C-squares – a new method for representing, querying, displaying and exchanging dataset spatial extents

Tony Rees

CSIRO Marine Research, GPO Box 1538, Hobart 7001, Australia

E-mail: Tony.Rees@csiro.au

Conventional metadata records (dataset descriptions) are intended for resource discovery and documentation. For spatial datasets, typically the dataset spatial extent (footprint) is indicated by a bounding rectangle (= bounding box) which thus forms a type of spatial index, and also supports spatial (geographic) searching using an overlapping rectangles test with an enquirer's designated search rectangle. However for many marine datasets (and for various terrestrial datasets as well), such representation is frequently a poor fit to the real data distribution. In the marine context, this can result from the nature of vessel-based sampling in particular (based around linear or irregular voyage tracks rather than filled rectangles) and the fact that significant unsampled regions (e.g. land, or unvisited areas of ocean) can exist within the dataset spatial boundary. The presence of substantial areas within the bounding rectangle which are devoid of data, for whatever reason, renders it likely that any list of hits resulting from a typical overlapping rectangles search will include an unknown number of false positives, where data is notionally present from the enquirer's area of interest but on closer inspection proves not to be so.

C-squares (concise spatial query and representation system) is a recently described (2002) method introduced to overcome this limitation for representing dataset spatial extents at the metadata level. Using c-squares, complex dataset footprints are represented more exactly, while at the same time not requiring the overhead of a GIS application to represent or query the dataset footprint. C-squares supports improved spatial querying as compared with bounding rectangles searches; rapid representation of the dataset footprint on a map of any portion of the earth's surface using an on-line utility (the c-squares mapper); and, since the representation is by a string of ASCII characters, the footprint is well suited to storage as metadata, and to metadata exchange as required (e.g. via XML).

Using c-squares, dataset footprints are represented by aggregations of squares (tiles) at a selected resolution such as 0.1 x 0.1 degrees, 0.5 x 0.5 degrees, 1 x 1 degrees, each of which is labelled according to a notation derived from subdividing 10 x 10 degree WMO (World Meteorological Organisation) squares, with the resulting set of codes assembled into a simple character string. Spatial searching then merely comprises looking for an agreement (text match) between any portion of the dataset's c-squares string and the enquirer's designated search area (itself represented by the equivalent c-squares code or codes).

This poster presentation gives further information about the basic principle of c-squares, together with examples of oceanographic dataset footprints as represented by this method, from the data holdings of CSIRO Marine Research (CMR). An example c-squares spatial search interface, for CMR's metadata system MarLIN, is accessible on-line at http://www.marine.csiro.au/marlin/csq-chooser.htm. Additional documentation on the c-squares system is available via the c-squares home page at http://www.marine.csiro.au/csquares/.