E

UPOV/DATA/KYI/00/1
ORIGINAL: English
DATE: May 17, 2000

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

WORKSHOP ON DATA HANDLING

organized by
the International Union for the Protection of
New Varieties of Plants (UPOV)
and
the State Commission of Ukraine for Testing and Protection
of Plant Varieties

Kyiv, June 9 and 10, 2000

ABSTRACTS

Document prepared by the Coordinator of the Workshop

UPOV Workshop on Data Handling

Arranged by TWC

Kiev, June 9 and 10, 2000

List of abstracts

Design and Layout of an Experiment

JULIA BORYS

Research Centre for Cultivar Testing, Słupia Wielka, Poland

1. Introduction

The design and layout of an experiment play an important role in the testing for distinctness, uniformity and stability. The aim of DUS trials is to achieve an optimum expression of the phenotype (characteristics) of the variety as a technical basis for the grant of protection and to establish a reliable description for each variety. It is essential to establish the general rules and recommendations for conducting the technical examination and define the best experimental conditions.

2. General rules applied for DUS testing

- The experimenter should have a good understanding of the aim of the test, a good knowledge of growing conditions for the species, so as to minimise the interaction between genotypes and the environment to avoid changes in the classification of varieties for the various characteristics.

- The testing must be performed at the identity sample. For better assessment of stability, seeds delivered by the applicant in different years can be compared.

- The tests normally should be conducted at one place.

- The minimum duration of tests should normally be two similar growing periods.

- The tests normally are performed according to guidelines developed by UPOV. In the case of the crops, where no UPOV guidelines exist, a full botanical description is made in the same order as the UPOV guidelines.

- All observation for assessment of distinctness, uniformity and stability should be made on the number of plants or parts of plants according to the guidelines for each crop.

- Varieties are grouped according to their grouping characteristics mentioned in the appropriate UPOV guideline. In this way the most appropriate varieties are compared with each other in one and the same trial.

- Varietal comparisons normally are undertaken using a combination of the following methods:
  - direct visual comparison
  - population or plot scores against a prearranged scale
  - observations of single plants.

3. Different types of trials

The type of trial depends on the species and guidelines. The propagating mode of the species (allogamous, autogamous, vegetatively propagated) and the structure of the varieties species (pure-line varieties, clone, hybrid, population variety, synthetic variety) determine the design.

There are carried out such types of trials as:

- spaced plant trials – for some horticulture crops (for ornamental plants very often not randomised trials);

- spaced (individual) plant trials – generally used for allogamous species (grasses and some forage crops); 60 plants per variety grouped in three to six replications of 20, 15, 12 and 10 plants. More replications are generally more efficient when fewer varieties are included in the test. The varieties are randomised within each block and subgroup. Characteristics should be measured on each plant in the trial so that a mean value per plot can be obtained;

- row plots – generally for some agricultural crops (e.g. cereals, legumes oil crops); each test should consist of rows arranged in two or more replications. The size of the plots should be such that plants or parts of plants may be removed for observation without prejudice to the visual assessment which must be made up to the end of the growing period. The number of plants or parts of plants depends on the guidelines for a crop (e.g. 20 plants for wheat). In the case of refreshing the identity sample, the new sample can be compared with the old sample in the row trials.

- ear-rows plots are used for the assessment of uniformity of characteristics on single ear-rows, plants or parts of plants (visual observation of a number of aberrance in e.g. cereals, legumes). The distance between rows and between plants within the rows should be adequate to enable observation on individual plants.

4. Trial designs

Trial designs should be set up in that way they will enable the data to be processed and statistically interpreted. The main statistical tool used is the analysis of variance. So the basic assumptions of such analysis must be fulfilled. One of such assumptions is randomisation. Trial designs are developed to optimise the expression of phenotype, especially discriminating characteristics. The design generally follows the randomised complete blocks with two or more replications and full randomisation within each block. Plots of compact shape are commonly used in experiments in randomised blocks.

The experimental area is divided into some compact, or approximately square, blocks, and that each of these is divided into plots running from end to end of the block, and lying side by side.

In order to compare certain varieties directly, variety couples are occasionally planted systematically in each replication. The reference varieties are included in the trial and compared side by side with the relevant candidate variety. The number of reference varieties in the trials may be limited by grouping or pre-selection.

Because of high number of varieties being compared in DUS trials it is expected that incomplete block designs can be more effective than complete ones.

Recording of Characteristics

Beate Rücker,

Bundessortenamt, Hannover, Germany

All observations of the expression of characteristics aim at the assessment of distinctness, uniformity and stability and the variety description. The dataset to be recorded depends on the demands of these aims.

The expression of characteristics can be assessed by visual observations or by measurements (including counts) each with assessment of single plants or plots. The appropriate method to assess a characteristic is determined by different factors. These are for example:

- Type of characteristic: possibility for visual assessment vs. possibility for measurements with acceptable efforts (length, shape, color, weight, date ...)
- Variability between varieties (range of the collection)
- Variability within varieties caused by genotype and/or environment
- Number of varieties in the collection
- Specific local conditions (equipment, stuff)
- Relation of expense and profit

In order to obtain reliable and comparable results in various member States, recommendations are given in the Technical Guidelines for plot size, sample size, number of replications and duration of tests. Usually, qualitative characteristics are assessed visually, while quantitative characteristics are measured. However, visual assessment may also be applicable for quantitative characteristics under certain conditions. If visual observation fulfils all requirements for DUS assessment it should be preferred because usually visual observations are quicker and cheaper than measurements.

Practical aspects of measurements will be demonstrated in the workshop with examples from different species:

- One measured record per plot (e.g. plant length in a self pollinating crop)
- Measurement of single plants (e.g. plant length in a cross pollinating crop)
- Determination of flowering date for plots or single plants
- Use of image analysis for measurements (e.g. size of cotyledons or petals)
- Measurement of bulk samples vs. single plants (e.g. thousand seed weight)

For measured characteristics the assessment of distinctness, uniformity and stability is usually based on the measured records, while the measured records are transformed into notes (qualitative scale) for the variety description.

If a characteristic is assessed visually the observations are coded by notes. The notes should be used as fixed in the Technical Guidelines. In order to adjust the meaning of the states of expression over different tests UPOV recommends the use of example varieties. For quantitative characteristics the crop expert has to "calibrate" his own scale of notes according to the recommended notes and to adopt it to the specific situation he is faced (e.g. differentiation in the collection, environmental influence).

Practical examples will be presented concerning:

- Visual observation of a group of plants (e.g. most characteristics in self pollinating or vegetatively propagated crops)
- Visual observations of single plants (e.g. growth type in cross pollinating crops)

Data validation

Erik A. Lawaetz

Department of Variety Testing, Danish Institute of Agricultural Sciences, Denmark

DUS-testing of agriculture varieties

To establish distinctness in the DUS-testing we need variation between varieties. But is the variety variation you do see between the replications or within the plot at a normal level? Is it an ”off type” plant or do we have some kind of error? To find and recognise these different types of variation we need *data validation* and the experience from the cropexperts to understand the sources and the level of the variation. It is very important to do the *data validation* during and immediately after the observations of the particular characteristic. In order to be able to revisit the field when plants still are at the same stage. To handle and make the *data validation* computers and statistical programs are essential tools for the cropexpert. Data from the species *Spring barley* and *Spring rape* are used to show some examples of *data validation*.

BEFORE THE ACTUAL DATA VALIDATION

What to look for: (possible errors)

1. Trialplan
2. Bagging and setting the bags in the right seeding order
3. Trialseeding
4. Test Guideline characteristics

HOW TO DO IT: (FIND AND VERIFICATION OF ERRORS)

1. Check plotkey (randomisation)
    - variety code
    - other codes
2. Seed chamber
    - unambiguous and clear label for the variety
    - only one bag open at the time
3. Keep the seedbags in the order the trial has been seeded
4. To be sure that the characteristics of the Test Guideline are understood and explained properly

THE ACTUAL DATA VALIDATION

What to look for: (possible errors)

Typing errors, variety variation between replications and within plots, ”off types”
- Systematically/methodically errors
    - Plot displacement of the characteristic scores
    - Misunderstanding of Test Guideline
    - The time for harvesting the characteristic

How to do it: (find and verification of errors)

Spreadsheet methods - qualitative characteristics

Spring Barley 1999, Growth habit

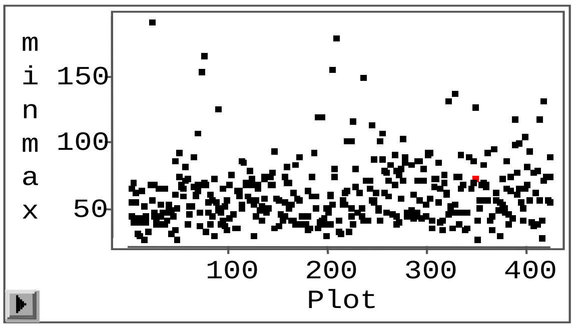
Variet y-		1 REP	2 REP	1 REP	2 REP	AVG	
code	kat99	plot	plot	___	___	_____	MINMA X
18150	36	86	345	6	7	7	1
18151	31	183	376	5	5	5	0
18152	31	108	468	6	4	5	2

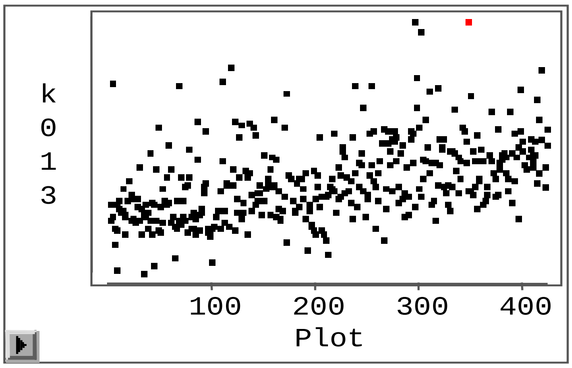
The variety sorted in order of varietycode or plotnumber in order of 1st, 2nd, or 3rd replication. Maximum minus minimum scores (minmax) and average for the variety are used to find typing errors and high difference of scores between replications. When a variety has a difference in scores greater than or equal to 2 we revisit the field in order to validate the observation. Sorting the table by 1st, 2nd or 3rd replication you are able to find a displacement of scores – if two or more plots next to each other have high minmax scores there can be a displacement of scores. In this case revisiting the field is necessary. The data validation is a good feedback for the cropexpert in connection to quality of performed work and keeping him updated during the growing season.

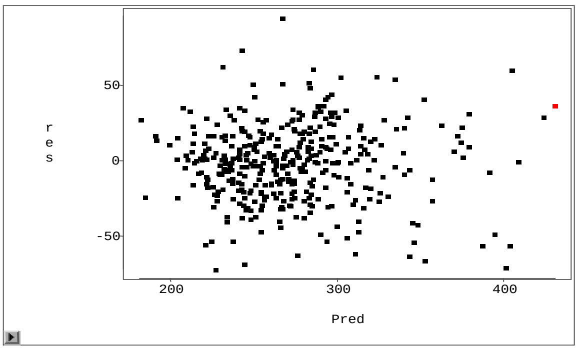
Graphical methods - quantitative characteristics

For DUS-testing of Spring rape Guideline TG/36/6 are used. The Spring rape characteristic length of 6th leaf harvested at stage 23-27 is used as an example. All 3 figures show the same data of the characteristic 'leaf length' (K013) in three different ways. The top figure shows the minmax recorded scores graphically. In our system you can trick on the observations and get the varietycode and at the same time the particular observation will be indicated in the other figures as well. The figure in the middle clearly shows that the 'leaf length' (K013) are getting larger from plot 1 to 426. In this case a systematically/methodically harvesting trend occurs which is not bad as long each replication are reasonably homogen.

The bottom figure shows the plot of the residual versus predicted values of 'leaf length'. This is a way to check the homogeneity of the variation. The data should be even distributed around zero.







IDENTIFYING SIMILAR VARIETIES

Wiesław Pilarczyk

Agricultural University of Poznan and Research Centre for Cultivar Testing at Słupia Wielka, Poland

1. Introduction

New varieties of many species are more and more similar to each other. So experimenters and institutions responsible for registration new varieties are deeply interested in identifying pairs of varieties that are too close to each other. In some cases mathematical measures of similarity based on characteristics of different types can help in detecting such pairs of varieties.

2. Measures of similarity between varieties

2.1 Experimenter eye
2.2 Mathematical measures of similarity
  Measures of similarity for continuous characteristics
  Measures of similarity for quantitative characteristics
 General measures of similarity of multidimensional objects
 Threshold values for coefficients of similarity
2.3 Advanced (computer aided) measures of similarity
  Measures of similarity based on electrophoretic data
  Computers and digitalisation of pictures

3. Supporting software

DUST package, General Statistical Packages (GENSTAT), VISOR, National solutions

Literature

Gower J.C., 1971, A general coefficient of similarity and some of its properties, Biometrics 27, pp.857-874
Heijden Gerie van der, 1999, Flores: A pictorial database for ornamentals using JAVA, TWC/17/5
Heijden Gerie van der, 1991, Calculating similarities between varieties using electrophoresis data, TWC/9/7
Heijden Gerie van der, 1994, Image analysis in variety testing, TWC/12/6
Heijden Gerie van der, 1998, Digital images in plant variety testing, TWC/16/11
Pilarczyk W., Application of Gower’s similarity coefficient to detect most similar varieties, TWC/14/2
Sokal R.R., Sneath P.H.A., 1963, Principles of numerical taxonomy, Freeman & Company
Weatherup S.T.C., 1992, Multivariate distinctness criterion, TWC/10/4
Weatherup S.T.C., 1995, Evaluation of most similar variety, TWC/13/6

Combined-over-years Distinctness Criterion (COYD)

John Law

National Institute of Agricultural Botany, United Kingdom

Document TC/33/7 gives details on the COYD Analysis.

To distinguish varieties on the basis of a measured characteristic we need to establish a minimum allowable distance between varieties so that a pair of varieties showing a difference greater than the minimum might be regarded as “distinct” in respect of that characteristic. There are several possible ways of establishing minimum distances from Distinctness, Uniformity and Stability (DUS) trials data. UPOV has agreed on what is known as the Combined-Over-Years Distinctness (COYD) criterion.

The COYD method involves:

– for each characteristic, taking the variety means from the two or three years of trials for candidates and established varieties and producing over-year means for the varieties;

– applying the technique of analysis of variance to the variety-by-years table in order to calculate a least significant difference (LSD) for comparing variety means;

– if the over-years mean difference between two varieties is greater than the LSD then the varieties are said to be distinct in respect of that characteristic.

The main advantages of the COYD method are:

– it combines information from several seasons into a single criterion in a simple and straightforward way;

– it ensures that judgements about distinctness will be reproducible in other seasons; in other words, the same genetic material should give similar results within reasonable limits from season-to-season.

– the risks of making a wrong judgement about distinctness are constant for all characteristics.

Document TC/33/7 describes:

– the principles underlying the COYD method;

– details of ways in which the procedure can be adapted to deal with special circumstances;

– UPOV recommendations on the application of COYD to individual species;

– the computer software which is available to apply the procedure.

The COYD method aims to establish for each characteristic a minimum difference, or distance, which if achieved by two varieties in trials over a period of two or three years, it should be possible to say that those varieties are clearly distinct with a specified degree of confidence.

The method uses variation in variety expression of a characteristic from year-to-year to establish the minimum distance. Thus, characteristics which show consistency in variety ranking between years will have smaller minimum distances than those with marked changes in ranking.

Calculation of the COYD criterion involves an analysis of variance of a variety-by-year table of means for each characteristic. Data for all candidate and established varieties which appeared in trials over the two or three years are included in the table.

COYD is recommended for use in assessing distinctness of varieties

– when observations are made on a plant (or plot) basis over two or more years;

– when there are some differences between plants (or plots) of a variety but, nevertheless, this variation is sufficiently small to allow us to distinguish between varieties;

– in general COYD is recommended for use in the testing of allogamous (cross-fertilized) varieties.

A pair of varieties is considered to be distinct if their over-years means differ by more than the COYD LSD in at least one characteristic.

It has been agreed to operate the COYD LSD at the 1% level for grass species for both two and three-year tests. Experience with spring onion has shown that a 5% level may be appropriate (Laidig 1988) and with leek the 1% level has been found to be acceptable (van der Heijden and van Marrewijk 1989).

Statistical methods for Uniformity tests

Sylvain Gregoire

GEVES, La Minière, France

Article 8 of the UPOV Convention reads as follows:

"The variety shall be deemed to be uniform if, subject to the variation that may be expected from the particular features of its propagation, it is sufficiently uniform in its relevant characteristics."

That means that the absolute level of uniformity required for vegetatively propagated varieties, truly self-pollinated varieties, mainly self-pollinated varieties, inbred lines of cross-pollinated varieties, cross-pollinated varieties, mainly cross-pollinated varieties, synthetic varieties and hybrid varieties is bound to be different.

Nevertheless despite of these many different situations **two main statistical methods are recommended by UPOV for Uniformity assessment**.

- In cases (crops, or within a crop depending of the genetic structure and/or way of propagation) **where it is assumed that all the plants of a given variety should be very much alike the detection of off-types is recommended**.

This can be applied to vegetatively propagated crops, lines and cereals for instance.

A general overview of this method can be found in TC/34/5

- In cases **where it is assumed that it exist variation from plant to plant within a given variety Uniformity is checked against the uniformity level of existing varieties of the same group.**
-

This can be applied to many forage crops for instance.

A general overview of this method can be found in TC/33/7

Usually the same trials are used to assess Distinctness, Uniformity and Stability. The same data collected in the field are used for both computations on Distinctness and Uniformity in the DUST program.

As a general policy UPOV recommend to use the same set of characteristics for Distinctness AND Uniformity.

- The more characteristics used, the more easy it is to distinguish on at least one characteristic, the more difficult it is to be uniform for each of the characteristics observed.

- On the other hand if too few characteristics are used, uniformity will be easier to be passed, but distinctness will not be easy to assess.

So countries have to establish an efficient selection of the characters they use from the UPOV Guide line of the crop, and National characters. The necessary balance between the ability to detect for Distinctness and Uniformity mentioned above is of course only part of the things to look at, when establishing which characteristics are to be used. Need for description purposes, efficiency for distinctness, cost of measures, etc…are some of the elements to consider.

The two types of methods (off-types and relative uniformity) will be presented with the following points:

- Principle of the method
- Way to proceed
- Practical example
- Elements for further deepening

Methods to be aware about when storing data

Uwe Meyer

Bundessortenamt Hannover, Germany

Data from measurements and visually assessed characteristics can be recorded electronically by some kinds of mobile data logger or in traditional paper form. Additionally it may be necessary to transfer some data from the personal computer to the program on the handheld computer (field plan with the order of the varieties for example) in order to control the recording process. The data structure is usually described by following variables:

- crop and variety
- year and location
- block, plot and plant
- characteristic

There are many different systems in UPOV member states for storing of data. Two frequently used ways are:

- storing of data in files for each stage of evaluation
- storing of data at first in files for transfer and then in a database for all stages of evaluation

Often it is necessary to use all available data from single plants. For further computations storing of averages, variances and other statistical parameter is almost sufficient.
In some cases it is important to have special routines to handle 'missing values' and always it is necessary to have a definite code for that kind of data. 'Zero' is generally not the method of choice for coding of 'missing values'.

In addition it is necessary to know the types of characteristics used in trial and the types of scales applied. One should distinguish between visual observed data and measurements, between qualitative and quantitative characteristics, and between nominal, ordinal and cardinal scales.

Besides the original trial data some descriptive information about the characteristics should be stored such as:

- short and long form of the description in at least two languages
- states of expression for the variety description
- input and output format
- type of characteristic and formulas for calculation if necessary (for automatic data processing)
- and other information

Usually databases or files are stored on a PC or on a special server. All calculation procedures need an interface between the program applied and the stored data. This interface may be:

- a direct access to files
- a direct access to a database (native connection)
- a general access to a database by using special interfaces (ODBC, JDBC,...)

Depending on the infrastructure you have to build up your own backup strategy. You should make sure to be able to recover completely all your data under all circumstances.

Available software for performing the tests and **How to run the DUST software**

Sally Watson , Biometrics Division, Department of Agriculture and Rural Development, Belfast, Northern Ireland
and
Adrian Roberts, Biomathematics and Statistics Scotland, University of Edinburgh, Scotland

Distinctness and Uniformity Software: The software available for performing distinctness and uniformity tests will be reviewed.

Using the DUST software:

OVERVIEW OF DUSTNT

DUSTNT is the Windows version of the software DUST, which helps manage and analyse data from DUS spaced plant variety trials for cross pollinating crops - such as herbage and vegetable crops.

Data are recorded on the DUS trials several times during the growing season. On each occasion a different set of characters are measured on all the plants. The data is stored in data files.

The DUSTNT software comprises a series of modules or programs linked by data files. DUSTNT allows the user to accumulate the data over the growing year, summarise and analyse them (see Fig. 1).

As the data are progressively accumulated, summarised and analysed, the nature of the data stored in the resulting data files changes, e.g. from measurements on individual plants to plot means to variety means. There are 10 different types of data file in the DUSTNT system depending on the sort of data the file contains. The principal types are:

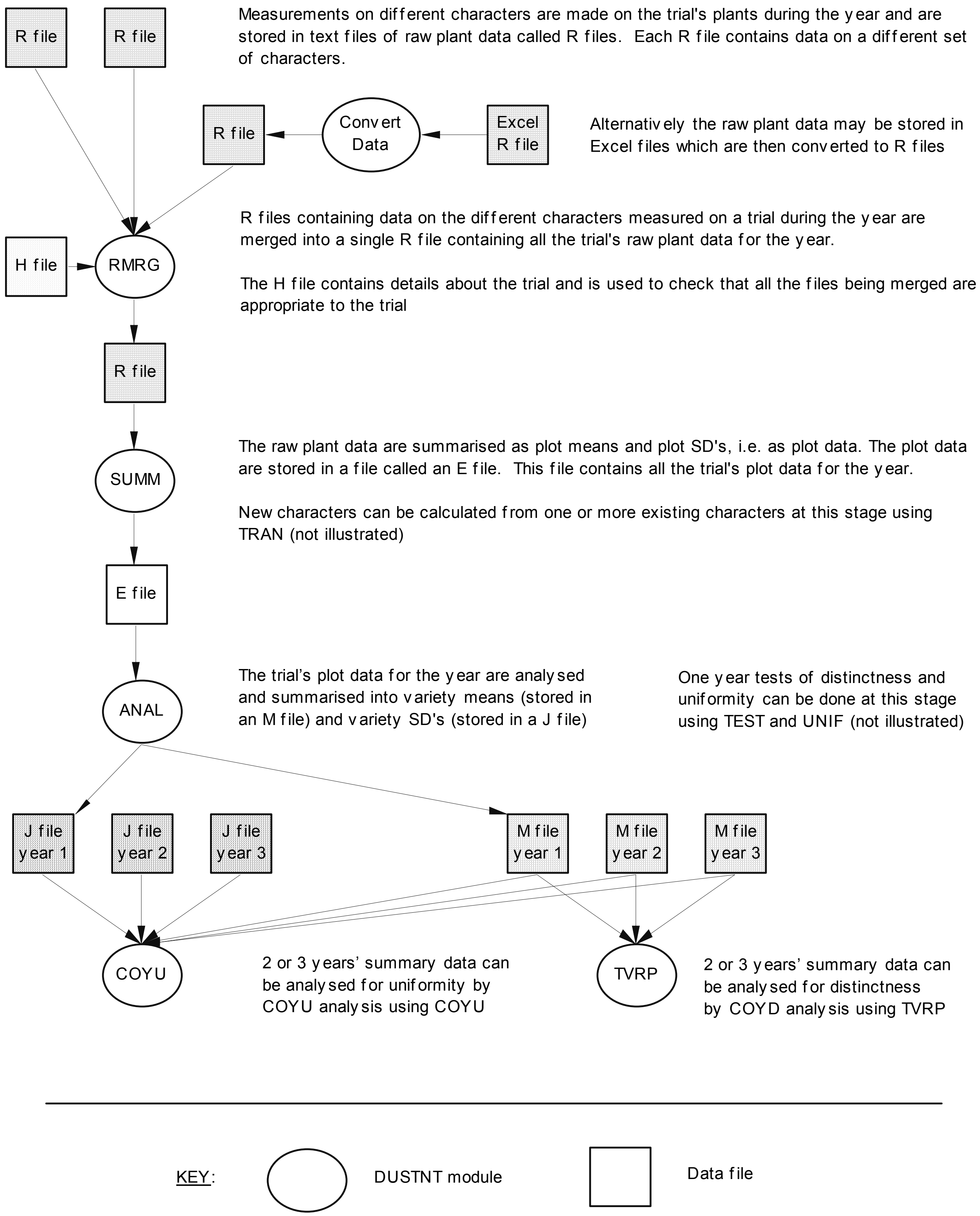
Type of file (first letter(s) of file name)	**Data the file contains**
H	Trial details (trial title, variety & character names)
R	Raw data on individual plants
E	Plot means accumulated over the growing year
M	Variety means
J	Variety within plot standard deviations
TX	Information for distinctness analyses
UX	Information for uniformity analyses

In order to accumulate, summarise and analyse a trial's data both in a single year and over years, great care must be taken to ensure that the correct data and information files are being used by the modules. The DUSTNT system uses two mechanisms to avoid mistakes, these are the *File Naming Convention* and the *Trial Number*.

Items to be covered:

- The flow of data; starting with measuring the plants in a trial and ending with COYU/COYD analyses
- The arrangement and identification of the data within the files
- Trial Numbers and File Naming Conventions
- Using DUSTNT to accumulate, summarise and analyse a trial's data for one year. This will include adjusting DUSTNT to suit your PC and viewing and printing files.
- Using DUSTNT to analyse a trial's data over 2/3 years (including COYD and COYU)
- Using Excel to create or alter DUSTNT data files
- Calculating new characters from existing characters
- Open forum session: an opportunity for participants to raise questions about their DUS problems.

Figure 1. The flow of data from plant measurement to COYD/COYU analysis using DUSTNT



[End of document]