

The Future of Climatic Data Management Systems

A Trip Report by P. Muraya

11 May 2000

There was a meeting of 15 data management experts – 2 from Africa (including myself), 1 from Australia and the rest from US and Europe – held in Geneva from 2nd to 5th may. The purpose of the meeting was to propose how the World Meteorological Organization (WMO) should proceed to meet the needs of a variety of its member countries who want to take advantage of the latest technology to manage their climatic data. Management of climatological data from national meteorological services is well ahead that of agricultural research data. The reasons are clear: the data are collected to meet the needs of future generations for reliable, homogenous, national climatological data; there is a long established tradition of sharing the data because all countries stand to benefit from the resulting improved forecasts. I was invited to this meeting for 3 reasons: 1) it is known that ICRAF, through the Logbook project, is heavily involved in using the latest technology to develop a data management toolkit for agricultural research; 2) I was a trainer at the **Climatic Computing** (CliCom) evaluation and training workshop held at the African Center for Meteorological Application and Development (ACMAD) in Niamey last year; 3) there is growing sensitivity to the need to link climatic data and related environmental data seamlessly, as more and more advanced predictive tools are developed.

A number of countries do not need assistance to take advantage of the latest technology in climatic data management because they already have well-developed systems. Some of them are offering to share their systems with member countries. There is a second group that is successfully using an existing system that was developed in the 1980's. They are crying for modernization. Zimbabwe is one of them; they are in a unique situation because they are the only country in Africa that – following the ACMAD workshop – have developed a system (using ICRAF's Logbook tool kit) that they are willing to share with member countries. There is also a third group that is using nothing. They are in need of a basic, easy to use system that requires minimum resources to purchase, operate and maintain. This is the group that does not have any formal way of organizing their data and exclusively uses spreadsheets to store them. ICRAF shares the same problem with such organization: and there lies our main reason why I was allowed to attend this meeting.

The challenge before the meeting was to decide on the **strategy** to use for testing the offered systems and to prepare an implementation **plan** that addresses the needs of those countries with no operational systems as well as those wishing to improve their existing systems. We did much on the testing **strategy** and little on the implementation **plan** – or at least I understood very little of it, this being very much a function of the agreement between the system donor, recipient, and WMO. So I report on the strategy.

We developed the evaluation criteria document, specified the characteristics of a common dataset to be used in the tests and specified the testing method.

Evaluation criteria

To develop the evaluation criteria we partitioned the data management system into 8 functional modules. Using some 11 general user requirements that had been specified in earlier CliCom meetings we identified performance indicators for each of the requirement in every module. Consider for example, the following user requirement: that the system has to be user friendly. For the data entry module (which is lacking in ICRAF's Logbook system), we needed to specify for user friendliness such indicators as need for direct mapping of paper form to computer form, presence of validation checks, field masking and provision of default sensible values, online help etc.

There was a requirement for the cost of operational and maintenance to be reasonable given the size of the data management task. This clearly would discourage the use of huge systems offered by the developed nations in the developing ones. The French and the Australians are countering this effect by developing scaled down versions of their systems to run on the Windows platform. This led to the requirement: offered systems have to allow for evolutionary development, be easily upgradeable, and use open source as far as possible.

There was a requirement that the system has to be adaptable to national preferences. This meant that if, for example, the French system is developed with this requirement in mind, then to move from a French to an English speaking country should require only a simple switch. So far, only the system offered by the Czech republic has this capability. Plans are already under way to implement it in Ghana and Macedonia.

Test dataset

The meeting suggested the composition of a test data set to complement the evaluation criteria document. (This was a familiar route: the distributed version of the Logbook is accompanied by a test dataset designed to illustrate how the system operates). The management of data from automatic weather station was very low in the requirements of most stations in Africa. Most inputs considered were derived from well-established data exchange formats on the Global Telecommunication System (GTS). Participants offered to send to WMO typical data sets selected from their central databases. ICRAF could not. The usual reasons: it is difficult to know whether any relevant data exist or not; it is not easy to access and retrieve it as we do not have central repository; and there is no clear policy about sharing it for such external applications. That's a pity, because we were the best-placed institution to present test data sets that exercise the "sensitivity to other related environmental data" portion of the user requirement.

Testing method

There was a previous idea of carrying out the tests at the WMO headquarters, but this was deemed too expensive. Instead a self-evaluation by the system developers was proposed. The evaluation criteria, the test data set and a list of required products will be packaged by WMO and sent to all member countries who offered to share their systems with other members, asking them to import the data into their systems and run the required products. The results will be sent to WMO, together with completed description of minimum and recommended requirements for hardware, commercial software including licensing, training, human resources etc. WMO will then hire 2 consultants to analyze the results of all the self-evaluation tests, compile a report and publish it on the Web for potential users to access.

What does this mean for us?

More work on the Logbook with increased potential to help data management, not just climatology, in many more African countries. Zimbabwe is a good example. It is becoming increasingly clear that a direct key entry module is important for quality controlled capture and preservation of data. If a true generic data entry module is added to the Logbook toolkit then this functionality will be useful to not only our researchers but also to our collaborators, especially those needing to integrate climatic and other data in agricultural, health or environmental applications.

The integration of data is becoming increasingly feasible. There is now general agreement that almost all climatological data has attributes of element (i.e. type of measurement), location of observation and date/time of observation (and for non-instantaneous elements, duration of observation). The Logbook generalizes this by expressing every observed value in terms of what, where and when – making it feasible to integrate data from different disciplines. Steve Palmer, of the UK meteorological office predicts that all one day, perhaps in 10 years time, all climatic data will be managed using the "Logbook" data model – when mass storage consideration will be inconsequential and the need to perform complex simulation using data across disciplines will become paramount. He supports further development work on the Zimbabwe system that is currently using a fully normalized data model – similar to the Logbook. Here there is real incentive to get on with the data entry module as ICRAF's contribution to the broader data management problem.