Source Data Management

The potential for hydrographic information in the holdings of today's Hydrographic Offices (HOs) extends far beyond its classical use: the production of charts and nautical publications to ensure safety of navigation.

Historically, HOs have collected far more data and information than has ever been published in charts and nautical publications. Since the establishment of the first Hydrographic Offices more than three centuries ago, a sensible archiving system has been sufficient for the management of source data. However, the move into the digital era and extension of the mission and objectives of HOs are changing the requirements for and possibilities of management of information sources. What is more, many HOs today receive data from sources beyond the positive control of the receiving HO, adding a further dimension to the source data management problem: assessment of the authenticity and validity of information.

The Modern HO
The (generic) mission of the HO is to provide adequate and timely hydrographic data, products and services and ensure their widest possible use. The objectives of the HO are to:

- promote the use of hydrography for safety of navigation and all other marine purposes
- improve coverage, availability, quality and access to hydrographic data, information, products and services
- use international standards for the quality and formats of hydrographic data, information, products, services and techniques
- provide authoritative and timely guidance on hydrographic matters for government agencies and institutions, and co-operate in hydrographic activities amongst regional States within the framework of the International Hydrographic Organisation.

While the top priority is still to ensure safety of navigation, bullets 1, 2 and 3 point to the other uses of the underlying information on charts. HOs all over the world are giving increased attention to these areas, and this is reflected in the new mission and objectives of the International Hydrographic Organisation (IHO) as recently agreed by IHO Member States in April 2005.

The Challenge
Modern HOs receive thousands of communications every year, many of which contain relevant information about the area that they have a responsibility to chart. These communications can be termed Source Data Documents (SDDs) and examples are maps, charts, aerial photographs and satellite imagery, photographs, documents, reports, books and manuscripts. In fact, any human communication about the physical and social world may be considered a SDD. The challenge is to keep track of, evaluate and make timely use of the hydrographic data contained in these SDDs, and to process them in accordance with the objectives stated above. HOs have traditionally done this manually through cataloguing, indexing and visually assessing paper versions of such SDDs. The increasing volumes of data being generated mean that data-centric methods have to be devised and used to cope with them.

Database-driven
Over the past twenty years T-Kartor has developed and implemented a database-driven approach to cartography, map-making and maintaining a geographic or hydrographic database, and this has now been extended to source-data management. The overall flow includes registering SDDs into appropriate database tables. Registration involves assigning a unique identifier to the SDD and deriving metadata such as author, publication date and place, thematic content and other elements of bibliographic data. Processing can include many steps, such as receipt of source data, translation (if needed), geodetic evaluation, datum conversion (if needed), geo-referencing, identification of other source information in the same area, distribution to cartographic specialists, assessment and updating of products, and archiving.
Then comes the most critical step: SDD Evaluation by Cartographic Specialists and, as shown in the diagram, this is the last step before updating of the Hydrographic Database. Part of this evaluation is assessment of the urgency of the information: is it critical enough to warrant immediate distribution to users, or can it be held back until subsequent systematic update of the products? If urgent, the information can be broadcast as a radio navigational warning before inclusion in Notices to Mariners.

Questions Arising
A number of questions must be asked and answered before a Radio Navigational Warning is issued and an update to the Hydrographic Database performed. Is the new source more reliable than the original one? If the source is a new survey and it shows more water than originally believed, can we risk an update or should we maintain the shallower depth? What has happened to the wreck? Has the surveyor missed it, has it been covered by sand, has it disintegrated? The list of potential questions is long. It is apparent that time spent on the evaluation is critical and whilst it needs to be kept as short as possible, quality must not be impaired. It is here that the database-driven approach finds its true potential, allowing the evaluator to search quickly through all available information relevant to the geographic area in focus. Evaluation of SDDs can result in updated information becoming a part of the Hydrographic Database, as well as deciding the impact of new SDDs on existent HO products. The strength of the database approach becomes apparent when it is understood to include publishing all products from one consistent HDB based on systematic analysis of all relevant SDDs.

Case Example
In the sample case the input is a paper chart from the Japanese HO. Here the Source Data Document (SDD) is analogue, which means that the chart has to be scanned and all metadata information collected and entered into the Source Data Management system (SDM). The majority of the information for the input of a foreign chart into a HO production system is more or less metadata, as would be expected

from any geographic dataset. Of particular interest is the assignment of two unique identifiers: one for the source data document and one for a Source Data Evaluation project (SEP).

Unique IDs are essential for the tracking and future relocation of the SDD, assessments carried out and action taken, which are recorded in the SEP. The SEP is the vehicle for collating all information pertinent to the evaluation and containing the justification and outcome of assessment; it is explained in more detail below. Another essential part is the generation of a geographic footprint, which may be any geometric primitive, in this case, an area. The geographic footprint will enable future queries to the underlying database to answer questions such as, which SDDs are in the area covered by my product and/or what assessment took place, by whom, and what was the action taken?

Evaluation: the SEP

An important part of the overall concept is the Source Evaluation Project (SEP). The SEP is the logical container of all assessment activity undertaken by the analyst. This can include linkages to products affected by a particular SDD, emails sent out to clarify SDD contents, and translations of the whole or parts of an SDD. The SEP becomes the working unit for further assessment of the information by the cartographic specialist, deciding upon what action is to be taken and recording the outcome of the assessment and the actions taken in the SEP.

This process allows other organisations to browse stored data for information on data holdings through the SDDs and subsequently request the information. It allows the HO to do the same thing, but through the SEP the HO can go back and examine past evaluations. Ultimately the SEP will enable the HO to track source data right down to the level of individual objects within products.

Added Benefits

An added benefit to an organisation choosing a database-driven approach to source data management is that it enables queries to be carried out systematically and fast, not only through key wording and traditional indexing, but also by geographical search. An example is the generation of a new edition of a chart. One important aspect of this process is to ensure that no source information is overlooked. Searching geographically for all SDDs in the area to be covered, with a reasonable buffer around the area, will ensure the retrieval of all SDDs relevant to that geographic area. This removes the feeling of uncertainty experienced by many cartographers when going through this process: have I overlooked potential source information?

Compare the database-driven approach to the more traditional approach, where a reference chart is annotated with information on incoming sources. The reference charts have to be reviewed and original source documents may have to be retrieved for archive and review. The power of databases in carrying out fast, complex searches offers significant time saving and, provided SDDs are held digitally, will make available all information relevant to the new edition of the product at the cartographer's work station. There will be a simultaneous sense of security that that no potential source has been overlooked - an often underestimated factor in the revision process.

Other benefits of the database-driven approach to source data management is that metadata captured and the geographic footprints generated enable the running of a Web service without the need to manually search the archives in answer to queries.

Further, it reduces the not uncommon risk of duplication of data-capture efforts, due to lack of readily available information on hydrographic source data holdings.

Summary

HO data holdings contain far more information than is published in navigation products. HOs are increasingly aware of this, and face the challenge of informing about and providing access to such information whilst at the same time increasing the speed of processing new information into new or updated navigation products. Many HOs are further faced with the requirement to maintain or even decrease their operating costs, and obviously no HO wishes to compromise on quality.

Exploiting the power of a database-driven approach, from handling and managing incoming Source Data to updating a central Hydrographic Database and creating products from this single, consistent Hydrographic Database, offers HOs a chance to rise to new challenges.