AGRICULTURAL SYSTEMS

ELSEVIER Agricultural Systems 70 (2001) 477–492

www.elsevier.com/locate/agsy

Agronomic data: advances in documentation and protocols for exchange and use

L.A. Hunt[a,*], J.W. White[b], G. Hoogenboom[c]

[a]*Department of Plant Agriculture, University of Guelph, Guelph, Canada*

[b]*International Maize and Wheat Improvement Centre (CIMMYT), Ap. Postal 6-641, 06600 Mexico DF, Mexico*

[c]*Department of Biological and Agricultural Engineering, University of Georgia, Griffin, GA 30223, USA*

Received 30 March 2001

Abstract

Data from agronomy experiments are typically collected and stored in a number of minimally documented computer files, with additional information being entered and archived in field books or diaries. Data manipulation is generally cumbersome and error-prone, and data loss is frequent. Modern database technology has the potential to resolve these issues. However, experience gained by an international network of experimenters and crop modellers (the International Benchmark Sites Network for Agrotechnology Transfer; IBSNAT) in using a database for agronomic experiments conducted by many workers at different sites highlighted problems of data entry, quality control, and changing requirements for storage and output variables. In an attempt to minimize these problems, IBSNAT reduced its focus on a central database, but considerably enhanced its effort on the design and use of a set of simple, standard experiment documentation and results files that could be established and edited easily, transferred directly among workers, used as inputs to analytical software and crop models, and read by database and spreadsheet software. The standard files which were developed, and which were used in a software package termed DSSAT V3, have recently been upgraded by a consortium of experimenters and modellers (the International Consortium for Agricultural Systems Applications; ICASA). These new files are described briefly here. The ICASA files constitute an advance in the potential for good documentation and storage of agronomic data, but only partly solve the problem of overall data management and use. There is still need for central and local databases that facilitate both the searching of information from different experiments, and the examination of relationships that may be apparent in a large array of data. A number of such databases have been developed for specific applications,

* Corresponding author.
E-mail address: thunt@uoguelph.ca (L.A. Hunt).

PII: S0308-521X(01)00056-7

and a few of these are briefly touched upon. In particular, recent work with one large database currently being developed by a number of international Agricultural Research Centers, National Research Organizations, and Universities, (the International Crop Information System, ICIS), is briefly described.

Keywords: Agronomy; Data; File standards; IBSNAT; ICASA; Documentation; Minimum data set; Databases; ICIS

1. Introduction

For agronomic experiments, information detailing the planning, the sequence of activities and events during the conduct of the experiment, and the results, is commonly recorded first in notebooks, diaries, handheld tape recorders, or palmtop computers. Subsequently, this information may be transferred to prepared reporting sheets, to computer spreadsheets, or to computer files, and processed using statistical and other analytical software. The computer files so created may, in turn, be reorganized and updated such that it often becomes unclear as to which files contain original data, why changes to data were made, and which of the modified data are the most recent.

Faced with such a problem, some workers (e.g. Van Evert et al., 1999a, 1999b) have argued that the information stemming from agronomic experiments should be stored in a computerized database, and managed using a database management system (DBMS). The computerized database per se would be a logically coherent collection of computer files that contain all the data (Elmasri and Navathe, 2000), and the DBMS, software to simplify the creation and management of such computer files. The concept of using computerized databases and management systems in agricultural research, however, is not new. Andrews et al. (1978) and Andrews and Hardwick (1982) described databases for use in plant breeding before the advent of the micro-computer, and Mulitze (1990) documented a commercial package that also included analytical software. More recently, CIMMYT (Fox et al., 1997) developed a package entitled 'The International Wheat Information System' for storing and managing genealogical and selected performance data. This system forms the base of the 'International Crop Information System' (ICIS) that has received some attention in recent years (IRRI, 1997).

One of the first attempts to develop and implement database and associated management software for an international array of experiments was made by a US Agency for International Development (USAID) project, the International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT; Tsuji et al., 1998). The database, based on concepts developed by Nix (1984) and Wallach (1984), was documented in detail by IBSNAT (1991). It was implemented using commercially available database management software and was distributed in a software package called DSSAT V2.1 (IBSNAT, 1989). As part of the implementation process, data sheets were prepared and distributed to those wishing to make use of the software to create a local database, or to submit data to a central database (IBSNAT, 1986).

Experience in using the database and software both at individual locations and at a central site handling data from many co-operators, highlighted a number of problems. These included: double handling of data, quality control, and changing requirements for storage and output variables. Such problems, however, merely reflected one overriding problem — viz. the considerable and continuing need for human resources both to enter and check data, and to maintain and update the management software. Given that competition for financial resources, and the need for change, will always exist, IBSNAT concluded that the fundamental problem in reality was not one of resources but one of resource disposition — one of failing to direct sufficient resources to activities targeted at developing a system that fully recognized how agronomists work, think, and interact with each other and their clients, the farming community. Effort was thus shifted from one with database design and maintenance uppermost, to one in which improvement of the notes and files already used by agronomists became pre-eminent, and in which attempts were directed at improving the usability, interchangeability, and general understandability of such files.

Following this change of thrust, a set of simple, standard experiment documentation and data files that could be established and edited easily, that could be transferred directly among workers, and that could be directly used by analytical software and crop models was developed. These standard files, which were used in a software package DSSAT V3, were recently upgraded by a consortium of experimenters and modellers (the International Consortium for Agricultural Systems Applications; ICASA). These files, together with a description of experiences gained in using earlier versions, and with a description of efforts to develop databases and software for using agronomic information in a broader context, are described here.

2. First efforts — the IBSNAT files

As with recent attempts to develop protocols for information exchange between email communities, IBSNAT recognized that three things were necessary for meaningful agronomic information exchange and interpretation — a common set of required information, a common vocabulary for presenting information, and a common file structure and format that would facilitate understanding of the contents of any file. Attention was paid to each of these aspects when designing a set of files. First, considerable time and effort was devoted to defining a 'Minimum Data Set' (MDS) that would allow for a balanced documentation (and hence interpretation) of field experiments. The MDS consisted of: site and soil information, conditions at the start of the experimental period, crop and soil management details, and weather and agronomic data (Table 1).

The second requirement for easy information exchange and interpretation — a common vocabulary — was addressed by IBSNAT by defining one set of codes for variates measured before or during the conduct of an experiment, and a second set of codes for operations and inputs that could be carried out or supplied before or during the experiment. The codes were short, and thus somewhat indecipherable to

Table 1
Contents of the International Benchmark Sites Network for Agrotechnology Transfer minimum data set for experimental documentation

1. *Site and Soil*
   Latitude and longitude, elevation; average annual temperature; average annual amplitude in temperature
   Slope and aspect; major obstruction to the sun (e.g. nearby mountain); drainage (type, spacing and depth); surface stones (coverage and size)
   Basic profile characteristics by soil layer: in-situ water release curve characteristics (saturated, drained upper limit, lower limit); bulk density, organic carbon; pH; root growth factor; drainage coefficient
2. *Initial conditions*
   Surface layer(s) measurements of bulk density, organic carbon, organic nitrogen, pH, P, and K
   Previous crop, root and nodule amounts; numbers and effectiveness of rhizobia (if for nodulating crop)
   Water, ammonium and nitrate by soil layer
3. *Management*
   Cultivar name and type
   Planting date, depth and method; row spacing and direction; plant population
   Irrigation and water management, dates, methods and amounts or depths
   Fertilizer (inorganic) and inoculant applications
   Residue (organic fertilizer) applications (material, depth of incorporation, amount and nutrient concentrations)
   Chemical (e.g. pesticide) applications (material, amount)
   Tillage (e.g. seed-bed preparation)
   Environment (aerial) adjustments
   Harvest schedule
4. *Weather*
   Daily global solar radiation, maximum and minimum temperatures, precipitation
5. *Crop performance*
   Date of emergence
   Date of flowering or pollination (where appropriate)
   Date of onset of bulking in vegetative storage organ (where appropriate)
   Date of physiological maturity
   Leaf area index or percentage light interception at three stages during the life cycle
   Canopy height and breadth at maturity
   Yield of appropriate economic unit (e.g. kernels) in dry weight terms
   Canopy (above-ground) dry weight at maturity (and at two earlier stages if practical)
   Harvest product individual dry weight (e.g. weight per grain, weight per tuber)
   Harvest product number per unit at maturity (e.g. seeds per spike, seeds per pod)
   Damage level of pest (disease, weeds, etc.) infestation (maximum)
   Number of leaves produced on the main stem
   N percentage of economic unit
   N percentage of non-economic parts

a new user, but most were constructed according to a standard convention that facilitated understanding once a few simple rules were learned.

The third requirement — a common file structure and format that would facilitate understanding — was handled by introducing a set of standard ‘mark-up’ symbols and headings for different file sections, data items, and comments. The symbols used included an '*' for a section heading, '@' for a header line specifying variables occurring below, and '!' for a comment. Of these symbols, that for the header line was highly significant. Information on this line identified the variates that were placed below, and thus ensured that the files were virtually self-explanatory. The use of these symbols, along with a standard convention that data be organized into rows and columns, with data items separated by spaces, also made it possible to write software that could 'read' the file without specific format statements.

The standard files, which were used in DSSAT V3 (Tsuji et al., 1994) and described by Hunt et al. (1994b), were sufficiently flexible both to allow for use for primary data entry and storage, and to allow for the addition of variables deemed essential to document specific experiments. They were also largely self-explanatory, thus facilitating not only the exchange of data among users, but also easy interpretation by the originator after the elapse of some years.

2.1. Experience with the IBSNAT files

The standard IBSNAT files have been widely used by experimenters and modellers using the DSSAT system (Tsuji et al., 1994) and have been adopted by the Global Change and Terrestrial Ecosystem project (GCTE) for use in documenting experiments and regional yield investigations. Within GCTE, the standards have greatly assisted model comparisons (Goudriaan, 1996; Jamieson et al., 1998) that have led to model improvements. The formats have also been used for direct recording of information usually entered in a researcher's diary or fieldbook, (e.g. within the Ontario, Canada, Winter Wheat Co-operation Performance testing system). Experience gained in using the file structures and standards revealed, however, that they contain inconsistencies and ambiguities and are not entirely suited for generic use. In addition, the standards made insufficient provision for entering specifically headed comments.

Widespread use of datasets facilitated by the file standards has emphasized that datasets almost invariably contain errors, inconsistencies, and gaps despite the best efforts of experimenters. Any protocol for data storage must thus allow for the checking and/or verification of data items, for the documentation of any checking activity, and for the easy updating of data. To this end, a structure that allows for visual evaluation of all of the data is important, and the standard and columnar set-up of the IBSNAT files facilitated this. Once obvious errors have been eliminated, however, problems often remain, particularly in the following aspects:

2.2.1. Moisture contents

Grain yield reported without a clear indication as to whether it represents fresh weight (i.e. an 'as-is' basis) or dry weight.

2.2.2. Units

Measurements reported using non-standard units, without a clear indication as to the units and basis of measurement used. In some cases, this type of problem is easy to recognize, but in others (e.g. when yield data have been reported on a 'per plot' rather than 'per unit' area basis), the problem is virtually impossible to detect.

2.2.3. Conventions

The conventions used to define the aspects measured are not adequately documented. For example, leaf area index reported without any indication as to whether senescent (yellow or partially yellow) leaves or parts of leaves have been included, or whether the area of non-lamina structures (e.g. stems or spikes) has been included. Similarly, specific leaf area is often reported without an indication as to whether mid-ribs, petioles or petiolules are included.

2.2.4. Discrepancies

Yield component data (e.g. grain weight and number) do not match the overall yield data; canopy component data (e.g. leaf weight, stem weight) do not match with overall canopy weight; and end of season measurements often do not match with the final determination of a sequence of measurement made throughout the growing season. In many cases such discrepancies are explainable (e.g. different sampling techniques) and if so, this should be explained. If not, a user is unsure as to which measurement(s) are most reliable.

2.2.5. Manipulations

Sometimes, the data in a file have been adjusted without any indication as to the reasons for the adjustment. Such adjustments could occur when a lattice design and analysis was used, but more frequently occur when plots were damaged in some way (lodging, animal trampling, etc.), or some of the harvest product was lost or damaged. Such changes must be clearly documented in the data file.

In updating the file standards, therefore, particular attention has been paid to each of these aspects. Some consideration was given to directly associating all variates with their units, as advocated by Fowler (1997), but this was rejected as being too cumbersome for file use. Instead, the current approach, in which the name of a variate is associated with a unit, has been somewhat 'tightened'.

3. Recent efforts — the ICASA files

To minimize the problems identified when using the IBSNAT files, ICASA has devoted considerable time and effort to update the original IBSNAT files. The resulting standards (an example file is presented in Appendix) retain the basic structure used in the IBSNAT files. They use ASCII characters and have a line length restriction of 254 characters, with information arranged in columns that are headed by associated codes. Specific symbols are used to identify comments, missing data, software commands, and file endings, as summarized in Table 2. As with IBSNAT files, the files use 'keys' and character strings to link information in different parts of a file and in different files and are based on 'relations' that could easily be used in a large database containing information from many experiments.

Table 2
Summary of some of the syntax and formatting rules for International Consortium for Agricultural Systems Applications (ICASA) files

Item	Rule
Comments	A '!' at the beginning of a line indicates a comment.
Missing data	'-99', '-' or '.'
Non-applicable data	A '-99' also indicates data that are not applicable as opposed to missing per se. For example, row width and spacing for a crop that is broadcast sown
Data 'flags'	To indicate outliers, data filled with an estimate, etc. place a reference number in a flag column and detail the problem and changes in a FLAGS section (group) of the file
Commands	The symbol '#' at the beginning of a line indicates the start of a file statement that can be interpreted as a command by software (e.g. '#INCLUDE C:/ICASA/Newdata/FileA.dat')
Blank lines	May be placed anywhere in a file
End-of-file	'=' as the only character on the last line of a file. Its use is optional but it is recommended for electronic file transfer to indicate whether a file has unintentionally been truncated

The new ICASA files have data organized into a more clearly defined hierarchy than was present in the IBSNAT files. This hierarchy has three levels: DATA CLUSTERS; DATA GROUPS, and DATA UNITS.

The DATA CLUSTER is the most basic (smallest) aggregation, consisting of a number of variates and/or level indicators ('keys') clustered together under one line of codes. For example, a DATA CLUSTER could define a crop and cultivar that is to be used in an experiment.

The DATA GROUP aggregates all variates as levels of a particular type. For example, a cultivar DATA GROUP would be followed by all cultivars that are in the experiment.

The third and highest level of aggregation is the DATA UNIT. This level of aggregation was not present in the IBSNAT files. Its use allows a complete set of related data to be kept together. Examples of data that might be so aggregated are: (1) details of a particular experiment; (2) weather information from different sites within a region; (3) soil information for different sites from one institute; and (4) results of a particular experiment that involves several sites. These units can be placed in separate files or be aggregated into one OVERALL file to facilitate data transfer. The naming convention for the ICASA files follows that of the IBSNAT DSSAT V3 files.

The reader is referred to the ICASA web-site (http://www.icasanet.org/) to obtain a detailed description of the file structure as well as a listing of all codes and their definitions.

3.1. Overall data management

While the ICASA file standards represent a major advance in terms of the accurate documentation of trials and the facilitation of data use and transfer, they only partially address the broad problem of efficient management and archiving of agronomic data. As noted earlier, relational data base software has been proposed for use in agronomic data management. The experience of IBSNAT and, more recently, ICASA would suggest, however, that database software and the associated specific databases could best serve to facilitate the searching and ordering of selected items of information drawn from the ICASA files, the primary storage units, and to aid the examination of relationships that may be apparent when a large number of data sets are considered, particularly relationships between performance characteristics and genetic information. Such an approach would allow for the construction of computerized databases and associated software for particular and well defined purposes, a fundamental tenet of computerized database design. A number of specialized databases for use in particular contexts in agriculture have already been developed and described (e.g. Townsend et al., 1994; Haley et al., 1999).

The need to sort through datasets was not ignored in the second stage of the IBSNAT project, and database software was developed for extracting specific pieces of information from the standard files, and for sorting and searching through this information for experiments that meet desired criteria. The potential of this approach is being maintained and enhanced by workers within GCTE, who are developing software that will extract metadata directly from the ICASA files, and then allow for the searching and sorting of the metadata. The capture of specific pieces of information for sorting and searching in centralized databases may, however, be rendered unnecessary by the development of approaches that facilitate the searching of files on local computer servers. One such approach could involve the use of the 'Extensible Markup Language' (XML; see www.w3.org/xml/1999/xml-in-10-points). At the present time, however, the question as to whether resources in the long term would be better directed to the development of software that could translate ICASA files into XML, rather than to the development of software that can extract data from ICASA files, place such data into a centralized database, and then search through the information is still moot.

Over and beyond the question of how best to search and sort through information in different files, however, there remains the need to aggregate information from many sources for use when exploring overall relationships. A general database for agro-ecological research that could aid in such an endeavour has been described and implemented by Van Evert et al. (1999a, 1999b), and a second database that can store both genealogical and experimental information is at the core of the International Crop Information System (ICIS Project, 1999). This latter could be of particular use in the exploration of performance-genotype relationships.

The ICIS development has been based on the premise that to obtain the full benefits of a system for managing crop research data, the system should cover as broad a range of data as possible, potentially including information from biotechnology to plant breeding to farm-level agronomy (ICIS Project, 1999). Furthermore, recognizing that researchers may be reluctant to share all of their data, the system has been developed to permit storage of data both in local and central systems. It has been built on the experience of the International Wheat Information System (IWIS; Fox et al., 1997), but has taken more than 4 years of design effort to arrive at what appears to be a robust physical structure for both genealogy and data management systems (ICIS

Project, 1999). The Genealogy Management System (GMS) handles out-crossing and self-pollinating crops as well as clonally propagated material and germplasm derived from specialized techniques such as mutation breeding, polyploidization and genetic transformation. The Data Management System (DMS) is used to store experimental details and crop-soil system performance data. The latter, which can be numerical or text, are linked to a trait management system that allows for different units and methods of measurement for a single trait. Supporting data on individuals, institutions, bibliographic references and locations are also readily stored.

The ICIS Version 1.0 CD-ROM (ICIS Project, 1999) contains implementations for nine crops (barley, common bean, chickpea, maize, potato, rice, soybean, triticale, and wheat) as well as a special implementation for farming systems data. This first release emphasizes genealogy data which, for wheat (the International Wheat Information System, IWIS), can be presented back for 18 generations. Data input is direct, but the ICASA standards are seen as key for data input and output for subsequent versions.

The development and implementation of a system as flexible and complex as ICIS has presented many challenges. Software development strives for modularization with reusable components, and ICIS developers have participated in efforts to specify object-oriented standards for agricultural decision support tools. They are involved to a degree with data entry and recognize, as did IBSNAT after DSSAT V2.1, that this is crucial to the success of a database. Their efforts in this area, as in the overall database and associated software design, parallel those of some major seed companies, who are developing in-house systems with similar scope. Such efforts reflect the consensus in the bio-informatics arena that effective data management will soon be a major challenge in crop improvement. However, efforts to develop extremely flexible and all encompassing systems may fall foul of the most basic tenets underlying effective database design and application — viz., the need for a requirements-driven approach. It is to be hoped, therefore, that subsequent developments will place most emphasis on 'adding value' to whatever is being recorded, handled, and analysed using existing systems and standards, prominent among which are the ICASA file standards, and to modelling the data for application in new and expanding areas of endeavour. Such a viewpoint is supported by the experience of a major agricultural research group (the Agricultural Production Systems Research Unit, APSRU, located in Toowoomba, Australia). This group had planned to develop a relational database of agronomic data for use with their crop modelling system, APSIM (McCown et al., 1996), but have recently abandoned efforts to develop a central database, leaving data organization to individuals with their own specific interests (Carberry, APSRU, pers. comm.).

3.2. Future needs

The ICASA experimental details file contains the names of the cultivars included in any given experiment and, as an optional item, general information on the nature of the cultivars. The data set specification, however, does not require detailed information on cultivar pedigree and characteristics. This is a major weakness insofar as the nature of the cultivars may well determine the overall performance characteristics of an experiment. ICASA should thus work in co-operation with the international Agricultural Research Centers and others towards the definition of an array of phenotypically invariant trait coefficients that could be used to characterize cultivars. Naturally, some of these characteristics would be species specific, and some of the definitions would become clearer as a result of the analytical work associated with crop modelling. However, any attempt to define invariant coefficients will inevitably involve a consideration of the genomics that underlie phenotypic expression. A pressing need, therefore, is for the development of databases in which performance information is aggregated with genomic information. The ICIS endeavour, or something with similar international orientation, could play a pivotal role in this domain.

An equally pressing problem stems from recognition that data defined for the current files, and used as input to a number of simulation models, have been inadequate to enable accurate simulation of the results of some experimental work (e.g. Landau et al., 1998; Van Oijen and Ewert, 1999). While there may be many reasons for such discrepancies, the possibility that the variates included do not encompass some aspects that are significant to crop performance should not be overlooked. Aspects that can be mentioned in this context are windspeed and air humidity (Vijaya Kumar et al., 1996), soil micro-nutrient content (e.g. Engel et al., 1998), soil-borne organisms (e.g. Collins Johnson et al., 1992), and soil compaction (e.g. Soane and Van Ouwerkerk, 1994; Lal, 1996). ICASA and others should thus continue work to refine and enhance standards for the documentation of experimental work, to ensure that results of field experiments can be effectively explained and realistically simulated.

Over and beyond the need for specific data to be added to the files, there is a need for tools to enter and manipulate data within the files, and for wider recognition and use of the standard file configurations. A number of such tools is available in the DSSAT V3 package, but with the introduction of the ICASA standards, these need modification and enhancement. Some such tools will be produced by those directly involved with ICASA, but as with all sets of standards, there will always be a need for additional and improved tools. It is hoped that many individuals not involved directly with ICASA will develop tools and share those with ICASA and others. As regards the need for wider recognition and use, it is perhaps best to let this develop as workers recognize the interchange opportunities afforded by adherence to a set of standards, much as with the Internet (Twomey, 2000). Some further publicity, however, could be achieved by bringing the ICASA standards to the attention of groups that foster and support the development of standards for computer applications. The Life Sciences Research (LSR) Group, which is open and non-profit making (http://www.omg.org/homepage/lsr/intro.html/), is one such group focusing on the definition of data objects.

4. Conclusions

The definition of a set of standards that can serve for experimental documentation, for data transfer, and for input to analytical programs, crop models, and specifically oriented databases has implications that extend well beyond the immediate issue of data handling. It has implications in crop modelling per se, where there is a recognized need for standardized inputs and outputs, both within the modelling community and among those who are not modellers themselves but are interested in the potential uses of models (e.g. Mutsaers and Wang, 1999). It also has implications that relate to the way that individual field experiments are conducted, on requirements for participation in joint experiments, and on requirements for publication of the results of experimental work. The contents of the set of standard files in essence constitute a 'check-list' that details what should be required of all those conducting field experiments. The work in developing the standards thus complements earlier efforts to define Minimum Data Sets for field experiments (e.g. Nix, 1984; Hunt et al., 1994a).

For the interpretation of data from multi-site trials, including plant breeding nurseries, the collection of the defined minimum set of background information at each site appears essential for all trials other than those distributed for disease virulence identification. For publication, both to obtain maximum utility from the experiments reported, and to allow for checking of conclusions, a minimum data set in standard form should be available to other workers. Publishing houses involved with field research should thus consider requiring the submission of a minimum data set in standard form along with the usual manuscript, and to posting this along with the manuscript on their electronic site. Naturally, password protection may be required in some cases, but the implementation of such limitations no longer presents problems.

Finally, we would like to re-emphasize the importance of a careful consideration of the files used by agronomists and plant breeders for initial data acquisition and note-taking, and of the value of standards that relate both to structure and to coding for such files. We contend that careful consideration of these aspects is important to allow for the rapid and yet structured addition of new variables at the prime data acquisition point, to facilitate data exchange amongst workers, and to simplify the implementation and continued maintenance of large agronomic databases, particularly those aiming at aggregating information from a wide array of collaborators. Without standardization, and without a design that allows for machine reading of the initial files, data verification and entry will inevitably become a bottle-neck demanding the time and effort of technicians and professionals beyond the reach of most if not all programs. The ICASA files represent an important step in the definition of standards for widespread use. They may need enhancement in the future, as some have argued (e.g. Van Evert, Plant Research International, pers. comm.), to allow for a more formal specification of the layout of an experiment, or of the measurements that were made. ICASA is taking steps to ensure that procedures are in place to formally upgrade its file standard, and will ensure that suggested enhancements are considered and introduced in a manner that does not compromise the simplicity and usability of the current files. Similarly, it may be necessary to develop a description of the files using formal notation. For the present, though, the description and demonstration files on the ICASA website (http://www.icasanet.org/) should serve for most applications, and these should be used widely by agronomists and plant breeders.

Acknowledgements

Many people have contributed to the development of the ICASA file standards, and we thank all of those who have made suggestions at one time or another. We particularly wish to thank Mr. D.W.G. Van Kraalingen of Alterra, Wageningen University and Research Centre, The Netherlands, who has devoted considerable time and effort to the files, and whose effort is reflected in a number of aspects. Dr. Frits Van Evert, Plant Research International, Wageningen University and Research Centre, The Netherlands, has also provided valuable insights, and we thank him for his contribution.

Appendix. Example of an ICASA experimental details file

$(2001)EXPERIMENT:KSAS8199WH Fictitious experiment based on KSAS8101

*GENERAL
@ NAME
Nitrogen response of wheat at 2 irrigation levels

@ MAIN_FACTOR	FACTORS	LOCAL_NAME
Nitrogen	4PL*2IR*3FE	Godwin

@ OBJECTIVES
To determine if response to N follows predicted pattern

@ PEOPLE	EMAIL	ROLE
Hunt,L.A.	thunt@uoguelph.ca	File maintainer
Wagger,M.G.	-99	Primary data collector
Kissel,D.	-99	Thesis supervisor

@ NOTES
File based on KSAS8101.WHX but with many modifications to illustrate useage.
Seed mixed for 2 plots in rep 1.
Some leaf flecking - could be micro-nutrients or virus or ?
@ PUBLICATIONS
Wagger,M.G., 1983 N cycling in the plant-soil system. Ph.D., Kansas State
@ DISTRIBUTION
Public
@ VERSION
12-02-2001(LAH,Guelph)

*TREATMENTS FACTOR LEVELS

@ TRNO	R#	C#	O#	S#	TREATMENT_NAME.........	CU	FL	SA	IC	PL	IR	FE	OM	CH	TI	EM	HA
1	1	1	1	1	0N,dryland + weeds Rep1	1	1	1	1	1	0	0	1	1	1	1	1
1	2	1	1	1	0N,dryland + weeds Rep2	1	1	1	1	1	0	0	2	1	1	1	1
1	0	1	1	2	0N,dryland + weeds	2	1	1	1	2	0	2	1	1	1	1	1
2	0	1	1	1	180N,dryland	1	1	1	1	1	0	3	1	1	1	1	1
3	0	1	1	1	0N,irrigated	1	1	1	1	1	1	0	1	1	1	1	1
4	0	1	1	1	60N,irrigated	1	1	1	1	1	1	2	1	1	1	1	1
5	0	1	1	1	180N,irrigated	1	1	1	1	1	1	3	1	1	1	1	1
6	0	1	1	1	180N,irrig(dap)	1	1	1	1	1	2	3	1	1	1	1	1
7	0	1	1	1	180N,irrig(automatic)	1	1	1	1	1	3	3	1	1	1	1	1
7	0	2	1	1	RedClover Post-harvest	3	1	0	0	3	0	0	0	0	0	0	0
7	0	2	2	1	RedClover2 Postharvest	4	1	0	0	4	0	0	0	0	0	0	0

*CULTIVARS

@ CU	CR	CUL_ID	CULTIVAR_NAME..........	CULTIVAR_NOTES
1	WH	IB0488	Newton	Hard red winter,Central Plains
2	GW	IB0001	Garytown	Central Gt.Plains quackgrass
3	RC	IB0001	Bigyield	General double cut red clover
4	RC	IB0002	Earlywonder	Single cut ecotype

*FIELDS

@ FL	FL_NAME		
1	Research park		
@ FL	FIELD_ID	WEATHER_ID	SOIL_ID
1	KSAS0001	KSAS	KSAS81IF1

*SOIL_ANALYSES

@ SA	SAYR	SADAY	
1	1981	260	
@ SA	SABL	SAPHW	SAPHB
1	15	5.85	5.21

*INITIAL_CONDITIONS

@ IC	ICRYR	ICRDY	ICRCD	ICRDP	ICRIP	ICRAM	ICRRT	ICRN%
1	1981	250	WH	10	50	2000	100	1.00
@ IC	ICYR	ICDAY						
1	1981	270						
@ IC	ICBL	ICH2O	ICNH4	ICNO3				
1	5	.205	3.4	9.8				
1	15	.205	3.4	9.8				
1	30	.170	3.2	7.3				
1	60	.092	2.5	5.1				
1	90	.065	2.2	4.7				
1	120	.066	2.7	4.3				
1	150	.066	2.7	4.3				
1	180	.066	2.7	4.3				

*PLANTING

@ PL	PL_NAME								
1	Early planting								
2	Late planting								
3	Planting 25 days after harvesting the previous crop								
4	Planting 45 days after harvesting the previous crop								
@ PL	PLYR	PLDAY	PLDOE	PLPOP	PLPOE	PLDS	PLRS	PLRD	PLDP	PLPH
1	1981	270	-99	162	162	PLD0R	17	90	5.5	1.0
2	1981	289	-99	40	40	PLD0R	17	90	3.5	1.0
3	-99	25	-99	400	400	PLD0B	-99	-99	0.0	1.0
4	-99	45	-99	400	400	PLD0B	-99	-99	0.0	1.0
@ PL	PLMA	PLMWT	PLMG%	PLMSOURCE						
1	PLM0S	120	97	University plant breeder						
2	PLM0S	20	97	University weed scientist						
3	PLM0S	5	97	University plant breeder						
4	PLM0S	10	97	University plant breeder						

*IRRIGATION AND WATER MANAGEMENT

@	IR	IR_NAME
	1	Irrigation on specified dates
	2	Irrigation on specified days after planting
	3	Automatic irrigation

@	IR	IREF	IRADP	IRATH	IRAEP	IRAOF	IRAOP
	3	1.0	60	90	85	80	IR001

@	IR	IRYR	IRDAY	IROP	IRVAL	IRDEP
	1	1982	96	IR001	65	0
	1	1982	110	IR004	78	0
	1	1982	117	IR005	70	20.0
	2	-99	170	IR001	65	0
	2	-99	200	IR001	78	0
	2	-99	210	IR001	70	0

*FERTILIZERS (INORGANIC)

@	FE	FE_NAME
	2	Single application at planting
	3	Applications at planting and in spring

@	FE	FEYR	FEDAY	FECD	FEACD	FEDEP	FEN
	2	1981	289	FE001	AP001	15	60
	3	1981	289	FE001	AP001	15	90
	3	1982	56	FE001	AP001	1	90

*ORGANIC_MATERIALS

@ OM	OMYR	OMDAY	OMCD	OMACD	OMDEP	OMINP	OMAMT	OMN%
1	1981	280	RE000	-99	10	80	1000	0.1
2	1981	280	RE000	-99	10	80	1500	0.1

*CHEMICALS

@	CH	CHYRCHDAYCHSTG		CHCD	CHACD	CHDEP	CHAMT	
	1	1981	300	-99	CH001	AP001	1.3	2.2

*TILLAGE

@ TI	TIYR	TIDAY	TIIMP	TIDEP	TIMIX	TILLAGE_NOTES
1	1981	250	TI005	15.0	30	Lea type plough with long moldboard.

*ENVIRONMENTAL_MODIFICATIONS

@ EM	EMYR	EMDAY	EMDYL	EMRAD	EMTMX	EMTMN	EMH2O	EMCO2
1	1982	180	0	0	0	0	0	R600
1	1982	220	0	0	0	0	0	R360

*HARVESTS

@ HA	HAYR	HADAY	HASTG	HACOM	HASIZ	HAPC	HABPC
1	1982	220	90	HAC0P	HAS0A	100	50

References

Andrews, D.J., Hardwick, R.C., 1982. A database management system for information retrieval and documentation of expeirments for plant breeders. Euphytica 31, 281–285.

Andrews, D.J., Hardwick, R.C., Hardaker, J.M., 1978. A computer-based retrieval system for plant breeding material. Euphytica 27, 849–853.

Collins Johnson, N., Copeland, P.J., Crookston, R.K., Pfleger, F.L., 1992. Mycorrhizae: possible explanation for yield decline with continuous corn and soybean. Agron. J. 84, 387–390.

Elmasri, R., Navathe, S.B., 2000. Fundamentals of Database Systems, 3rd Edition. World Student Series. Addison-Wesley.

Engel, R.E., Bruckner, P.L., Eckhoff, J., 1998. Critical tissue concentration and chloride requirements for wheat. Soil Sci. Soc. Am. J. 62, 401–405.

Fowler, M., 1997. Observations and measurements. In: Analysis Patterns: Reusable Object Models. Addison-Wesley, Menlo Park, California, pp. 35–55.

Fox, P.N., Magana, Lopez, C., Sanchez, H., Herrera, R., Vicarte, V., White, J.W., Skovmand, B., Mackay, M.C., 1997. International Wheat Information System (IWIS), Version 2. CIMMYT, Mexico, D.F. (CD-ROM publication).

Goudriaan, J., 1996. Predicting crop yields under global change. In: Walker, B.H., Steffen, W.L. (Eds.), Global Change and Terrestrial Ecosystems. International Geosphere-Biosphere Programme Book Series, Vol. 2. Cambridge University Press, Cambridge, England, pp. 260–274.

Haley, S.D., May, R.D., Seaborn, B.W., Chung, O.K., 1999. Relational database system for summarization and interpretation of hard wheat regional quality data. Crop Sci. 39, 309–315.

Hunt, L.A., Jones, J.W., Tsuji, G.Y., Uehara, G., 1994a. A minimum data set for field experiments. In: Uhlir, P.F., Carter, G.C. (Eds.), Crop Modelling and Related Environmental Data. A Focus on Applications for Arid and Semiarid Regions in Developing Countries. CODATA, International Council of Scientific Unions, Paris, France, pp. 27–33.

Hunt, L.A., Jones, J.W., Hoogenboom, G., Godwin, D.C., Singh, U., Pickering, N., 1994b. General input and output file structures for crop simulation models. In: Uhlir, P.F., Carter, G.C. (Eds.), Crop Modelling and Related Environmental Data. A Focus on Applications for Arid and Semiarid Regions in Developing Countries. CODATA, International Council of Scientific Unions, Paris, France, pp. 35–73.

IBSNAT (International Benchmark Sites Network for Agrotechnology Transfer), 1986. Decision Support System for Agrotechnology Transfer (DSSAT) — Level 1: User's Guide for the Minimum Data Set Entry, Version 1.1 (Technical Report 3). Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI.

IBSNAT (International Benchmark Sites Network for Agrotechnology Transfer), 1989. Decision Support System for Agrotechnology Transfer, (DSSAT) Version 2.1 (Technical Report 4). Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI.

IBSNAT (International Benchmark Sites Network for Agrotechnology Transfer), 1991. Documentation for the IBSNAT Data Base Management System for the Decision Support System for Agrotechnology Transfer (DSSAT). Version 2.1 (Technical Report 4). Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI.

ICIS Project, 1999. The International Crop Information system [Online]. Available at http://www.cgiar.org/icis/ (posted 24 December 1998; verified 16 May, 1999).

IRRI, 1997. ICIS Technical Development Manual, Version 1. International Rice Research Institute, Manila, Philippines.

Jamieson, P.D., Porter, J.R., Goudriaan, J., Ritchie, J.T., Van Keulen, H., Stol, W., 1998. A comparison of the models AFRCHWHEAT2, CERES-Wheat, Sirius, SUCROS2 and SWHEAT with measurements from wheat grown under drought. Field Crops Res. 55, 23–44.

Landau, S., Mitchell, R.A.C., Barnett, V., Colls, J.J., Craigon, J., Moore, K.L., Payne, R.W., 1998. Testing winter wheat simulation models' predictions against observed UK grain yields. Agric. For. Meteor 89, 85–99.

Lal, R., 1996. Axle load and tillage effects on crop yields on a mollic ochraqualf in northwest Ohio. Soil Till. Res. 37, 143–160.

McCown, R.L., Hammer, G.L., Hargreaves, J.N.G., Holzworth, D.P., Freebain, B.M., 1996. APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. Agric. Systems 50, 255–271.

Mulitze, D.K., 1990. Agrobase/4: a microcomputer database management and analysis system for plant breeding and agronomy. Agron. J. 82, 1016–1021.

Mutsaers, H.J.W., Wang, Z., 1999. Are simulation models ready for agricultural research in developing countries? Agron. J. 91, 1–4.

Nix, H.A., 1984. Minimum data sets for agrotechnology transfer. In: Kumble, V. (Ed.), Minimum Data Sets for Agrotechnology Transfer. ICRISAT, Hyderabad, India, pp. 181–188.

Soane, B.D., Van Ouwerkerk, C., 1994. Soil compaction problems in world agriculture. In: Soane, B.D., Van Ouwerkerk, C. (Eds.), Soil Compaction in Crop Production. Elsevier, Amsterdam, The Netherlands, pp. 1–21.

Townsend, M.D., Henning, J.A., Carrier, C.J., 1994. The alfalfa catalog software package. Agron. J. 86, 337–339.

Tsuji, G.Y., Uehara, G., Balas, S., 1994. DSSAT Version 3: A Decision Support System for Agrotechnology Transfer. University of Hawaii, Honolulu, HI, USA.

Tsuji, G.Y., Hoogenboom, G., Thornton, P.K., 1998. Understanding Options for Agricultural Production. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Twomey, P., 2000. How internet standards emerge. Red Herring, January 2000, p. 167.

Van Evert, F.K., Spaans, E.J.A., Krieger, S.D., Carlis, J.V., Baker, J.M., 1999a. A database for agroecological research data. 1: data model. Agron. J. 91, 54–62.

Van Evert, F.K., Spaans, E.J.A., Krieger, S.D., Carlis, J.V., Baker, J.M., 1999b. A database for agroecological research data. II. A relational implementation. Agron. J. 91, 62–71.

Van Oijen, M., Ewert, F., 1999. The effects of climatic variation in Europe on the yield response of spring wheat cv. Minaret to elevated CO_2 and O_3: an analysis of open-top chamber experiments by means of two crop growth simulation models. European J. of Agron. 10, 249–264.

Vijaya Kumar, P., Srivastava, N.N., Victor, U.S., Gangadhar Rao, D., Subba Rao, A.V.M., Ramakrishan, Y.S., Ramana Rao, B.V., 1996. Radiation and water use efficiencies of rainfed castor beans (*Ricinus communis* L.) in relation to different weather parameters. Agric. For. Meteor. 81, 241–253.

Wallach, D., 1984. The organization of agronomic experiment data for crop modeling. In: Kumble, V. (Ed.), Minimum Data Sets for Agrotechnology Transfer. ICRISAT, Hyderabad, India, pp. 147–154.