DATA MANAGEMENT TO INFORMATION PORTALS

L. Belbin
D. Connell
D. Watts
Australian Antarctic Data Centre
Australian Antarctic Division
Kingston, Tasmania, Australia
Dave.Connell@aad.gov.au

ABSTRACT: Scientific research is driven by data. Data are valued by the scientific community for their ability to prove or disprove hypotheses, but can also be valued by the cost of collection and processing. On the latter scale, Antarctic data are extremely expensive. It is therefore anomalous, that up to the last decade, little thought has been given to preserving these data in a form that facilitates discovery and effective use in perpetuity.

Until the Internet, the primary repository of scientific information was the library. One had to get to a library physically, then use the catalogues to find relevant information. Now, vast volumes of data are freely available from the Internet, but the major difficulty is actually finding the relevant information. Web search engines such as Google are highly effective, but not sufficient by themselves to address practical information sharing in science.

While everyone knows that the library is a repository of publications with a catalogue, the same has not been true of the data that generated the publications. Scientists have assumed that they knew who was doing what, and that data could often be obtained via the author. The web has clearly demonstrated that there is more, relevant information out there than most astute observers could have imagined. The increasing emphasis on multinational and multidisciplinary research also highlights the need to expand our horizons.

The Antarctic Treaty is a document with amazing foresight. Article III.1.c of the Antarctic Treaty states that, "scientific observations and results from Antarctica shall be exchanged and made freely available." The Australian Antarctic Data Centre (AADC) was established in 1995 as the Australian node within an international network of National Antarctic Data Centres. Our strategies for data and information management have evolved with the web and emerging standards.

Compared to a library catalogue, our metadata are more like an abstract. They contain fields such as spatial and temporal extents, data quality, format and usage constraints. While libraries catalogue and manage available material, our data centres have the primary problem of educating the scientific community to a level where data management is valued as an integral component of quality scientific research. We have therefore developed a 'cradle to grave' strategy. For example, the AADC wrote the first web-based science application proposal system so that information about proposed research was available even before it was approved. Now, metadata are automatically generated from approved proposals, while data must be submitted to the AADC within two years of collection. Failure to submit data can result in future funding rejections. We believe this is the first instance internationally of enforcement of such a policy, even though such science policies exist in many countries.

Recently, the AADC has developed an "Enterprise Information Portal" (EIP). This portal provides a single, personalised gateway to our users, from which they can search and access internally and externally stored information. The portal requires only a browser and Internet connection, and can thus be used regardless of the user's software and geographic location. The portal also simplifies administration by allowing access to certain information to be restricted if need be, and tracking of user activities. In effect, an EIP is a one-stop-shop for access to information and administration applications.