A new field-tested electronic system for data gathering, recording, transfer and dissemination via the World Wide Web

Taciana Barbosa Cavalcanti[1], Abrahão Rezende[2], Roberto Togawa[2], Patrícia Rodrigues[2], Luciano Magalhães Favilla[2] & Goran Neshich[3]

Summary

Cavalcanti, T. B., Rezende, A., Togawa, R., Rodrigues, P., Favilla, L. M. & Neshich, G.: A new field-tested electronic system for data gathering, recording, transfer and dissemination via the World Wide Web. – Taxon 47: 381-386. 1998. – ISSN 0040-0262.

One of the major challenges for taxonomists today is the organisation of collection data by making use of modern computer technology for data storage and retrieval. Recent developments enabling data accessing in the field are here discussed. A field-tested electronic system for data gathering, recording, storage, transfer, formatted printout (e.g. for labelling) and, finally, dissemination by on-line access, called ELCEN (Electronic CEN Herbarium), is described. It makes use of new, easily portable, hand-held "personal data assistants" (PDAs) that connect easily with personal computers (PCs) and workstations, and results in global data availability via Internet's World Wide Web. Among the advantages of the new system are its ease of use, the elimination of most typing errors during data entry, and the reduction of the time lag between data input and data availability.

Plant scientists use herbarium collections to inventory and document the known plant species. The plant taxonomic community has therefore set the organisation of information linked to plant material that has been collected so far as one of its major priorities. The contribution of taxonomists to conservation has been unduly neglected in the past, one of the reasons being the unsuitability of non-computerised, paper-based systems to record, process, and use the data collected in the field (Stuessy & Sohmer, 1996). This is particularly true when it is not feasible to sample a given plant, yet it may be possible to store images of it, and a complete description, in electronic form. Herbaria are increasingly storing digitised images and collection data. Photography and computerised video applications combined with a geographic information system (GIS) are now used to visualise plants and their relation to the physical environment.

Our new field-tested data accessioning system ELCEN (Electronic CEN Herbarium) enables the efficient storage of information as well as its processing and retrieval. It effectively meets the demand for availability of value-added electronic data, because the data can easily be ported onto the universally used World Wide Web (WWW). Electronic publication in a searchable and retrievable format greatly enhances and expedites the communication between taxonomists and users of their work.

Background

The National Centre for Genetic Resources and Biotechnology (CENARGEN) of the Brazilian Enterprise for Research in Agriculture (EMBRAPA) has been using for

[1] Department of Germplasm Collecting, Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Centro Nacional de Pesquisa de Recursos Genéticos e Biotecnologia (CENARGEN), Parque Rural, C.P. 02372, CEP 70849-970, Brasília, DF, Brazil (e-mail: taciana©cenargen.embrapa.br).

[2] Laboratory of Informatics, EMBRAPA-CENARGEN (same address).

[3] Laboratory of Strucutural Biophysics, EMBRAPA-CENARGEN (same address).

the last 15 years a program, written in Basic, to record and organise information collected during field expeditions. This program has been written in non-object-oriented programming language and allows only sequential access to the data. Called "Sistema Coleta", it organises the data in 32 pre-defined fields that correspond to the fields in our paper-based field notebooks. Subsequent data processing suffered from the bottle-neck of data transfer from the notebooks to digital format. Prior to the keying-in process, all data had to be verified by appropriate personnel for consistency and spelling errors, and again afterwards so as to remove any typing errors. "Sistema Coleta" provides for the printing of herbarium labels, listing of exsiccata, as well as consulting and updating of data.

"Sistema Coleta" is now being replaced by an up-to-date solution using equipment and software that are generally available on the world market, to permit not only data recording but also implementation of a relational database with a WWW interface, to enable access to the data not only by our own staff (by intranet) but also by interested researchers around the globe, via the Internet. Our new system provides for on-site data accessing during field expeditions, thus eliminating the need of later transcription from field notebooks. To minimise the risk of keying errors and the need for proof-reading, the system offers pull-down menus with ad-hoc selections of standardised data elements (see below). ELCEN also reduces the time gap between data input and availability to the scientific community to a minimum. Depending on geographic location of an expedition, the time lag between the recording of data and data accessibility may be a matter of hours.

The CEN herbarium considers electronic publishing of the data on the WWW a routine task and also envisages the edition of CD-ROMs. Adding to traditional plant sample preservation, a new form of biodiversity information is thus being provided: electronic data supplemented with photographs and video material – all this with minimal delay between data gathering and availability to the general public.

The number of accesses to any particular database and/or WWW site likely depends on how interesting the data are and on how well they are advertised. Global access to data is not only becoming an essential means of information exchange, but also brings credit to the work performed when its merits are judged based on the number of Internet accesses.

The new system we propose and have tested is a most promising educational tool. Following the pathway from initial recording of data to publication on the WWW, is an excellent way to interactively teach taxonomy – e.g. by correlating the distribution of taxa, and means for their preservation, with environmental parameters – and to demonstrate how to catalogue biodiversity.

System-design and implementation

Most systems for electronic data recording which have been designed in the past are either too cumbersome or too complicated to be easily implemented for our purposes on a long-term basis. The technology of hand-held computers and the correlated software for recording and organising data is well suited to satisfy our needs.

The ELCEN system design was based on the premise that paper-based devices, while useful for reference purposes, are inconvenient to carry in the field and unwieldy for transmission to others. Personal data assistant (PDA) technology makes it possible to record data electronically in any environment and to store the contents of

a large field note-book on a small, light-weight memory card about the size of a credit card. The advanced technology of PDAs makes them much easier to use for the technically untrained than older equipment that required extensive background knowledge.

In order to span the whole range of processes from data recording in the field to information availability via the WWW, the system requirements are complex enough to necessitate a thorough planning of hardware and software choices, as well as field tests.

Hardware. – With respect to hardware, we have chosen to use the following items of equipment:

- hand-held computers (PDAs) for data recording in the field (see below);
- a lap-top computer for daily data storing, to improve data security on long expeditions;
- a photo-camera for close-up images, with a circular flash, for traditional photo-documentation;
- a video Hi-8 camera for video-taping;
- a global positioning system (GPS) for the identification of locality co-ordinates;
- a digital photo-camera for fast photo-documentation;
- a universal desk-top computer station for data processing (PC- and Mac-compatible);
- a laser-printer suited for the printing of herbarium labels;
- a Unix workstation with advanced capacity of graphic data manipulation;
- a CD-ROM writer.

ELCEN data input builds on the availability of a hand-held computer (PDA) with sufficient memory to hold the daily data and able to interface with both a PC and a GPS. The PDA must be compact and sufficiently robust to withstand transport by hand or attached to a belt under field-work conditions in rough terrain and at high temperatures. The power supply must be light and last for 4-12 hours of use. We find the Newton Message Pad (110, 120, 130 or 2000) of Apple Computers, Inc. (Anonymous, 1995), a suitable choice. It weighs just 635 grams, provides three to six days of battery life based on typical usage, instant boot-up, and it is easy to learn its use. Its simple LCD display, easily programmable interface with pen-based input, and ability to interpret hand-writing are additional positive features.

We are currently working on a system to transfer data from a GPS receiver to a Newton PDA using a PCMCIA card, which slots into the Newton PDA, so as to feed in geographical co-ordinates directly, without user intervention.

Further data processing is done after data transfer via the Ethernet, using the TCP/IP protocol, to one of our workstations. This is the second data transfer, after the download from the PDA to the lap-top. WWW publishing is done completely on workstations, as is the data editing for the laser-printing of herbarium labels of a standardised size.

Bar codes will be generated and printed on herbarium labels to assist collection management. A bar code reader will be connected to a Pentium PC station.

Software. – Functionally, software used for ELCEN can be divided into three basic categories: data input and user interface generation software; data transfer; and electronic publishing.

Data input and user interface generation. – Software requirements include the storage and handling of potentially large quantities of data and the ability to display a frame for daily data entry that is easily understood. The user interface of the Newton PDA fulfils these requirements. In order to minimise confusion during data entry, an interface with sequentially arranged data fields in a format that even the inexperienced can easily use has been designed.

An additional factor we had to bear in mind is the limited amount of memory available for storing data. Compact data structures had therefore to be developed. Some data fields, such as a "comments", are of variable length. Our WWW site [http://www.cenargen.embrapa.br/rec_gen/herbarium/photos.htm] depicts some of ELCENs interface screens for data accessing (Fig. 1).

We used the commercial package Leverage for user interface generation and maintenance in the Newton PDA environment. This software allows free programming of individual database formats and field descriptions. We designed a menu-based system for data entry in order to minimise typing errors when data are entered under field conditions. Fig. 1 shows a Newton screen from an ELCEN application developed for data annotation using the Leverage database generator. Whenever possible, data entry is by means of choices from pre-defined menus, which increases efficiency and reduces error rates. The pen-based input of Newton PDAs provides for easy data entry through multiple-choice pull-down menus. However, some data are of a complex nature and have to be entered on-site in form of sequential text (e.g. comments) or sketches (e.g. maps). For entering such data we tested two options: the screen keyboard provided by Newton's operating system 2.0, or Graffiti SM (Anonymous, 1994a) software for handwriting recognition. Both methods gave satisfactory results after a short training.

Data transfer. – The Leverage package has an option for data transfer from a Newton PDA to a PC using a serial cable. Once transferred to our lap-top computer, the data are further formatted and converted to the desired font using text and desktop publishing software – e.g. for formatting and printing herbarium labels. In a Windows 95 environment, the data are transferred to a database structure (MS Access). Transfer of digitised images from video-tape has been achieved using Iris capture version 1.3.

Electronic publishing. – We used a number of commercially available software packages (3D studio Max, Netscape gold, Photoshop, Corel Draw) for making our data publicly available on the WWW in appropriate text, visual, and photographic presentation. A large, indexed database is thus available [gopher://gopher.cenargen.embrapa.br/0/coleta/coleta]. Another page describes our technology of data collecting and sample identification [http://www.cenargen.embrapa.br/rec_gen /herbarium/scan.htm].

We have furthermore assembled a compilation of video films [http://www.cenargen.embrapa.br/rec_gen/herbarium/movies.html] that includes short movie sequences from germplasm collecting expeditions.

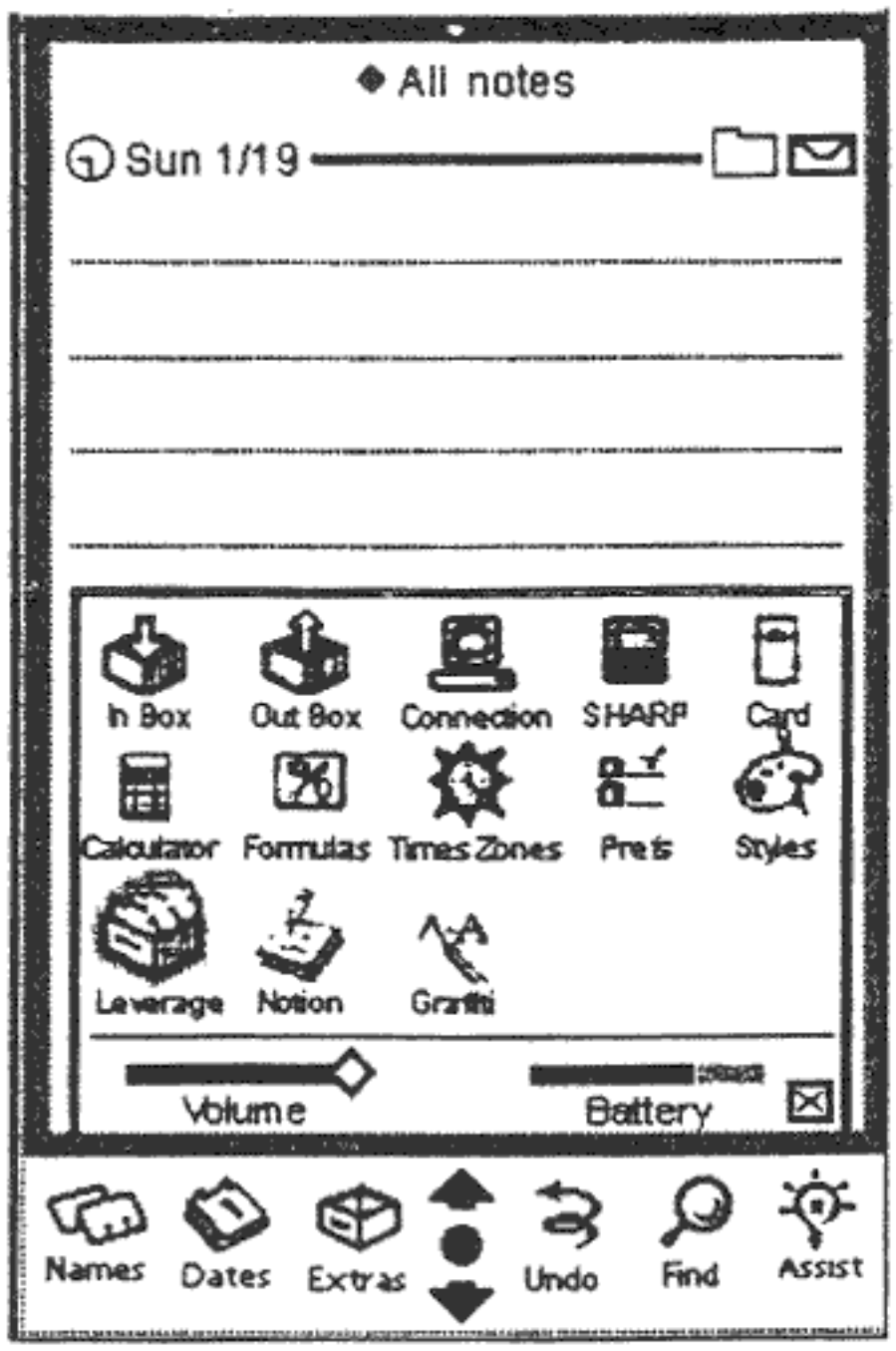

Fig. 1. Some of the data fields as displayed on the Newton Message Pad screen. – A, first screen for selecting the desired software, e.g. the database Leverage; B, to enter data in the field, a keyboard can be displayed on the screen by double-clicking the respective field; C, data entry has been initiated for the field "collected by" (now hidden); a pull-down menu offers the names of those participating in that particular expedition, and a simple click on one of the names will place it in the "collected by" field; D, display of a list of all registered entries.

Discussion

There are several aspects of ELCEN that are of general relevance:

- our system demonstrates that advanced technology can facilitate the recording and further processing of data, making them generally available;
- it provides a computer interface for easy data entry under field conditions, minimising the input error rate;
- it uses a flexible database design that can be extended by the user with minimal professional assistance by means of low-cost commercial products, yet its features are complex enough to satisfy the needs of most users and of any particular project;
- it successfully integrates diverse equipment and software platforms: PDA, lap-top computers, Windows and Macintosh operating systems, Unix workstations, relational database features, and WWW publishing including multimedia components; careful attention however is needed to ensure efficient and seamless intercommunication of these systems.

Future development. – We are preparing a graphics presentation that integrates information such as taxon distribution and its relationship with geographic features such as rivers. Users will be able to view locality information plotted on maps by a geographic information system (GIS) for any type of distributional data. It is expected that ELCEN will soon join efforts with the Brazilian Information System for Genetic Resources (SIBRARGEN; Anonymous, 1996) for database generation and maintenance, using the U.S. Department of Agriculture's GRIN model (Anonymous, 1994b).

We also plan to implement the Newton Message Pad 2000 platform; to complete the control of exsiccata flow by using bar code readers; to produce a CD-ROM; to have all CEN herbarium data tailored not only for reference needs but also for interactive multimedia teaching in the area of biodiversity cataloguing, germplasm collecting, environment protection, plant taxonomy, etc.

Literature cited

Anonymous, 1994a. *Graffiti: user's guide. Recognition that works.* Palm computing, Inc. California.

Anonymous, 1994b. *GRIN: user manual. GRIN TRAINING agenda.* Beltsville.

Anonymous, 1995. *Newton, Apple Message Pad Handbook.* Apple computer, Inc. California.

Anonymous, 1996. *Sistema Brasileiro de Informação de Recursos Genéticos. Dicionário de dados do SIBRARGEN.* EMBRAPA-CENARGEN. Brasília.

Stuessy, F. & Sohmer, S. H. 1996. *Sampling the green world - innovative concepts of collection, preservation, and storage of plant diversity.* New York.