Computers and Electronics in Agriculture 96 (2013) 1–12

Contents lists available at SciVerse ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards

Jeffrey W. White [a,*], L.A. Hunt [b], Kenneth J. Boote [c], James W. Jones [d], Jawoo Koo [e], Soonho Kim [e], Cheryl H. Porter [d], Paul W. Wilkens [f], Gerrit Hoogenboom [g]

[a] USDA, ARS, US Arid-Land Agricultural Res. Center, 21881 N. Cardon Lane, Maricopa, AZ 85138, USA
[b] Dept. Plant Agriculture, Univ. Guelph, 50 Stone Rd. E., Guelph, Ontario, Canada N1G 2W1
[c] Agronomy Dept., Univ. Florida, PO Box 110500, Gainesville, FL 32611-0500, USA
[d] Dept. Agric. and Biological Eng., 243 Rogers, PO Box 110570, Gainesville, FL 32611-0570, USA
[e] IFPRI, 2033 K St., NW, Washington, DC 20006-1002, USA
[f] IFDC, PO Box 2040, Muscle Shoals, AL 35662, USA
[g] AgWeatherNet, Washington State Univ., 24106 North Bunn Rd., Prosser, WA 99350-8694, USA

ARTICLE INFO

Article history:
Received 3 August 2012
Received in revised form 27 March 2013
Accepted 1 April 2013

Keywords:
Climate change
Databases
Data management
Crop modeling
Software

ABSTRACT

Agricultural research increasingly seeks to quantify complex interactions of processes for a wide range of environmental conditions and crop management scenarios, leading to investigation where multiple sets of experimental data are examined using tools such as simulation and regression. The use of standard data formats for documenting experiments and modeling crop growth and development can facilitate exchanges of information and software, allowing researchers to focus on research per se rather than on converting and re-formatting data or trying to estimate or otherwise compensate for missing information. The standards developed by the International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) project and subsequently revised by the International Consortium for Agricultural Systems Applications (ICASA) were of considerable value for describing experiments. However, the resulting ICASA Version 1 standards did not consider important management practices such as tillage and use of mulches, lacked descriptors for certain soil and plant traits (especially related to nutrient levels), and contained minor logical inconsistencies. The ICASA standards have evolved to allow description of additional management practices and traits of soils and plants and to provide greater emphasis on standardizing vocabularies, clarifying relations among variables, and expanding formats beyond the original plain text file format. This paper provides an overview of the ICASA Version 2.0 standards. The foundation of the standards is a master list variables that is organized in a hierarchical arrangement with major separations among descriptions of management practices or treatments, environmental conditions (soil and weather data), and measurements of crop responses. The plain text implementation is described in detail. Implementations in other digital formats (databases, spreadsheets, and data interchange formats) are also reviewed. Areas for further improvement and development are noted, particularly as related to describing pest damage, data quality and appropriate use of datasets. The master variable list and sample files are provided as electronic supplements.

Published by Elsevier B.V.

1. Introduction

Efficient interchange of data among researchers, especially for use in simulation models and other decision support tools, requires use of a common vocabulary and strategy for organizing data. The agricultural research community increasingly encounters research problems that require interdisciplinary collaboration. Physiologists and molecular biologists work together to develop a better understanding of the genetic control of productivity-related traits. Agronomists, soil scientists and irrigation specialists combine efforts in order to increase the efficiency of crop water use. In such

Abbreviations: ACE, AgMIP Crop Experiment; AgMIP, Agricultural Model Intercomparison and Improvement Project; ASCII, American Standard Code for Information Interchange; DSSAT, Decision Support System for Agrotechnology Transfer; IBSNAT, the International Benchmark Sites Network for Agrotechnology Transfer; ICASA, International Consortium for Agricultural Systems Applications; JSON, JavaScript Object Notation; NASA/POWER, NASA Prediction of World Energy Resource; WISE, World Inventory of Soil Emission Potentials; XML, eXtended Markup Language.

* Corresponding author. Tel.: +1 520 316 6368.
E-mail address: jeffrey.white@ars.usda.gov (J.W. White).

0168-1699/$ - see front matter Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.compag.2013.04.003

collaborations, efficient data interchange is essential. Furthermore, as data increasingly find use in applications that can affect the livelihoods of producers and other stakeholders, there is an ethical imperative for researchers to allow other people to examine their data. Taken together, these arguments create a powerful consensus for the need to promote data sharing in agricultural research (White and van Evert, 2008).

Genomic data are widely available through publicly accessible databases (Blanchard, 2004; Jung et al., 2011), and daily weather records and soil profile data are increasingly available through the Internet. The International Research Institute for Climate Prediction (IRI) developed two daily weather data download sites (IRI, 2006 IRI, 2009). Similarly, the NASA Prediction of World Energy Resource web site (NASA/POWER; Stackhouse, 2012) includes an option for downloading modeled daily weather in the ICASA format. The "World Inventory of Soil Emission Potentials" (WISE) database developed by the International Soil Reference and Information Centre in The Netherlands has been formatted for crop model applications (Gijsman et al., 2007; Romero et al., 2012). Field research data, however, are seldom available through public databases. Furthermore, information that is accessible often lacks key details such as on soil initial conditions and the timing and amounts of irrigations. Although various initiatives have developed systems for reporting and storing data from field research (e.g., van Evert et al., 1999a,b; Bostick et al., 2004; McLaren et al., 2005), no single system is widely used.

Among the largest, sustained efforts to promote the use of standards in relation to field research was that of ICASA and one of its predecessors, the IBSNAT project. As early as 1983, IBSNAT began developing data standards for documenting field experiments. A key event was the International Symposium on Minimum Data Sets for Agrotechnology Transfer, with representatives from over 20 countries (ICRISAT, 1984; Uehara and Tsuji, 1998). The standards and their implementation in plain text ("ASCII" or American Standard Code for Information Interchange) files facilitated interactions among experimenters and simulation modelers (Hunt et al., 1994; Jones et al., 1994). The plain text format was implemented in the Decision Support System for Agrotechnology Transfer (DSSAT) software system (Tsuji et al., 1994; Jones et al., 2003; Hoogenboom et al., 2011) and was adopted by the Global Change and Terrestrial Ecosystem project (GCTE) for use in documenting experiments and regional yield investigations (GCTE, 1996). Within GCTE, the standards greatly assisted comparisons among simulation models (Goudriaan, 1996; Jamieson et al., 1998; McMaster et al., 2008) that led to model improvements. Recently, the global Agricultural Model Improvement and Intercomparison Project (AgMIP; www.agmip.org), which seeks to improve characterizations of the effects of climate change on agriculture, has adopted the ICASA standards as a means of harmonizing and managing the data used by the numerous participating crop modeling teams.

Experience with the IBSNAT standards and files showed that they contained ambiguities and lacked descriptors for characterizing certain crops and management practices. Members of ICASA and other organizations began revising the standards to reduce ambiguities, facilitate processing by a wider range of software, and include more types of crops and crop management practices. Use of the resulting ICASA Version 1.0 standards (Hunt et al., 2001, 2006) highlighted additional issues that needed attention such as providing more extensive metadata and describing tillage and mulching practices. Furthermore, stakeholders requested that the standards permit other digital formats such as spreadsheets and relational databases. Thus, the ICASA standards have been further revised and expanded, with emphasis on standardizing vocabularies, clarifying relations among variables, and providing for description of additional management practices and types of measured data.

The goal of the ICASA standards, following the thrust of the earlier IBSNAT standards, is to provide a reliable and flexible structure both for documenting field experiments (or their equivalents in greenhouses or growth chambers) and for specifying realistic conditions for dynamic simulations or other applications. As an admittedly idealized goal, based on the information provided in a data set, one would be able to replicate perfectly the described experiment. To fully document an experiment, detailed information is required on weather, soil, crop cultivars, weeds, diseases, pests, and crop management, along with measurements of crop growth and of dynamic soil characteristics. In practice, of course, it is exceedingly difficult to reproduce an experiment under field conditions due to the inherent season-to-season variability in weather, soil conditions, pest populations and crop management. Thus, a consensus has evolved over a level of description time that balances completeness, meaningful level of detail, and feasibility for data recording and management.

The standards were originally developed under the "minimum data set" concept, which was elaborated at the 1983 symposium at ICRISAT (Nix, 1984). We emphasize, however, that because of the diversity of experiments that are considered, the list of variables documented in the standards is closer to a "maximal data set." Indeed, since the standards can be extended to accommodate new variables, and a given implementation can be limited to a small subset of variables, the standards are better viewed as describing an "open data set" concept where databases are structured to satisfy the needs of specific lines of research or decision support.

While interest in standards largely originated to support crop modeling, the standards are applicable in many disciplines. Most notably, this is through meta-analyses such as have been conducted for crop response to water (French and Schultz, 1984) and to elevated CO_2 (Kimball et al., 2002; Lam et al., 2012), fate of nitrogen in cropping systems (Gardner and Drinkwater, 2009), and environmental impacts of organic farming (Tuomisto et al., 2012). Meta-analyses are often profitably extended through inclusion of simulation studies of generated variables, as illustrated by the review of 74 studies on impacts of no-till management on soil carbon where the Century model was used to estimate soil organic carbon stocks (Ogle et al., 2012).

This paper describes the ICASA Version 2.0 standards, which include much more detailed and structured metadata, provisions for describing tillage and use of mulches, and more flexible identification of variables. We present examples as implemented in plain text and as a spreadsheet, describe an implementation developed for AgMIP, and review options for other digital formats. The standards will continue to evolve, both through addition of new variables and new architecture to manage, for example, indicators of data quality. Updates to the definitions will be posted through the DSSAT Foundation web site (dssat.net/data/standards).

2. Overview of the standards

The basic organization of the data is intended to allow description of essentially any field experiment or commercial crop production situation. A dataset thus may describe an experiment involving multiple sites and years as well as various crop and weed species, initial conditions and management practices. Data are organized through a central group of data that specifies the treatment structure, and which indexes replicates (if any), sequences within a crop rotation (if applicable), species components in multiple-cropping (if applicable), cultivars grown, crop management, and weather and soil conditions. This indexing also links the data from corresponding subsets that describe measurements of crops, soils or the aerial environment in a given experiment.

The relations among classes of data are presented as an entity-relationship diagram (Pedersen, 2004) in Fig. 1.

Measured field data typically include crop developmental stages, yield and yield components, and growth analysis data such as leaf area index (LAI), stem, leaf, aboveground and grain biomass, but they can include measurements of soil water content, soil nutrient levels, pest damage or other variables deemed relevant. Measurements may be recorded in two separate groups within the subset MEASURED_DATA. The group SUMMARY contains data recorded once for a given treatment, rotation, or crop component during the course of the experiment, such as the date a specific growth stage was attained, the grain yield, or the season total of water applied through irrigation. The group TIME_SERIES consists of measurements that are recorded at specific times throughout the experiment, thus representing a time series and being referenced to the specific dates when the measurements were made. Growth stages may be reported both as summary values (e.g., as time of anthesis or physiological maturity) or as time series (e.g., as stages attained at specific dates, most often using codes to identify specific stages). Variables of similar types (e.g., plant growth measurements, crop water balance descriptors, or pest effects) are arranged by sub-groups. Weather data and general descriptions of soil profiles are assigned to separate datasets because a single weather or soil data set may apply to multiple experiments.

2.1. Data items

The fundamental unit is a data item that contains a name plus one or more values, which may be numerical data, identifiers, codes, or descriptive text. The names are character strings with no distinction between upper and lower case. The information in a data item can be either for a variable (a data attribute), which contain data pertinent to the experiment/situation documented, or level indicators that are used to link related assemblages of data items (relational attributes). Examples of variables include grain yield, the date an experiment was planted, the name of a cultivar used, and the name of a specific treatment. Level indicators are specified as integers and typically relate treatments in an experiment or components of management treatments such as for irrigation or fertilization. The dictionary of variables recognized in the standards is provided in Supplement A. The dictionary somewhat parallels efforts to develop data ontologies in other branches of plant and agricultural sciences (The Plant Ontology Consortium, 2002; Shrestha et al., 2010; FAO, 2012).

Allowed data types for variables are numeric (integer or real) or character strings. A given variable must be of one data type. Units of measurement for numeric variables largely follow the International System of Units (SI), but centimeter (cm) and hectare (ha) are permitted in order to conform to dominant practices in agricul-

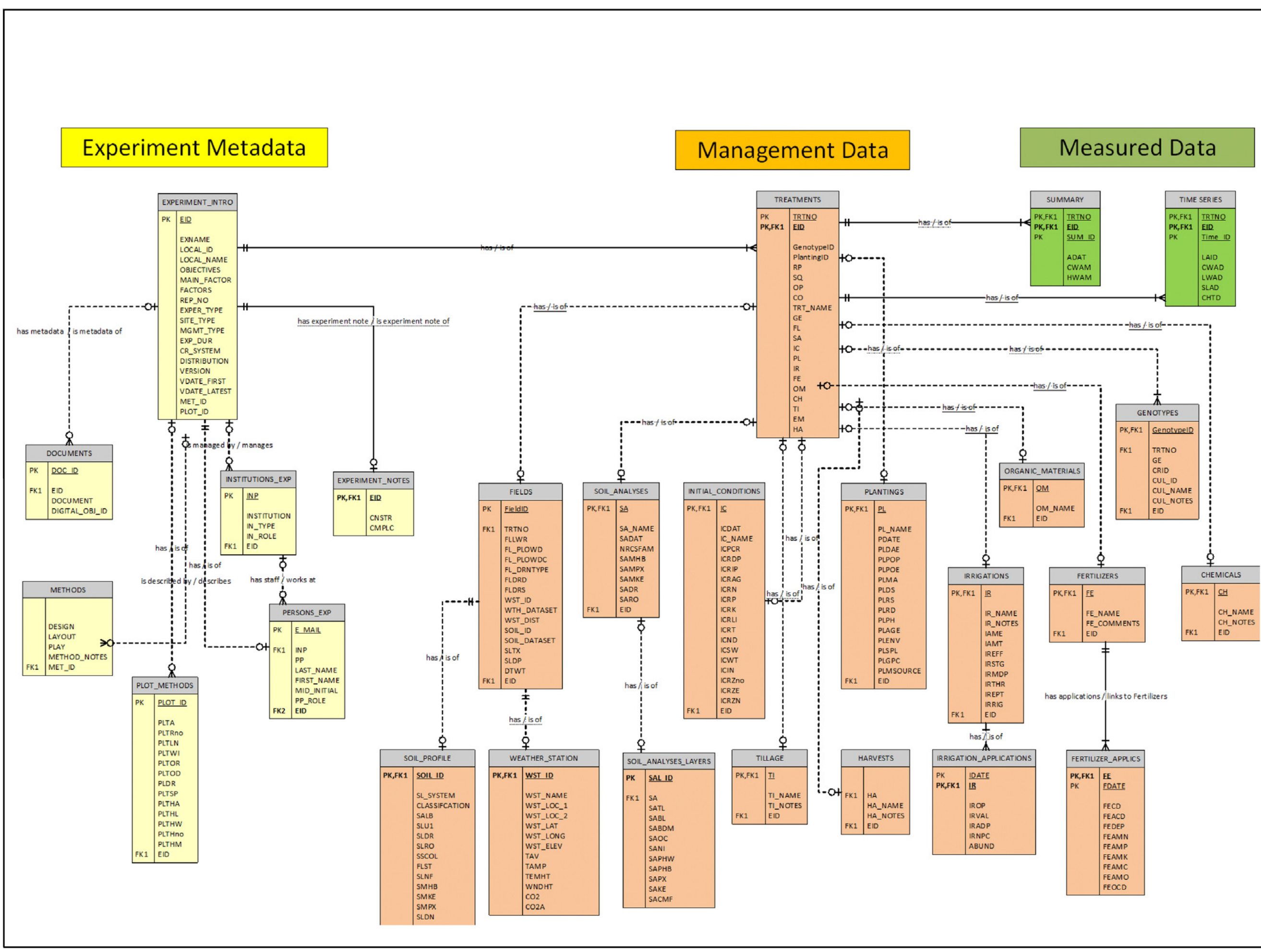

Fig. 1. Entity-relation diagram (Pedersen, 2004) for the ICASA Version 2 standards. Entities are represented by two-part boxes. The top, gray part contains the entity name. The bottom section lists the attributes. The connectors (lines) indicate relations among entities, with beginning and ends of connectors characterizing the type of relation (e.g., mandatory; one-to-one or one-to-many). Captions above the different groups of entities represent the major subsets of an experiment dataset. Various types of crop management are simplified by not showing the sub-groups describing individual events (e.g., for applications of organic materials, agrochemicals or for specific tillage practices). Attributes of summary and time series will vary greatly according to the experiment being documented.

tural research. Codes are provided for non-numeric variables where some degree of standardization is convenient or required, such as for describing fertilizer types, irrigation methods, or planting methods. Examples of codes are presented in Table 2, and the complete list of codes is included within Supplement A.

2.2. Datasets, subsets, groups and sub-groups

To facilitate the management of different types of information, items are organized according to thematic categories or expected use. Four levels of hierarchy are recognized, i.e., datasets, subsets, groups (the term "group" is used in place of the more cumbersome "sub-subset") and sub-groups.

Datasets represent the highest level of aggregation within the standards. An experiment dataset contains the metadata, treatment structures, crop management and any measurements of systems responses (e.g., crop growth and changes in soil nutrients or moisture). A second dataset may contain one or more soil profile descriptions, usually from a single location or geographic region or data source, and the third dataset contains daily weather data from one or more recording stations.

Subsets allow connected but not necessarily related data to be kept together. Three types of subsets within an experiment are recognized (Table 3). The metadata subset describes the objectives of the experiment, identifies the responsible researchers and institutions, and provides other background information. The management subset contains groups for treatment combinations, initial conditions, and details of crop management practices. The measured data subset contains measured crop and environmental responses, including yield, phenology and plant growth details. Both management details and measured data are linked independently to the specific details of the treatment (i.e., number, replicate, sequence and crop component).

Data groups are comprised of related data items such as those describing one type of field operation, one soil profile characterization or one weather station. The group describing field locations provides the links to the associated soil profile and weather datasets. Within the management subset, most groups correspond to management activities (e.g., planting or irrigation) or to field measurements (Table 3). Sub-groups are used for types of events occurring at multiple times (e.g., for irrigation) or for data related to layers in a soil profile.

2.3. Naming conventions

Experience from managing data from large numbers of experiments demonstrated the value of identifying datasets, subsets, groups and variables in a consistent manner. Furthermore,

Table 1
Examples of variables used to describe inputs, crops, management practices, environment, and other aspects of an experiment as specified in the master list of variables for the ICASA standards.

Long name	Abbreviated name	Description	Units or type
Cultivar_name	CUL_NAME	Cultivar name	Text
Field_elevation	FLELE	Elevation of field site	m
Field_name	FLNAME	Field name	Text
Fertilizer_level	FE	Fertilizer level in treatment structure	Number
Fertilizer_level_name	FE_NAME	Fertilizer level name	Text
Fertilizer_date	FEDATE	Fertilizer application date	Date
Application_depth_fert	FEDEP	Fertilizer application/mixing depth	cm
Fertilizer_applic_method	FEACD	Fertilizer application, code for method	Code
Irrigation_level	IR	Irrigation level in treatment structure	Number
Irrigation_date	IRDATE	Irrigation application date	Date
Irrig_amount_depth	IRADP	Irrigation amount as depth of water applied	mm
Irrigation_operation	IROP	Irrigation operation (e.g., furrow, drip, etc.)	Code
Planting_date	PLDATE	Planting date	Date
Plant_pop_at_planting	PLPOP	Population at planting (vs. at emergence)	Number/m^2
Planting_material_age	PLAGE	Planting material age (also dormancy)	Day
Planting_material	PLMA	Planting material (e.g., seed, tubers, etc.)	Code
Anthesis_date	ADAT	Anthesis date	Date
Zadoks_21_growth_stage	Z21D	Zadok's growth stage 21 as date	Date
Grain_dry_wt_area_maturity	GWAM	Grain dry weight at maturity	kg/ha
Grain_N_area_maturity	GNAM	Grain N (total amount) at maturity	kg/ha
Harvest_index_maturity	HIAM	Harvest index (tops only) at maturity	kg/kg
Leaf_area_index_maximum	LAIX	Leaf area index, maximum in season	m^2/m^2
Leaf_number_per_stem	LNOSD	Leaf number per stem on a given day	Number
Leaf_area_index	LAID	Leaf area index on a given day	m^2/m^2
Tops_dry_weight	CWAD	Tops dry weight on a given day	kg/ha
Grain_number_area	GNOAD	Grain number on a given day	Number/m^2
Grain_dry_weight	GWAD	Grain dry weight on a given day	kg/ha
Harvest_index	HIAD	Harvest index on a given day	kg/kg
Pod_dry_weight	PWAD	Pod dry weight on a given day	kg/ha
Stem_dry_weight	SWAD	Stem dry weight on a given day	kg/ha
Tuber_dry_weight	UWAD	Tuber dry weight on a given day	kg/ha
Tiller_number	TNOAD	Tiller number (area basis) on a given day	Number/m^2
Grain_N_conc	GNPCD	Grain N concentration, percent on a day	%
Plant_P	PLPAD	P content on a given day	kg/ha
NO_3_soil_by_layer	NOSLD	NO_3 [N] conc. for a given layer and day	g/Mg
Soil_CO_2_emission	SCO2D	Soil CO_2 (as C) emission on a given day	g/m^2 d
Soil_organic_C_perc_layr	SLOC	Organic carbon as 100 × g/g dry soil, by layer in profile	%
Soil_pH_in_water	SLPHW	pH of soil in water, from layer in profile	Number
Soil_bulk_density_moist	SLBDM	Soil bulk density when moist	g/cm^3
Soil_water_lower_limit	SLLL	Soil water conc., lower limit for extraction	cm^3/cm^3
Temperature_maximum	TMAX	Temperature of air, daily maximum	°C
Rain_snow_fall	RAIN	Daily total precipitation, including snow	mm
Wind_speed_daily	WIND	Wind speed (run), daily	km
Ambient_CO_2_conc	ACO_2	Ambient CO_2 concentration, daily average	vpm

Table 2
Examples of codes (variable codes) used to represent specific inputs, crops, management practices, implements or other aspect necessary to characterize an experiment.

Category	Code	Description
Chemicals	CH001	Alachlor (lasso), metolachlor (dual) [herbicide]
	CH022	Malathion, mercaptothion [insecticide]
	CH051	Captan [fungicide]
Crop	ALF	Alfalfa/lucerne (*Medicago sativa* L.)
	MAZ	Maize (*Zea mays* L.)
Pest organism	CEW	Corn earworm (*Heliothis zea*)
	VBC	Velvetbean caterpillar (*Anticarsia gemmatalis*)
Application methods	AP001	Broadcast, not incorporated
	AP006	Foliar spray
	AP999	Application method unknown/not given
Drainage	DR000	No drainage
	DR001	Ditches
	DR002	Sub-surface tiles
Fertilizers	FE001	Ammonium nitrate
	FE006	Diammonium phosphate
	FE999	Fertilizer type unknown/not given
Irrigation methods	IR001	Furrow, depth to be given in mm
	IR003	Flood, depth to be given in mm
	IR005	Drip or trickle, depth to be given in mm
Organic materials	RE001	Crop residue
	RE 002	Green manure
Planting material	S	Dry seed
	T	Transplants
	R	Ratoon
Plant distribution	R	Rows
	H	Hills
	U	Uniform/broadcast
	RB	Rows on beds
Soil P analysis methods	SA001	Olsen
	SA003	Bray no. 2
	SA004	Mehlich I (double acid, 1:5)
Tillage implements	TI003	Moldboard plow
	TI004	Chisel plow, sweeps
	TI009	Tandem disk

relatively short names facilitate manipulation in software such as spreadsheets or statistical packages. Recognizing, however, that flexibility is necessary in naming datasets and subsets in order to accommodate user needs and established local practices, users may propose identifiers that deviate from the preferred convention or establish translation lists that allow their data to be mapped onto the Version 2 dictionary (Supplement A).

2.3.1. Dataset names and identifiers

Datasets are named as being of one of three types, EXPERIMENT, SOIL or WEATHER (Table 3). Individual datasets are further assigned a specific identifier that indicates the contents of the set. These are constructed differently for experiments, soils, and weather. For experiments, an identifier is constructed by combining an institution or region code (three characters, e.g., "UFL" for "University of Florida", "CAN" for "Canada"), a code for the site or set of sites (three characters, e.g., "GNV" for "Gainesville"), a year code (four characters usually representing the year in which the experiment was initiated or harvested), an experiment number or code (four characters), and a crop, multi-crop (for mixed cropping or crops with weed populations) or sequence (for rotation experiments) code (three characters). Thus, the third experiment (0003) conducted by the University of Florida (UFL) at Gainesville (GNV) in 2006 with soybeans would be identified as "UFLGNV20060003SBN".

For weather datasets, identifiers can be constructed from Institution and Site codes plus, if desired, four digits to indicate the starting year (e.g., UFLGNV2006). Optionally, a fourteen-character code may be used where the first four digits indicate the number of years of data, and the last four characters identify other characteristics of the set. Thus, "UFLGNV196825M1" might indicate a 25-year series from Gainesville, Florida that started in 1968 and that used a method "M1" for estimating daily solar radiation.

For soil data, specific set identifiers can be constructed using a three-character code for the institution or region, plus a three-character code for the site or collection of sites. Longer names (within a fourteen character limit) can be used to provide information on the contents of the set. Thus, "ARIZONA" could be used as a general name for a dataset containing soil profile descriptions from diverse sites in Arizona.

2.3.2. Names and identifiers for data subsets and groups

For experiments in which each dataset is restricted to data from that experiment, the general names (Table 3) identify each subset (e.g., METADATA, MANAGEMENT or MEASURED DATA) and group (e.g., TREATMENTS, PLANTING or TILLAGE). Weather and soil datasets often contain information for multiple weather stations or soil profiles, so specific subset or group identifiers may need to be appended to the basic name to uniquely identify the data items. For weather subsets that contain only a part of an overall dataset, e.g., a single year or portion of a year, the subset should be identified with the general name plus a suffix that provides more specific information, such as WEATHER_STATION:UFLGNV2004S1, in which the suffix "S1" might indicate "Season 1". For soil subsets (single soil profile descriptions), the specific identifiers are 14-characters long. The Institution and Site codes occupy the leading six positions, the year the profile was described is in the next four positions, and a specific profile identifier occupies the remaining four positions. Thus, a valid soil subset identifier would be SOIL_PROFILE:UFLGNV20040001.

2.3.3. Names for variables

The length of many variable names in the ICASA Version 1 standards were limited to four or five characters to permit displaying a name as a label over a column of data that contained no more than five digits. The extension of the standards to other digital formats has reduced the need for short variable names, so the ICASA V2 standards allow two primary name formats, a long variable name and an abbreviated name (Table 1). The long name is generally 12–24 characters long and uses complete words as much as possible, with the words being separated by the underscore ("_") character. Abbreviated names for most data are four or five characters long to permit their use as compact column headings. The names often correspond to those used in DSSAT, but some variables have been re-named to improve clarity.

The naming strategy for abbreviated names of measured data emphasizes consistency in use of each character position in the name string. For observed data relating to plant organ weights or nutrient content, the first character indicates the tissue type, and the second describes the quantity being measured. Thus, for LWAD, "L" is for leaf, and W is for dry weight ("F" for fresh weight). The third and sometimes the fourth character(s) indicate, the measurement reference, such as "A" for land area basis, "PC" for percentage and "NO" for number (counts). The final character indicates the time or frequency of measurement or observation, such as "D" for specific sampling or observation dates and "H" for data recorded at harvest.

For display purposes, variable labels are also defined based on the abbreviated names. The main differences are that variables where values are given as percentages are indicated through use of "%" in place of "PC" and as counts through use of "#" in place of "NO".

Most variables names are unique, even when the same basic variable is described in different datasets or subsets. This ensures

Table 3
Descriptions of the datasets, subsets and principal groups.

Dataset	Subset	Group	Description
EXPERIMENT			Complete description of management and initial conditions for a real or synthetic experiment (or very closely linked set of experiments). Data measured during or at the end of the experiment. The information presented should be sufficient to allow thorough interpretation or analysis of the results and for simulation of the experiment
	METADATA		General information describing the experiment
		GENERAL	Objectives, publications, distribution controls, ICASA version, experiment types (e.g., station vs. on-farm; irrigated vs. rainfed)
		PERSONS	Names, roles and contact information for persons involved in the experiment or preparation of the dataset
		INSTITUTIONS	Names, roles and contact information for institutions involved in the experiment
		DOCUMENTS	Any publications or other media associated with the experiment, especially journal articles or project reports
		METHODS	Experimental design, layout, field plot information
		PLOT_DETAILS	Information on the plot size and characteristics
		NOTES	Information on constraints or complications affecting the usability of the dataset
	MANAGEMENT		Descriptions of different management practices
		TREATMENTS	Treatment names and level codes for crop rotations, crop components and experimental factors
		GENOTYPES	Crop and cultivar identifiers, names, and background
		FIELDS	Field description including links to weather station and soil profile information
		SOIL_ANALYSES	Details of classical soil surface layer analyses
		INITIAL_CONDITIONS	Starting conditions including above- and below-ground residues, and water and nitrogen in the complete profile
		PLANTINGS	Planting date, seed and initial plant populations, seeding depth, and plant distribution (row spacing, etc.)
		IRRIGATIONS	Irrigation dates and amounts, flood and water table depths, thresholds for automatic applications
		FERTILIZERS	Fertilizer amounts, types, and dates of applications
		ORGANIC_MATERIALS	Details of straw, manure and other organic material applications
		MULCHES	Details of soil covers such as plastic sheeting or fabrics
		CHEMICALS	Herbicide, pesticide, or growth regulator application details
		TILLAGE	Dates and types of tillage operations
		ENV_MODIFICATIONS	Adjustments to weather variable such as those that could be made during growth chamber, CO_2 enrichment, or rainout shelter studies
		HARVESTS	Harvest dates, components harvested, and percentages removed
	MEASURED_DATA		Data from measurements of system performance, potentially including crop or weed growth, soil constituents, and gas fluxes or concentrations
		SUMMARY	Summary data (may include data from means over replicates or from individual plots) that are collected during the course of or at the end of an experiment
		TIME_SERIES	Time-course data (may include data from means over replicates or from individual plots) collected during the course of an experiment
SOIL			Set of soil profile data for pedons from one or more sites in a region
	METADATA		Descriptions of the sources of soil data
		GENERAL	General information on the set of soil profiles such as methods used to obtain data and possible limitations on distribution
		PERSONS	Persons involved in preparing the soil data
		INSTITUTIONS	Institutions involved in preparing the soil data
		DOCUMENTS	Documents describing the source or applications of the soil data
	SOIL_PROFILES		Soil profile data for individual pedons
		PROFILE_METADATA	Background information on a single profile
		SOIL_PROFILE	Data describing a single pedon, including surface and soil layers
WEATHER			Daily weather data from one or more stations
	METADATA		Descriptions of the sources of the weather data
		GENERAL	General information on the set of weather stations such as methods used to obtain data and possible limitations on distribution
		PERSONS	Persons involved in preparing the weather data
		INSTITUTIONS	Institutions involved in preparing the weather data
		DOCUMENTS	Documents describing the source or applications of the weather data
	WEATHER_STATION		Daily weather data typically for a single experiment or season, but preferably for a complete year or longer
		STATION_METADATA	Description of a single weather station
		WEATHER_GENER_PARMS	Parameters for weather generators that are applicable to the station
		WEATHER_DAILY	Daily weather data from a single weather station

that the variable can be used independently of its assigned dataset or subset. Situations of multiple instances of a variable name primarily involve metadata (e.g., e-mail address of data providers), dates, soil depths and application methods.

3. Implementations of the standards

The standards are designed to permit implementations in whatever digital formats meet the needs of users. The primary requirement was that the dictionary of variable names, identifiers and codes should include definitions and units of measurement, be comprehensive enough to embrace experiments conducted with different objectives, and allow use in different digital formats. Thus, the Version 2 master list of variables includes the variable names, definitions, units of measurement, data types and suggested allowable minimum and maximum values (Supplement A). An additional requirement was to define the relations among variables so that treatment structures and sampling regimes (e.g., dates or positions within a soil profile) could be described accurately. This required that the master list also assign variables to their corresponding positions within the hierarchy of datasets, subsets, groups, and if applicable, sub-groups.

3.1. Implementation in the plain text format

When the standards are implemented in plain text, a single file may correspond to a dataset, subset or group. Data items are formatted in columns that are headed by the abbreviated variable name. The names are presented above the variables or level indicators on a header line that begins with the symbol '@' (Fig. 2). Multiple names can occur on one line. Headers for long text variables are extended to the expected length by appending a string of periods after the variable name to indicate the full allowed length of values for the respective variable (e.g., "TRT_NAME..." under treatments in Fig. 2). Blank lines may appear anywhere in the file. Comments are indicated by an explanation point ("!") as the first character in the line and may appear in any part of a dataset. Fig. 2 presents a dataset for a single experiment, without the measured crop, weather and soil data. Supplement B contains the complete description.

3.1.1. File organization and naming

The usual file organization is for data from a given dataset or subset to reside within a single file. The file is named using a conventional name-plus-extension format where the name identifies the specific set or subset, and the file extension specifies the type of dataset. The name is simply the identifier as described in Section 2.3.1. File extensions may be of three main types. The extension "EXP" identifies an experiment, "SOL" identifies a soil profile dataset, and "WTH" identifies a weather dataset. The subset for observed and measured data from an experiment is often assigned to two additional, separate files for summary and time-course groups. Their respective extensions are "EXA" for summary data and "EXT" for time-course data. Large weather datasets can be subdivided to simplify file handling. In such cases, the subset names indicate the time period covered and are used as the file names. Thus, a dataset with the name UFLGNV19750020, having 20 years of daily data starting in 1975, could be broken into two 10-year subsets, identified as UFLGNV19750010.WTH and UFLGNV19850010.WTH.

3.1.2. Special characters for file processing

Within individual files, specific symbols are used to guide digital parsing as well as visual inspection. Datasets are identified with a dollar sign ($) in the first column of the first line. This symbol is followed by the general name and the specific identifier. Three valid examples are $EXPERIMENT:UCADAV19960001SBN,$WEATHER:UFLGNV2001 or $SOIL:ARIZONA. Subsets and groups are identified by placing two dollar signs ($$) for subsets and an asterisk (*) for groups in the character positions before the identifier, for example, *TREATMENTS, *IRRIGATIONS or $$WEATHER_STATION:UFLGVN2004S1. An "@" symbol indicates the start of a row of column headers given as abbreviated variable names.

3.1.3. Linking data across datasets and subsets

Numeric level identifiers and character strings link information across datasets, subsets, groups and sub-groups in a manner analogous to how keys link tables in a relational database. Links among assemblages of data require multiple identifiers in order to accommodate combinations of treatments, replicates, crop rotations and crop or weed associations as specified in the TREATMENTS subset (Table 4 and Fig. 2). Treatment levels and the associated links are identified with integers, preferably numbered from 1 to the maximum number of levels within a given experiment. A value of 0 for the replicate implies that the values represent means across replicates.

The TREATMENTS group in Fig. 2 illustrates cases of single replicates and means and of different plant species components (wheat and weeds). Note that while data for individual replicates are given for wheat, the weed component is only specified at the level of means of the two replicates.

```
$EXPERIMENT:KANKSTS200401WHT

$$METADATA
*GENERAL
@ EXNAME
  Fictitious experiment based on KSAS98101 in version 1 format
@ LOCAL_ID.. LOCAL_NAME
  WHEX200401 Godwin
@ OBJECTIVES
  To determine the degree of interaction between N and water supply
@ MAIN_FACTOR  FACTORS.........................  REP_NO
  FE(N)        4GE,4PL(Date),3IR,3FE(N)               4
@ EXPER_TYPE SITE_TYPE MGMT_TYPE EXP_DUR
  ET001      ST001     MT001           1
@ METHODS
  Standard field-plot technique using cone seeder.
  Standard breeding observations for phenology. Hege combine for yield.
  MDAT2 obtained using dye absorption technique (MDAT loss of green)
  Small sub-plots for TWAD.
  Gravimetric procedures for SW1D
@ CR_SYSTEM
  Winter wheat, conventional tillage
@ DISTRIBUTION
  Use at will but acknowledge source. No secondary distribution
@ VERSION... REV_DATE_FIRST RE_VDATE_LATEST
  ICASA02.00 2004175        2013002
@ REVISION_NOTES
  Revised by L.A. Hunt for compliance with ICASA 2.0 standard.

*PERSONS
@ PP LAST_NAME............. FIRST_NAME.......... MID_INITIAL PP_ROLE
   1 Doe                    John                 H           PR001
   2 Smith                  Maria                Q           PR002
   3 Lin                    McCall               J           PR004
   4 Hunt                   L.                   A           PR006
@ PP IN E_MAIL
   1  1 John.Doe@ars.usda.gov
   2  1 Maria_Smith@kstate.edu
   3  2 mlin7@acmedata.com
   4  3 (none)

*INSTITUTIONS
@ IN INST_NAME............................ IN_TYPE IN_ROLE
   1 Kansas State Univ.                    IT001   IL001
   2 ACME Data Miners, Inc.; Manhattan, KS IT005   IL002
   3 University of Guelph                  IT001   IL004

*DOCUMENTS
@ DOC_ID DOCUMENT
       1 Smith, M.Q., 2004. N cycling in a wheat-soil system. Ph.D.,Kansas State
       2 Doe et al., 2004. Nitrogen management in wheat. J. Irrev. Crop Sci. 3:123
@ DOC_ID DIGITAL_OBJ_ID
       1 -99
       2 10.1016/j.ics.2004.03.009

*PLOT_DETAILS
@ PLTA PLTR# PLTLN PLTWI PLTOR PLTOD  PLDR PLTSP PLTHA PLTH# PLTHL PLTHW PLTHM
  12.0 15.00  4.00  3.00     0 90.00    90    20   4.8     8   3.0   1.6  hand
@PLLAY
  RCBD
!If plot organization cannot be documented as above, describe in 'Notes'

*NOTES
@ CONSTRAINTS
  Constraints could relate to: Soil physical conditions
  (surface compaction, duripan); Moisture deficit; Moisture
  excess; Nitrogen; Phosphorous; Potassium; Salinity; pH;
  Aluminum/Manganese; Boron; Other micro-nutrients; Above-ground
  diseases; Below-ground diseases; Nematodes; Aphids; Insects;
  Weeds; Other
@ COMPLICATIONS
  Complications could deal with: Stand loss (e.g.,winter kill);
  Wind damage (->leaf shredding); Sandblasting; Hail; Lodging;
  Bird feeding; Shattering; Sprouting; Combine settings;
  Seed mixing; Animal grazing; Other.

!And the following specific information is necessary in most cases. Data
!subsets can be omitted, however, if the particular aspect (e.g., irrigation)
!is not a factor in the experiment.
```

```
$$MANAGEMENT
*TREATMENTS                                -----------FACTOR LEVELS-----------
@TRTNO RP SQ OP CO TRT_NAME.............. GE 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

*GENOTYPES
@ GE CRID CUL_ID CUL_NAME.... CUL_NOTES
   1  WHB IB0488 Newton       Hard red winter,Central Plains
   2  GRW IB0001 Garytown     Central Gt.Plains quackgrass
   3  RCL IB0001 Bigyield     General double cut
   4  RCL IB0002 Earlywonder  Single cut

*FIELDS
@ FL    ID_FIELD..  FL_NAME
   1    KSTASH0001  Research Park F105
@ FL    FL_LOC_1    FL_LOC_2  FL_LOC_3
   1    USA         KS        Research Park, Ashland
@ FL    FL_LAT      FL_LONG   FLELE  FLSL FLSLL FLSLA FAREA  FLOB FLLWR
   1      37.1167   -90.4521     85    20   200    90   226   -99   -99
@ FL    FL_PLOWD    FL_PLOWDC  FL_DRNTYPE FLDRD FLDRS
   1          25         0.4  DR000          0     0
@ FL    WTH_SUBSET. WTH_DATASET WST_DIST SOIL_SUBSET... SOIL_DATASET
   1    KANKSTS2004 KANKST           -99 KANKST20040001 KANSAS
@ FL    SLTX  SLDP  DTWT
   1     -99   -99   -99

*SOIL_ANALYSES
@ SA    SA_NAME
   1    Science Associates Lab
@ SA    SAYR SADAY SAMHB SAMPX SAMKE SADR  SARO
   1    2004   260   -99   -99   -99  -99   -99
@ SA    SATL  SABL SABDM  SAOC  SANI SAPHW SAPHB  SAPX  SAKE SACMF
   1       0    15   -99   -99   -99  5.85  5.21   -99   -99   -99
   1      15    30   -99   -99   -99  5.97  5.60   -99   -99   -99

*INITIAL_CONDITIONS
@ IC    IC_NAME
   1    10 days before planting
@ IC    ICRYR ICRDY ICPCR ICRDP ICRIP ICRAG  ICRN  ICRP  ICRK ICRLI  ICRT  ICND
   1    2004   260  WHB      10    50  2000  1.00  1.00   -99  10.0   100   -99
@ IC    ICYR ICDAY  ICSW  ICWT  ICIN ICRZC ICRZE
   1    2004   260   300   100   100   -99   -99
@ IC    ICTL  ICBL ICH2O ICNH4 ICNO3
   1       0     5 0.205   3.4   9.8
   1       5    15 0.205   3.4   9.8
   1      15    30 0.170   3.2   7.3
   1      30    60 0.092   2.5   5.1
   1      60    90 0.065   2.2   4.7
   1      90   120 0.066   2.7   4.3

*PLANTINGS
@ 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 PLDOY PLDOE PLPOP PLPOE  PLMA  PLDS  PLRS  PLRD  PLPH  PLDP PLMWT
   1    2004   270   -99   162   150     S     R    17    90   1.0   5.5   120
   2    2004   289   -99    40    35     S     R    17    90   1.0   3.5    20
   3     -99    25   -99   400   370     S     B   -99   -99   1.0   0.0     5
   4     -99    45   -99   400   375     S     B   -99   -99   1.0   0.0    10
@ PL    PLAGE PLENV PLSPL  PLG% PLMSOURCE
   1        0   -99   -99    97 Univ. breeder
   2        0   -99   -99    90 Univ. weed scientist
   3        0   -99   -99    95 Univ. breeder
   4        0   -99   -99    95 Univ. breeder
```

```
*IRRIGATIONS
*IRRIGATIONS
@ IR    IR_NAME
   1    Furrow irrigation on specified dates
   2    Drip irrigation on specified days after planting
   3    Automatic irrigation
@ IR    IAME  IAMT IREFF IRSTG IRMDP IRTHR IREPT IRRIG
   1    IR001  -99   1.0   -99   -99  0.03   -99     1
   2    IR005  -99   1.0   -99   -99   -99   -99     1
   3    IR001  60    1.0   -99   -99   -99   -99     1
@ IR    IRYR IRDAY  IROP IRVAL IRADP  IRN%
   1    2005    96 IR001    65     0   -99
   1    2005   110 IR001    78    20   -99
   1    2005   117 IR001    70    20   -99
   2     -99   170 IR005    65    20  0.03
   2     -99   200 IR005    78    20  0.05
   3    2005   200 IRA01    60   -99  0.01
   3    2005   250 IRA00    60   -99  0.05

*FERTILIZERS(INORGANIC)
@ FE    FE_NAME
   2    Single application at planting
   3    Applications at planting and in spring
@ FE    FEYR FEDAY  FECD FEACD FEDEP FEAMN FEAMP FEAMK FEAMC FEAMO FEOCD
   2    2004   289 FE001 AP001    15    60   -99   -99   -99   -99   -99
   3    2004   289 FE001 AP001    15    90   -99   -99   -99   -99   -99
   3    2005    56 FE001 AP001     1    90   -99   -99   -99   -99   -99

*ORGANIC_MATERIALS
@ OM    OM_NAME
   1    Low lignin
   2    High lignin
@ OM    OMYR OMDAY  OMCD OMACD OMDEP OMINP OMAMT
   1    2004   280 OM000 AP002    10    80  1000
   2    2004   280 OM000 AP002    10    80  1500
@ OM    OMYR OMDAY OMH20  OMC%  OMN%  OMP%  OMK% OMLI%
   1    2004   280   0.1  40.1   4.1   -99   -99    10
   2    2004   280   0.1  40.1   4.1   -99   -99    20

*CHEMICALS
@ CH    CH_NAME
   1    Pre-plant weed control
@ CH    CH_NOTES
   1    Poor control of leafy spurge due to dry conditions
@ CH    CHYR CHDAY  CHCD CHACD CHDEP CHAMT CH_TARGETS
   1    2004   260 CH001 AP001   1.3   2.2 BRLFWD

*TILLAGE
@ TI    TI_NAME......................... TI_NOTES
   1    Standard preparation for wheat    Lea type plough with long moldboard
@ TI    TIYR TIDAY TI_OP_NAME... TIORD TIIMP TIDEP TIMIX
   1    2004   250 Ploughing       -99 TI005  15.0    30

*ENV_MODIFICATIONS
@ EM    EM_NAME
   1    CO2 supplementation early
@ EM    EMYR EMDAY ECDYL EMDYL ECRAD EMRAD ECMAX EMMAX ECMIN EMMIN
   1    2005   180     0     0     0     0     0     0     0     0
   1    2005   220     0     0     0     0     0     0     0     0
@ EM    EMYR EMDAY ECRAI EMRAI ECCO2 EMCO2 ECDEW EMDEW ECWND EMWND
   1    2005   180     0     0     R   600     0     0     0     0
   1    2005   220     0     0     R   380     0     0     0     0

*HARVESTS
@ HA    HA_NAME................... HA_NOTES
   1    Harvest with no screening   Hand cut, ALMACO plot thresher
@ HA    HAYR HADAY HASTG HACOM HARM  HAREA HASIZ  HAP%  HAB%
   1    2005   273 GS090 HAC0P HM001   100     M    95    50
```

Fig. 2. Example of an ICASA experimental details dataset implemented as a flat file. Lines that start with "!" are comments.

Table 4
Indicators required to link data in different datasets and subsets.

A. Indicators required to link subsets and groups in different datasets		
Dataset 1	Dataset 2	Links required
Experiments	Weather	Weather set and subset identifiers
Experiments	Soil	Soil set and subset (soil profile) identifiers
B. Indicators required to link groups in the same (experiment) dataset		
Main group	Subsidiary group	Links required
Treatments	Genotypes	Genotype level no.
Treatments	Fields	Field level no.
Treatments	Soil analysis	Soil analysis level no.
Treatments	Initial conditions	Initial conditions level no.
Treatments	Plantings	Planting level no.
Treatments	Irrigations	Irrigation level no.
Treatments	Fertilizers	Fertilizers level no.
Treatments	Organic materials	Organic material level no.
Treatments	Chemicals	Chemicals level no.
Treatments	Tillage	Tillage level no.
Treatments	Environmental modifications	Environmental modifications level no.
Treatments	Harvests	Harvests level no.
Treatments	Summary	Treatment no., replicate no., sequence no., option no., component no.
Treatments	Time course	Treatment no., replicate no., sequence no., option no., component no.

3.1.4. Special syntax and formatting rules

To simplify parsing by software and to facilitate viewing and manual checking of the files, the plain text implementation limits the line length to 254 characters. Variable name abbreviations, variables, level indicators, and dataset and subset identifiers are allowed up to 31 characters. Specific rules for comments, missing data, non-applicable data, data flags, sub-samples, and end-of-file markers are summarized in Table 5. Missing numeric data are identified by −99, and missing character strings or text, by the string "−99". Missing information such as for application methods, irrigation types or fertilizer types are identified with specific codes (Table 2). Dates for growth stages are presented using four digits for the year and three digits for the day ("year-day of year" format such as "2009253" for day 253 of 2009).

3.1.5. File additions and modifications

To incorporate new data items in a dataset, additional abbreviated names are defined, and the corresponding columns of data items are added within an existing or new group. Adding additional data items at the end of existing long rows of data items is discouraged. Two exceptions are when the variable is needed as a link to data elsewhere in the file, e.g., additional factors in the treatment subset, or when adding data at the end of existing rows of data items might disrupt the overall configuration of the file (e.g., for a new daily weather variable).

Table 5
Special syntax and formatting rules for the ICASA standards.

Item	Syntax or formatting rule
Comments	Information on data quality, problems with treatments, or aspects of an experiment that are difficult to quantify should normally be recorded under the metadata subset for an experiment. Note that specific comment fields are provided for methods, production constraints and unexpected complications that arise during experiments. However, comments can also be inserted immediately after a dataset or subset identifier, or after a line of data using '!' as the first character in the line (used only in the plain text implementation)
Missing data	Indicated by a value of −99 for numeric data and the string "−99" for text data
Non-applicable data	Indicated by a value of −99. For example, row width and spacing for a crop that is broadcast sown
End-of-file	The symbol '=' can be entered as the only character on the last line of a file to indicate the end of a file. Its use is recommended to indicate whether a file has unintentionally been truncated (used only in the plain text implementation)

3.2. Implementing the standards in other digital formats

3.2.1. The AgMIP Crop Experiment database as an alternative digital implementation

The AgMIP Crop Experiment (ACE) database contains data from detailed field experiments as well as less-detailed data from variety trials conducted by international agricultural research centers, universities and the private sector (Villalobos, 2012). The database is designed to hold descriptions potentially of thousands of field experiments and to permit concurrent access by a global user community.

ACE is implemented in a non-relational database using the Riak platform (wiki.basho.com). Riak is an open source, hybrid, key-value data storage system that is deployed in a clustered fashion where data are distributed across multiple nodes (locations). Riak is intended to be simple to operate, to tolerate failures at individual nodes, and to scale easily from prototypes to large databases. Data are sorted into "buckets" that store keys and "values". The values are type-agnostic, so digital formats may range from integers to text files to complex objects.

In ACE, experiment-related data are divided into the buckets, with each experiment assigned a unique key. A separate metadata bucket stores and indexes a searchable subset of the experiment data, enabling fast searching within the database. Since information from a given experiment is stored in a single bucket, rather than multiple tables, fast retrieval times for large amounts of data are possible.

In ACE, the actual data that describe an experiment are stored using JSON (JavaScript Object Notation, www.json.org) structures, so this is the main component of ACE where the ICASA standard is applied. A sample JSON structure representing a maize experiment is presented in Fig. 3. The experiment has two treatments, rainfed and irrigated. Data from treatment 1 are used in treatment 2, except where explicitly overridden by the new data entries, in this case, irrigation records. Irrigation data are stored as nested JSON structures.

3.2.2. Other digital formats

Relational databases are widely used to store and query large sets of data. The groupings of data recognized in the ICASA standards readily permit implementation as a relational database. Groups and sub-groups correspond to tables in a relational database. Individual data items are stored in database fields within tables. Relations among tables would follow those defined in the entity-relation diagram given in Fig. 1. While the ICASA standards focus on ensuring accurate characterization of single experiments, the standards are readily extended to storing multiple experiments using a relational model.

Spreadsheets are widely used for data entry, are available on most personal computers, and can be read with a range of other software. Similar to relational databases, individual sheets can be created that correspond to a group or sub-group and are named accordingly. Each column in such a sheet would contain a variable name as a header and the corresponding data values in the cells below the header. Supplement C provides an example spreadsheet based on the wheat dataset used in Fig. 2 and Supplement B. In this example, three header rows are given, providing the full variable name, the units of measurement and the display name.

Data interchange formats such as JSON (used in the ACE database) and XML (eXtensible Markup Language; XML Core Working Group, 2008) provide further examples of options for implementa-

```
"data_source": "DSSAT",  /* data were input from DSSAT-format input files */
"crop_model_version": "v4.5",
"exname": "UFGA8201MZ_1",
"local_name": "NIT X IRR, GAINESVILLE 2N*3I",
"in": "UF",
"people": "BENNET,J.M. ZUR,B. HAMMOND,L.C. JONES,J.W.",
"institution": "UNIVERSITY OF FLORIDA, GAINESVILLE, FL, USA",
"site": "IRR.PARK,UF.CAMPUS 29.63;-82.37;40.;FLA",
"tr_name": "RAINFED LOW NITROGEN",
"id_field": "UFGA0002",
"wst_id": "UFGA",        /* weather station id */
"soil_id": "IBMZ910014", /* soil id            */
"fl_lat": "29.63",       /* latitude           */
"fl_long": "-82.37",     /* longitude          */
"flele": "40",           /* field elevation    */

"initial_condition": {
    "icpcr": "MAZ",      /* previous crop was maize       */
    "icdat": "19820225", /* date for initial conditions   */
    "icrt": "100",       /* initial root residue kg/ha    */
    "icrag": "1000",     /* initial surface residue kg/ha */
    "icrn": ".8"         /* initial residue N %           */
    "soilLayer": [       /* init soil layer data:         */
    /* depth (cm), moisture (frac),     NH4 (ppm),   NO3 (ppm) */
    {"icbl":  "5", "ich2o": ".086", "icnh4": ".5", "icno3": ".1"},
    {"icbl": "15", "ich2o": ".086", "icnh4": ".5", "icno3": ".1"},
    {"icbl": "30", "ich2o": ".086", "icnh4": ".5", "icno3": ".1"},
    {"icbl": "60", "ich2o": ".086", "icnh4": ".5", "icno3": ".1"},
    {"icbl": "90", "ich2o": ".076", "icnh4": ".6", "icno3": ".1"},
    {"icbl":"120", "ich2o": ".076", "icnh4": ".5", "icno3": ".1"},
    {"icbl":"150", "ich2o": "0.13", "icnh4": ".5", "icno3": ".1"},
    {"icbl":"180", "ich2o": ".258", "icnh4": ".5", "icno3": ".1"}
    ]},

"management":{
      "events":[
         {"event":"fertilizer",       /* Fertilizer application  */
                  "date":"19820225",  /* Feb 25, 1982            */
                  "fecd":"FE001",     /* Ammonium nitrate        */
                  "feacd":"AP001",    /* Broadcast, incorporated */
                  "fedep":"10",       /* 10 cm deep              */
                  "feamn":"27"},      /* 27 kg[N]/ha             */

         {"event":"tillage",
                  "date":"19820225",   /* Feb 25, 1982            */
                  "tiimp":"TI003",     /* Moldboard plow          */
                  "tidep":"35",        /* 35 cm depth             */
                  "ti_name":"Moldboard plow"},

         {"event":"planting",
                  "date":"19820226",   /* Feb 26, 1982            */
                  "crid":"MAZ",        /* maize planted           */
                  "cul_name":"McCurdy 84aa",  /* cultivar name    */
                  "plpop":"7.2",       /* plant population        */
                  "plme":"S",          /* dry seed planted        */
                  "plds":"R",          /* planted in rows         */
                  "plrs":"61",         /* row spacing 61 cm       */
                  "pldp":"7",},        /* planting depth 7 cm     */

         {"event":"fertilizer",
                  "date":"19820412",
                  "fecd":"FE001",
                  "feacd":"AP001",
                  "fedep":"10",
                  "feamn":"35"},

         {"event":"fertilizer",
                  "date":"19820517",
                  "fecd":"FE001",
                  "feacd":"AP001",
                  "fedep":"10",
                  "feamn":"54"},

         {"event":"harvest",           /* crop harvested          */
                  "date":"19820703"}   /* on July 3, 1982         */
      ]}
```


Fig. 3. Example of a portion of a crop experiment described using a Java Script Object Notation (JSON) data structure as used by AgMIP. The dataset corresponds to a maize experiment which was planted on 26 Feb 1982. Explanatory notes are given in comments delineated by "/* comment */".

tions of the standards. The preferred approach with XML is to define an XML schema or document type definition (DTD), so that XML-compatible software can read the file, perform basic validation checks, and correctly interpret the data for subsequent processing (XML Core Working Group, 2008). Typically, parties having a shared interest in a discipline develop an XML standard for the subject matter of interest. We are currently examining XML standards proposed by other groups involved in agricultural research or industry and expect to develop an XML prototype that is consistent with the ICASA standards.

4. Applications of the standards

The ICASA standards are intended to support any activity that requires detailed descriptions of a field experiment or commercial production situation. To date, the foremost use of the standards has been in simulation modeling where datasets have been used to specify initial conditions, weather inputs, soil inputs, and management for specific cropping situations and to provide measured data on crop responses that are used in model development and cross validation (e.g., White et al., 2008). Nonetheless, datasets can also be used for meta-analysis using regression and other techniques. For example, to compare a large set of measured and modeled datasets of solar radiation, White et al. (2011) prepared the daily data in weather datasets.

5. Standards documentation, revisions and expansions

Complete definitions of variable names, variable codes, dataset and subset names, and relations are maintained as a set of tables accessible at the DSSAT Foundation web site (dssat.net/data/standards). The standards allow for new datasets, subsets, groups and variables to be defined according to the needs of specific users, but the utility of the standards requires a uniform vocabulary. Users are encouraged to propose new variable names and codes, as well as subsets, groups and sub-groups, together with concerns over inconsistencies in the overall standards.

Use of the standards by AgMIP is illustrative of the types of revisions made in producing Version 2. Recognizing the need to provide better background descriptions of experiments, new variables were defined in the METADATA subset to indicate the experiment duration, whether it was conducted on a research station or on a farm, and whether it was managed by researchers, a producer or others. Also, descriptors for persons, locations and citations were expanded.

A second example of the revision process is taken from a project by the University of Florida to develop a prototype data management system for the Florida Department of Agriculture and Consumer Services. In reviewing the data to be reported, an agronomist noted that although plastic mulches are widely used for vegetable production, they were not considered in the standards. To introduce a new category of management practices, the category is defined as a group and provided a link to the TREATMENTS group of the MANAGEMENT subset. Individual events need to be described in terms of dates, methods and materials. Thus, the group MULCHES was created with level indicators that linked to the TREATMENTS group. Since mulch types differ in thickness, portion of ground cover provided, and other characteristics, variables were defined for mulching dates, type of mulch applied, mulching application method, and date of mulch removal. Subsequently, mulch removal proved to be a substantially different operation from mulch application, so separate sub-groups for mulch addition and removal were created. We anticipate that the variable list and codes for mulching will be modified as additional field data are processed using the standards. One unresolved issue is how best to deal with organic mulches that are expected to decompose and contribute to soil nutrient and carbon pools since the standards also recognize the group ORGANIC_MATERIALS.

One area for improving the standards concerns documentation of grazing, pest and other types of damage such as defoliation by wind or hail. The sub-group PEST_POPS_EFFECTS under TIME_SERIES of MEASURED_DATA currently defines pest population effects, which includes variables for pest populations (e.g., of unspecified root worms) and for levels of damage in terms of mass consumed, leaf area reduction or other effects. For insects, variables have been defined for specific species and instars, so the potential for proliferation of variables is large. Most likely, damage needs to be indexed by species (if known) and developmental stage of the pest in a manner similar to how soil data that linked to profile layers. However, it is unclear how best to link types of damage to causal species or mechanisms because crops are often affected by multiple problems.

Another area requiring attention is indicators of data quality. Statistical indicators such as the standard error should be associated with values of quantitative field measurements such as grain yield or volumetric soil moisture content. A second example is when data describing crop management practices are based on estimated dates or amounts in lieu of detailed measurements. Fertilizer applications and irrigation amounts are often reported as total season amounts and numbers of applications (e.g., "a total of 150 kg ha^{-1} of N was applied preplant and at two dates after emergence"). While fertilizer type, dates of application, amounts per application, and depths of incorporation might be inferred accurately based upon local knowledge, any such estimations should be documented within the dataset. One strategy is to indicate the completeness of the original data by providing single-letter flags for completeness of data for dates, amounts, methods, or other characteristics.

A related issue is to how to indicate appropriate usage or value of a given dataset. This requires more precise metadata relating to the objectives of individual experiments, measurement protocols and presence of field problems such as weed, insect or disease problems. The value of a given dataset might be assessed by the experimenter or a subsequent user of the dataset. Additional thematic areas where expansions are under discussion include ratoon and perennial crop management, characterization of germplasm (e.g., for genetic loci, growth habit and phenology), socioeconomic data (monetary costs and labor), carbon costs, life cycle assessments, and near-instantaneous measurements (e.g., leaf gas exchange), including measurement conditions and protocols.

Suggestions for additions or modifications to the ICASA standards are reviewed frequently and updates to the standards are made as necessary. Backward compatibility will be maintained by retaining synonyms and presenting these in the master list of standard variable names, which is downloadable from the ICASA site. Comments or suggestions on the ICASA standards may be submitted through the DSSAT site (dssat.net/contact-us) or to the corresponding author of this paper. An associated web site for the ICASA Data Exchange (dssat.net/data/exchange) allows users to store and share datasets.

6. Concluding remarks

The ICASA V2.0 standards for documenting field experiments have the goal of facilitating exchange of information and software tools. Wider use of the standards can help to focus research on science issues rather than on re-formatting shared data and to promote greater consensus in documenting field experiments.

Too often in working with data from secondary sources, one finds that key data on initial soil conditions or crop management are not provided even though the original study likely recorded the data. In the past, incomplete reporting was an accepted practice because researchers lacked efficient mechanisms to manage and store the data and the journals sought brevity in methodology. With improved software and declining costs of electronic data storage, the main obstacle to more efficient management of research data appears to be a lack of consensus on how to organize data. The ICASA standards represent an attempt to promote such a consensus.

We emphasize that the ICASA V2.0 standards are meant to be flexible and dynamic. Given the potential diversity of research applications for datasets, it is unlikely that the agricultural research community would agree to a single digital format or that this would even be desirable. Similarly, new types of variables continually arise that merit inclusion in descriptions of experiments. Nonetheless, the foundation provided by an existing standard is expected to facilitate orderly evolution of such descriptions and promote more efficient use of research datasets.

Acknowledgements

The authors thank the numerous people who have provided feedback on the standards, dating to the IBSNAT and GCTE collaborations. Daniel van Kraalingen and colleagues at Wageningen Agricultural University were important early contributors, Oxana Uryasev developed the first implementation of the standards in a relational database, and Christopher Villalobos is leading efforts to use the standards with JSON under AgMIP. USDA is an equal opportunity provider and employer. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA, the University of Florida, the University of Guelph, Washington State University or IFDC.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compag.2013.04.003.

References

Blanchard, J.L., 2004. Bioinformatics and systems biology, rapidly evolving tools for interpreting plant response to global change. Field Crops Res. 90, 117–131.
Bostick, W.M., Koo, J., Walen, V.K., Jones, J.W., Hoogenboom, G., 2004. A web-based data exchange system for crop model applications. Agron. J. 96, 853–856.
FAO, 2012. Agricultural Ontology Service Project (AOS) [Online]. <www.fao.org/aims/aos.jsp> (accessed 03.01.13).
French, R.J., Schultz, J.E., 1984. Water use efficiency of wheat in a Mediterranean-type environment 1. The relation between yield, water use and climate. Austral. J. Agric. Res. 35, 743–764.
Gardner, J.B., Drinkwater, L.E., 2009. The fate of nitrogen in grain cropping systems: a meta-analysis of 15N field experiments. Ecol. Appl. 19, 2167–2184.
GCTE, 1996. Global Change and Terrestrial Ecosystems Report No. 2, second ed. GCTE Focus 3 Wheat Network: 1996 Model and Experimental Metadata. GCTE Focus 3 Office, Wallingford, UK, 1996.
Gijsman, A.J., Thornton, P.K., Hoogenboom, G., 2007. Using the WISE database to parameterize soil inputs for crop simulation models. Comput. Electr. Agric. 56, 85–100.
Goudriaan, J., 1996. Predicting crop yields under global change. In: Walker, B.H., Steffen, B.H. (Eds.), Global Change and Terrestrial Ecosystems, International Geosphere–Biosphere Programme Book Series #2. Cambridge University Press, Cambridge, UK., pp. 260–274.
Hoogenboom, G., Jones, J.W., Wilkens, P.W., Porter, C.H., Boote, K.J., Hunt, L.A., Singh, U., Lizaso, J.L., White, J.W., Uryasev, O., Royce, F.S., Gijsman, A.J., Ogoshi, R., Tsuji, G.Y., 2011. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.5. University of Hawaii, Honolulu, HI.
Hunt, L.A., Jones, J.W., Hoogenboom, G., Godwin, D.C., Singh, U., Pickering, N.B., Thornton, P.K., Boote, K.J., Ritchie, J.T., 1994. 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, University of Hawaii, Honolulu, HI, pp. 35–72.
Hunt, L.A., White, J.W., Hoogenboom, G., 2001. Agronomic data: advances in documentation and protocols for exchange and use. Agric. Syst. 70, 477–492.
Hunt, L.A., Hoogenboom, G., Jones, J.W., White, J.W., 2006. ICASA Version 1.0 Data Standards for Agricultural Research and Decision Support. <www.icasa.net/standards/index.html> (accessed 10.01.13).
ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), 1984. Proceedings of the International Symposium on Minimum Datasets for Agrotechnology Transfer. 21–26 March 1983. ICRISAT Center, India, Patancheru, A.P. 502 324, India: ICRISAT.
IRI, 2006. ANEEL prcp_sta: Precipitation station data for South America, primarily Brazil <http://iridl.ldeo.columbia.edu/SOURCES/.ANEEL/.prcp_sta> (accessed 03.01.13).
IRI, 2009. NOAA NCDC DAILY GLOBALSOD: Global daily WMO weather station data. iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCDC/.DAILY/.GLOBALSOD/ (accessed 3 Jan. 2013).
Jamieson, P.D., Porter, J.R., Goudriaan, J., van Ritchie, J.T., Keulen, H., Stol, W., 1998. A comparison of the models AFRCWHEAT2, CERES-wheat, sirius, SUCROS2 and SWHEAT with measurements from wheat grown under drought. Field Crops Res. 55, 23–44.
Jones, J.W., Hunt, L.A., Hoogenboom, G., Godwin, D.C., Singh, U., Tsuji, G.Y., Pickering, N.B., Thornton, P.K., Bowen, W.T., Boote, K.J., Ritchie, J.T., 1994. Input and output files. In: Tsuji, G.Y., Uehara, G., Balas, S. (Eds.), DSSAT v3, vol. 2–1. University of Hawaii, Honolulu, HI, pp. 1–94.
Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18, 235–265.
Jung, K.-H., Jeon, J.-S., An, G., 2011. Web tools for rice transcriptome analyses. J. Plant Biol. 54, 65–80.
Kimball, B.A., Kobayashi, K., Bindi, M., 2002. Responses of agricultural crops to free-air CO_2 enrichment. Adv. Agron. 77, 293–368.
Lam, S.K., Chen, D., Norton, R., Armstrong, R., Mosier, A.R., 2012. Nitrogen dynamics in grain crop and legume pasture systems under elevated atmospheric carbon dioxide concentration: a meta-analysis. Global Change Biol. 18, 2853–2859.
McLaren, C.G., Bruskiewich, R.M., Portugal, A.M., Cosico, A.B., 2005. The international rice information system. A platform for meta-analysis of rice crop data. Plant Physiol. 139, 637–642.
McMaster, G.S., White, J.W., Hunt, L.A., Jamieson, P.D., Dhillon, S.S., Ortiz-Monasterio, J.I., 2008. Simulating the influence of vernalization, photoperiod and optimum temperature on wheat developmental rates. Ann. Bot. 102, 561–569.
Nix, H.A., 1984. Minimum datasets for agrotechnology transfer. In: Proceedings of the International Symposium on Minimum Datasets for Agrotechnology Transfer, 21–26 March 1983, ICRISAT Center, India. ICRISAT, Patancheru, India, pp. 181–188.
Ogle, S.M., Swan, A., Paustian, K., 2012. No-till management impacts on crop productivity, carbon input and soil carbon sequestration. Agric. Ecosyst. Environ. 149, 37–49.
Pedersen, A.A., 2004. Entity relationship modeling. Dev Articles. <www.devarticles.com/c/a/Development-Cycles/Entity-Relationship-Modeling> (verified 10.01.13).
Romero, C.C., Hoogenboom, G., Baigorria, G.A., Koo, J., Gijsman, A.J., Wood, S., 2012. Reanalysis of a global soil database for crop and environmental modeling. Environ. Modell. Softw. 35, 163–170.
Shrestha, R., Arnaud, E., Mauleon, R., Senger, M., Davenport, G.F., Hancock, D., Morrison, N., Bruskiewich, R., McLaren, G., 2010. Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature. AoB Plants 2010. http://dx.doi.org/10.1093/aobpla/plq008.
Stackhouse, P/W., 2012. NASA Climatology Resource for Agroclimatology <http://power.larc.nasa.gov> (verified 10.01.13).
The Plant Ontology Consortium, 2002. Conference report. the plant ontology consortium and plant ontologies. Comp. Funct. Genom. 3, 137–142.
Tsuji, G.Y., Uehara, G., Balas, S., 1994. DSSAT version 3 Volumes 1, 2, and 3. International Benchmark Sites Network for Agrotechnology Transfer. Honolulu, HI, University of Hawaii.
Tuomisto, H.L., Hodge, I.D., Riordan, P., Macdonald, D.W., 2012. Does organic farming reduce environmental impacts? – A meta-analysis of European research. J. Environ. Manage. 112, 309–320.
Uehara, G., Tsuji, G.Y., 1998. Overview of IBSNAT. In: Tsuji, G.Y., Hoogenboom, G., Thornton, P.K. (Eds.), Understanding Options for Agricultural Production. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 1–7.
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.
Villalobos, C., 2012. Using JSON & Riak for AgMIP <http://research.agmip.org/download/attachments/1212592/using_json_and_riak.pdf?version=1& modificationDate=1339620113767> (accessed 12.04.12).
White, J.W., van Evert, F.K., 2008. Publishing agronomic data. Agron. J. 100, 1396–1400.

White, J.W., Herndl, M., Hunt, L.A., Payne, T.S., Hoogenboom, G., 2008. Simulation-based analysis of effects of Vrn and Ppd loci on flowering in wheat. Crop Sci. 48, 678–687.
White, J.W., Hoogenboom, G., Wilkens, P.W., Stackhouse Jr., P.W., Hoell, J.M., 2011. Evaluation of satellite-based, model-derived daily solar radiation data for the continental United States. Agron. J. 103, 1242–1251.
XML Core Working Group, 2008. Extensible Markup Language (XML) 1.0, fifth ed. <http://www.w3.org/TR/2008/REC-xml-20081126/> (accessed 10.01.13).