

It Is on the Web

"It is on the Web" has become one of the most common responses that I hear whenever I ask for information about almost anything. The only problem with this has to be where on the Web? Typing the word 'Hydrography' into Google produces about 152,000 hits. 'Hydrographic' is even more voluminous, with more than 282,000 hits. This illustrates the tremendous volume of data that is available to us now.

Data overload is a very real problem. It is not only quantity but also quality that is important. Whilst I can only applaud having so much information freely available, I must admit that I am also very overwhelmed by the volume and have great difficulty in verifying quality. It is a fact of life nowadays that very poor data can be made to look very good (i.e. attractive) simply by slick presentation. Equally, there is some excellent information that is being avoided by many professionals simply because the presentation is poor.

I think that one of the challenges facing us as an industry is to find a common definition for quality and then apply this to the data available on the market. 'Retro-fitting' of quality criteria is such an expensive and time-consuming effort that I fear it will never happen. I know from experience gained working in the arena of data management that often it is only the 'old timers' who can apply the necessary quality flags to certain types of data. I know that many of the 'newer professionals' have never worked with 'old-fashioned' positioning systems such as Decca Mainchain, HiFix or other hyperbolic systems. Even the very concept of 2-range systems is now virtually obsolete, but probably more than 60% of available historical hydrographic data is based on these types of systems. For many parts of the world, data are based on sextant-derived Lat/Longs – how can we apply quality flags to this sort of information?

Merging of these very disparate datasets is a common way to obtain a comprehensive overview but this can be very dangerous. History has often shown us the folly of making decisions on inadequate intelligence. Using the Web as a tool for data mining is just another form of gathering intelligence about an area or location, but one that is only as good as the quality of the information gathered.

Maybe it is naïve of me to want to have a quality flag for all data. But there is a very real danger of people assuming that all electronically available information must be correct. I am not advocating going out and reacquiring worldwide hydrographic data coverage, but simply raising the warning flag: just because data is available on a computer screen does not mean that it is accurate. The Web is a marvellous tool and a huge advance to all of society, not just Hydrographers, but the information available is only as good as the original observations. One of the early computer terms was GIGO. (Garbage In = Garbage Out) - that phrase is still valid today.