data relevant for control on validity at registration, the definition and how to make the observation and calculate the data.

Along these lines data dictionaries for three crop were constructed. Analysis of these descriptions showed that for many variates the same or similar attributes were used. For instance the observation method yield by trial harvester was used many times. Small differences in description could occur. However if such a method is described in a standard way and referenced, it is much easier to conclude that the same observation method is used. Therefore the data dictionary was normalised meaning that all attributes were classified and put in standard tables. The description of variates was done by reference to these standard tables.(table 2) In this way each variate is determined in a unique and traceable way.

This data definition should be used to exchange data between partners. The data definition is a common standard, locally a particular description might be in use. Therefore a conversion of the name and sometimes of the value will be needed because of the use of another dimension (table 3). However it is also feasible that a conversion is not possible because the definition is different or the method of measurement is different. For instance in Denmark the root yield is inclusive of top tare in contrast to the other countries where root yield is without top tare.

Table 2 Example of normalised description of a variate

ADVA_00007	Net weight topped roots sample	
Definition	Weight of the harvested roots sample cleaned and topped	
Purpose	To determine tare and net weight	
Subj. Observ	ADSO_00003	storage root
Type Observ	ADTO_00010	fresh weight sample
Time Stamp	ADTS_00130	harvest
Meth. Sampling	ADMS_00002	sub-sample
Meth. Observ.	ADMO_00004	weighing
Calculation	ADCA_00000	none
Dimension	ADDI_00021	weight

Table 3 Conversion table for the variate root yield.

ADDA_ID	ADVA_00111
English name	Yield of roots
Dutch name	Wortelopbrengst
Danish name	Rod
French name	Poids de racine
Spanish name	Peso de racine
Eng->ADDA	1
Dut->ADDA	0.1 from quintal --> tonne
Dan->ADDA	not possible
Fre->ADDA	1
Spa->ADDA	1

4. Discussion

The advantage of normalisation is that it is necessary to define exactly how a property is measured and stored. One has to define and describe lists of attributes. Because of these descriptions conversion of data is a relatively simple task. A good description facilitates the exchange of data but is valuable too for the exchange of research tasks between researchers.

The maintenance of the norm tables needs special attention. In particular the entry of a new attribute or the change of the description of an attribute should be done at a central institute to maintain compatibility. Co-ordination is needed to maintain compatibility. But once implemented the exchange of data will be facilitated and mutual gain could be realised. In the evaluation of varieties a larger number of trials will increase the accuracy and a possible interaction with environment could be determined.

5. Literature

Escriou H., X. Le Bris, H.Thoré and G. Waksman, 1996. Two proposals for overcoming limitations of classical data dictionaries. In: Information and Communication Technology applications in Agriculture: state of the art and future perspectives, Ed.:Lokhorst, C., A.J. Udink ten Cate, & A.A. Dijkhuizen, Agro-informaticareeks Nr. 10 Wageningen. pag: 267-271

Dindorp, U., M. Flensted-Jensen & A.P. Ravn, 1996. A standard Format for data exchange- SFD. In: Information and Communication Technology applications in Agriculture: state of the art and future perspectives, Ed.:Lokhorst, C., A.J. Udink ten Cate, & A.A. Dijkhuizen, Agro-informatica reeks Nr. 10 Wageningen. pag: 368-373

Kristensen, J.T, M. Gudmand-Høyer & L.Thygesen, 1993. SFD Standard Format for Data Exchange. User's Manual. Dina Publication series no. 2