ANNEX 6

Some views of Steve Palmer on database models

The UK Met. Office developed a new climate database from first principles over the last six years. The design was done in LBMS SSADM Engineer so that it was portable to any commercial RDBMS. The database was first implemented in CA-IDMS, but has recently been ported to Oracle. There is interest in the database team in trying the port to MS-Access, so I am hoping that will be tried very soon.

I think that the idea of separating the logical design from the physical implementation is important for NGCDMS, also the use of commercial tools. Similarly the toolset used for querying the database need not be those provided by the supplier of the RDBMS (e.g. we use MS-Access and VB to build user queries against an Oracle database, because the MS-Office report design tools are much better. We also use other tools such as GQL where these are better).

Outcome from the ACMAD workshop: Logical design of a climate database splits into two areas, the metadata, and the main data tables. There is very little difference in the metadata design however you start. The main data tables are the interesting area, because of the potentially huge size. There is general agreement that almost all climatological data has attributes of element, location of observation, date/time of
observation (and for non-instantaneous elements, duration of observation).

The original CLICOM design assumed that there was always a periodicity - i.e. one observation per day across a month, but this is not a useful assumption, and makes it very difficult to extract data, and also to store data where duration is irregular (e.g. rainfall measured on working days only).

That leaves two possible designs, one which has total flexibility, by storing each datum with all its attributes including element in a single record - Peter Muraya's Logbook takes this approach, and it was considered for the UKMO design. The disadvantage is the huge overhead in storing all the extra attributes and indexes, and the huge size of the (single) data table. Another problem is ensuring consistency of definition of the elements, and stopping these from proliferating.

The third design is that in the UKMO MIDAS database, to store observation elements in fixed, named entities grouped in tables where the location, date/time (and period) attributes are common. This gives optimal storage volume and simplicity in building queries and reports. The disavantage is that adding a new observation element requires a change to the database model.

Hence one gets into arguments about the boundaries of the 'required system' - should climatology cover all the possible observations within the area of environmental monitoring, or should the scope be limited to those elements which are required for the business of the particular NMS?