Virtualising large digital terrain models

G.G. Spoelstra
ATLIS, the Netherlands

Abstract

Chart producing agencies like Hydrographic Offices face major challenges today in keeping up with the ever-growing amount of data that is produced by modern echo sounders. At the same time users are demanding products faster and yet more reliable to keep pace with the more powerful and accurate systems they use. To support Hydrographic Offices and other chart producers, new innovative technologies need to be developed that will bring a great deal of efficiency into the current survey to chart work processes ensuring better products and a much faster time to market. Above challenges can be met by the introduction of V*irtual Continuous Model (VCM)* technology. The ATLIS SENS (Spatial Enterprise Nautical Solutions) Bathymetry product is utilising this technology in such a way that data only has to be stored once; all users use the same data and can define their own terrain models without the need of copying the often large volumes of data to an individual workspace. The number of models (both up-to-date and historical) is virtually unlimited thus providing a maximum of flexibility to expert organisations. Using SENS Bathymetry, large digital terrain models are virtualised by just storing model definitions based on available meta data instead of redundant storage of the data itself. The underlying Oracle Spatial technology that holds the archive of survey data ensures fast retrieval of seamless models that can be used for a wide variety of applications. Using proven database technology is a great advantage over technologies that are based on traditional file systems which are reaching their limits on today's operating systems.

1. Introduction

Hydrographic offices face increasing pressure to deliver products faster, yet more reliable than ever. With source data coming in at growing speed and volume, bathymetric data management is changing into a major bottleneck between data acquisition and chart production. At the same time users are demanding products faster and yet more reliable. Organisations that manage bathymetric data often use the data for various reasons: safe navigation, morphology, off-shore planning to name a few. The challenge with managing large volumes of bathymetric data is to keep everybody happy without the need to (re)build and manage models for each of these communities. A nautical cartographer requires a navigational safe model that is as up-to-date as possible whilst a marine morphologist might be interested in a series of historical models to analyse sediment transport. The problem gets even more complex as most organisations are under the obligation to archive their bathymetric models for liability reasons. In addition they also have to make data available as part of national and international data infrastructures.

Mandatory ECDIS (Electronic Chart Display and Information System) carriage requirement, announced by the International Maritime Organisation (IMO) will force ship owners into investments that are easier to accept when the Electronic Navigational Charts fuelling ECDIS provide a much richer and more up-to-date content than today's paper charts. Content that is derived from seamless databases avoiding the traditional paper chart patchwork and offering high density bathymetric information to the end user in a smooth and unambiguous way. The size and draft of modern vessels make it almost impossible to bring these ships into ports without the use electronic chart systems showing go/no go areas with sub-meter detail. Demand of contour intervals of less than a meter in electronic charts is therefore no longer an exception but require advanced algorithms and high performing technology.

In section 2, this paper will look into the changing demands for high density bathymetric data and explains why today's production processes of hydrographic organisation form major bottlenecks when it comes to building large digital terrain models compiled from many different sources. The associated challenge is illustrated by describing work of marine morphologists and the liability issued faced by National Hydrographic Offices. The concept of the Virtual Continues Model as a solution is discussed in section 3. Section 4 highlights some future requirements for further improvement of handling massive point clouds using modern data base technology.

2. Changing demands

2.1 Different user needs

Bathymetric information is collected for various reasons. Safe navigation is probably the most prominent one but dredging, environmental protection, morphology, engineering, off-shore explorations, archaeology, fishery and coastal zone management are other examples of applications that require a detailed and up-to-date picture of the seabed. Satisfying all these users requires careful planning of the bathymetric workflow. Laying pipe lines using bathymetric data prepared for safe navigation is not a wise thing to do, investigating environmental impact at a world scale using high density multi beam surveys a very costly and time consuming exercise.

2.2 New demands for bathymetric data

Most organisations responsible for collecting bathymetric data usually performed their task with a relatively narrow scope. Hydrographic offices are responsible for navigational charts, survey companies performing route surveys only collect their data to check the status of the pipeline or cable. Introduction and growing attention for national and international data infrastructures have created a new data awareness amongst data producers. Suddenly they not only collect the data for their own prime task but have to make the data available to a wide range of users. A resent example in the hydrographic domain is the European EMODNET project [1] that will make bathymetric, chemical, physical and biological data of the European sea basin available to the general public. Frameworks underlying data infrastructures not only specify the why of these initiatives, they also specify the how by defining standard formats and protocols. Data producers now see themselves forced to comply with these frameworks and this requires a different view on the management of their data for more and different users.

2.3 The data management bottleneck

Today, software development in relation to modern multibeam echo sounders is primarily focussed on processing and analysis of (backscatter) data. These tools (e.g. Fledermaus from IVS-3D) provide a wealth of functionality that operate on the source data. Semi-automatic data cleaning and generation of surface hypotheses using the CUBE (Combined Uncertainty and Bathymetry Estimation) algorithm are just a few. CUBE is an error-model based, direct DTM generator that estimates the depth plus a confidence interval directly on each node point [2]. Most of these tools have a 3-D user interface (see Figure 1).

To face the ever increasing volumes of data, hydrographic organisations have invested in new tools to convert the raw sensor data into manageable work sets by applying some form of data reduction technique. Major characteristic of these tools is that they operate on survey files. For project based work (e.g. the processing of a single survey or the production of a paper chart) this is often sufficient but the limits of bathymetric file processing are coming closer. Not because (64 bit) operating systems can't handle very large files but the storage management of the data and the interaction with the stored data itself becomes a problem. To maintain acceptable read/ write access times for very large files advanced index and partitioning technology becomes inevitable. Bathymetric file management, even if it is just for single projects, therefore, will eventually evolve into data base management which is the traditional domain of the RDMS vendors.

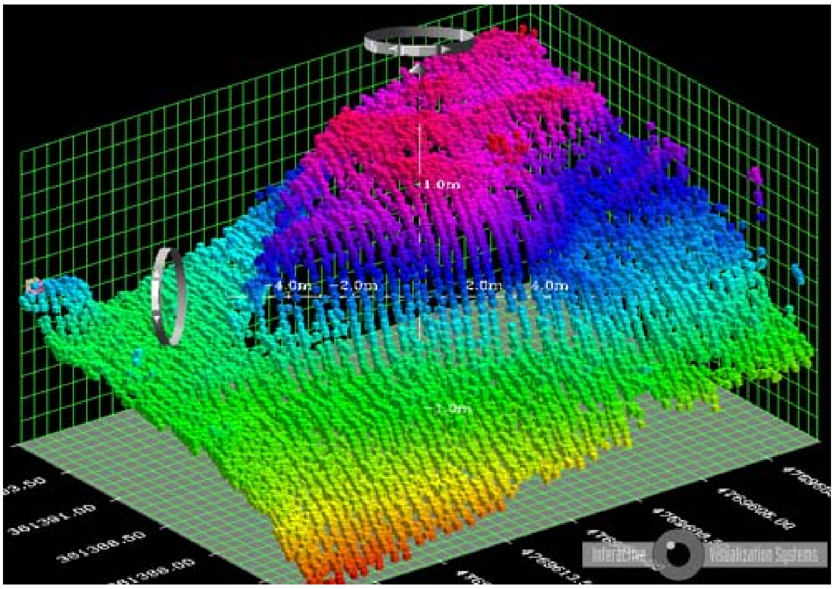

Figure 1. Advanced 3D editing using Fledermaus.

The huge volumes of (sensor) data do however create challenges for them too and the geo-spatial community must encourage data base vendors to continuously improve their systems.

2.4 Large digital terrain models

What is expected to become a problem for project based work is already pressing when it comes to building large high resolution digital terrain models (DTM) built from different raw point cloud sources. Creation of the DTM involves cleaning and processing of point cloud data but not automatically data thinning. Managing these models is therefore just as complex and challenging as managing the point clouds itself. As mentioned before the introduction of spatial infrastructures forces data owners to provide their collection of single data sets in a seamless and quality controlled way to the end-user. As long as these models have a relatively low resolution (e.g. 30 arc-second spacing) and up-to-dateness is not so much an issue, little problems exist. Examples of such data sets are the ETOPO1/GEBCO DTM (see Figure 2) and the Global Topography data set from Smith and Sandwell [3].

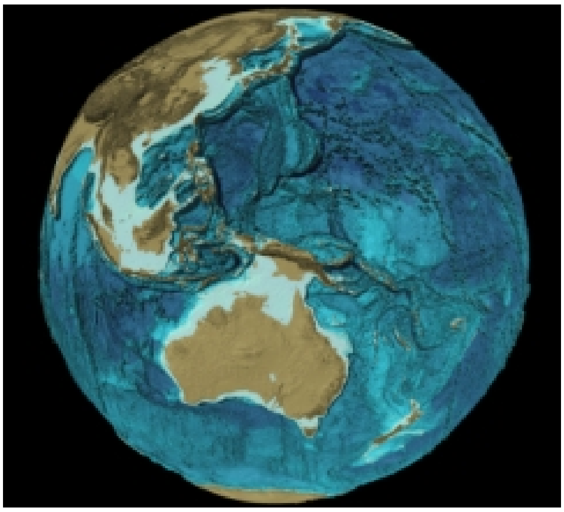

Figure 2. GEBCO(08) DTM.

Spatial data infrastructures however require more and better data. This is a key issue for SDIs and the reason they are being initiated and mandated (e.g. the European Inspire Directive [4]). Without better data decision makers can't make better decisions and they can't make them faster. Data sets have to be provided for a wide user community and this requires a careful planning of the management of the data. Provision of data in a spatial data infrastructure is an ongoing responsibility. It is not a one time project but end-users will expect almost direct access to all new data that comes available. This will be the challenge for many bathymetric data producers in the near future.

2.5 Morphological research

The challenge can probably be best explained by providing an example. Marine morphologists that study sediment transport in estuaries require many data sets representing the seabed topography at various moments in time. Various statistical sampling methods may be used to build the DTMs for this purpose but the main characteristic of these data sets is the fact that the data set represents a snapshot in time. If this would not be the case the risk exists that the same (moving) sediment is found multiple times in a single DTM.
The time factor is, therefore, the key issue and is dominant over the quality of the survey. In other words IHO S-44 – the survey (quality) standard of the International Hydrographic Organization [5] – quality is something a marine morphologist is less worried about if there are no alternative (better) data sets available. This implies that the models used by marine morphologists are different than the ones a marine cartographer is using for the production of a navigational chart. The expert in the latter example would not use lower quality data which is more recent if there is a slightly older higher quality data set available for the same area that provides better coverage.

To satisfy the needs of all users in a spatial data infrastructure flexibility will be the key issue. Given the fact that data volumes are increasing rapidly it will become a serious problem if an organisation collects various (end-user) requirements and tries to derive at agreed specifications for all these users.

2.6 Liability

Especially for National Hydrographic Offices responsible for producing navigational charts of their home water, liability is an important issue. Governments are responsible for the information they provide and can not easily wave that responsibility as most private (data) organisations do. In case of accidents caused by faults or omissions in navigational charts, the hydrographic office must provide evidence that the fault or omission can not be blamed on them. The only way to do that is to keep details records of their work and to operate a system that allows them to reproduce their knowledge of the situation before the accident occurred. Apart from archiving individual source documents and surveys, they also have to archive every change and process step associated with that change in the DTMs used for the production of the navigation charts.

Given the density of the datasets traditional backup procedures should be avoided as much as possible. So not just the various (end) user requirements call for efficient solutions but also the liability issue is heavily contributing to today's data management challenge.

3. Virtual Continuous Models

To overcome some of the problems associated with the management of large digital terrain models some solutions try to simplify the data as much as possible in order to reduce the size of the derived models and consequently increase the manageability. Although this may sound straightforward and workable, reality is different. Hydrographic data reduction by modelling the data sets into grids or TINs is lossy and the result is a compromise and does not always satisfy all users.

Concluding that storing many copies of high density models is not an option and building simplified models does not provide the solution either, calls for a solution that combines the advantages of both methods without incorporating the disadvantages. Virtual Continuous Models (VCM) as introduced by ATLIS combine the best of both worlds. A VCM defines the seabed topography based on rules defined by the user of the model. The rules are translated into standard SQL queries and describe how the individual surveys from the archive should de-conflicted and combined into a DTM without actually copying all the required data points into the DTM. An example of a simple rule could be: give the highest priority to the most recent survey and the lowest to the oldest. In order to increase the processing efficiency, each defined VCM is pre-processed by storing the outer limit or survey (concave) hull of those parts of the individual surveys that participate in the DTM. A concave hull follows the distribution of the point better than a convex hull. A concave hull can have so called dents bending inwards thus avoiding unnecessary “empty” spots inside the hull (see Figure 3 [6]). The hulls are stored unprojected as latitudes and longitudes to support potential world coverage of the VCM.

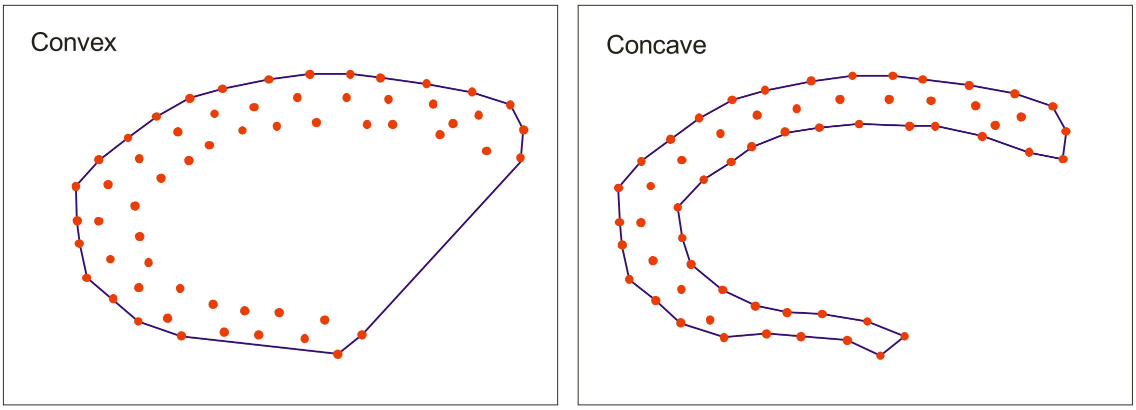

Figure 3. Convex versus concave hull.

The VCM definition consists of the rule used for de-conflicting and combining the individual modelled surveys and the set of concave hulls that together form the outer boundary of the DTM. This set can best be imagined as a source diagram that is often shown on paper navigation charts (see Figure 4).

The set of concave hulls for a given VCM can become very complex especially if many surveys have been conducted in the same area. The rule definition determines the stacking order of the individual surveys and also which data points from the source data participate in the result (e.g. shallowest, mean, CUBE hypotheses etc).

The number of VCM that can be defined using this mechanism is almost unlimited and the advantage is clear: the actual data points are only stored once.

3.1 Survey archive

The technology of VCM is based on the assumption that hydrographic organisations always archive the original source data. The idea is that if something goes wrong or turns out to be wrong in a later stage of the work process one can always revert back to the original data. Also the original data often contains much more than just depth information. Backscatter information can be geo-coded and used to determine bottom classification. This assumption is important as it does not in any way lead to more inefficiency in the solution as it requires no additional storage space. The source data is stored anyway but until now the archive was often difficult to access. Using VCM will change this. The only requirement for archiving the source data in order to make the data usable in the VCM concept is a common geodetic reference system. As a

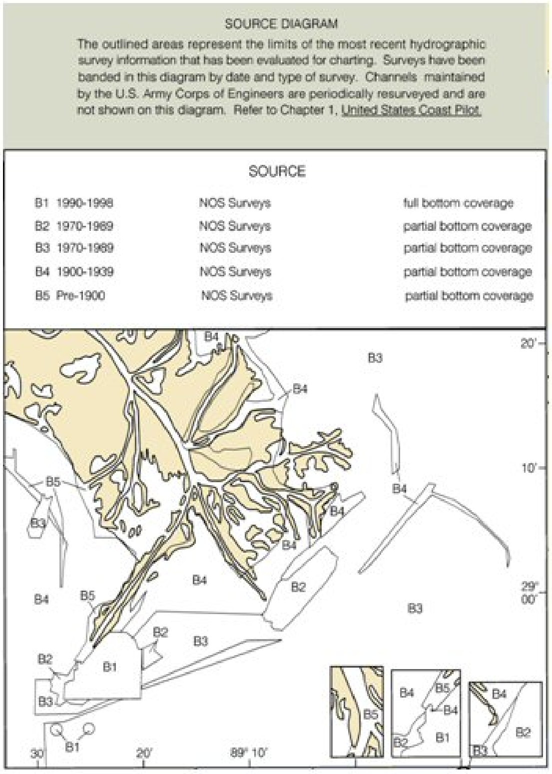

Figure 4. Example source diagram (courtesy NOAA).

DTM can cover the whole world, the system is based on a world coordinate system (lat/lon) and a single horizontal datum. Depths are negative and expressed in standard unit. Conversion of different coordinate systems and datums is performed as part of the archiving process. In theory this could be done on the fly but for performance reasons it is efficient to do this up-front. Vertical datum adjustment is also part of the modelling process.

The source data archive is catalogued by adding metadata to the individual surveys that allow easy access to the data based on international metadata standards (ISO 19115 and the OGC CSW standard).

3.2 Modeling process

Depending on the requirements of the organisation (and its customers) using VCM technology, the individual surveys can be modelled. Modelling may include to following steps:

– Adjustment (transformation) to a common geodetic model.
– Adjustment of vertical datum (based on separation models if available or fixed offsets).
– Data validation (e.g. using CUBE).
– Data thinning.
– Gap management (applying attributes to gaps in the data set describing the reason of the gap and the scale threshold for interpolation in products).
– Adding node level meta data (e.g. statistically derived quality information).

The result of the modelling process is called an Individual Model (IM) within the VCM concept. The IM acts as the building block of the VCM and has the same coverage as the original individual surveys. The IM is stored as a point cloud but is easier to manage as a result of the modelling process. This leads to a better overall performance of the VCM technology.

If required, the original (raw) data can be archived together with the modelled version of the individual survey.

3.3 Combining individual models

Different users require different rules for VCMs. This will result in different DTMs after applying the rules to the archive. Derived DTMs may even contain interpreted objects (e.g. interpolated depths) if the interpretation algorithm is added to the VCM definition. Figure 5 illustrates a survey archive consisting of 5 overlapping (conflicting) individual models.

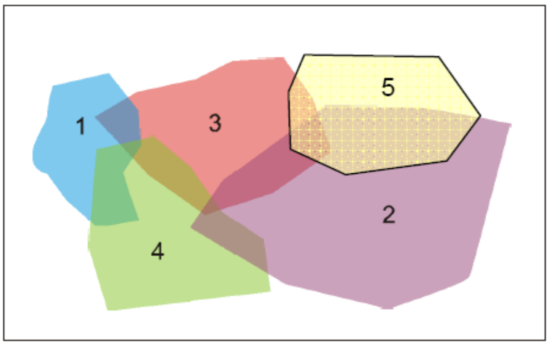

Figure 5. Archive of individual overlapping models.

After applying a rule set to the archive a VCM as shown in Figure 6 can be built.

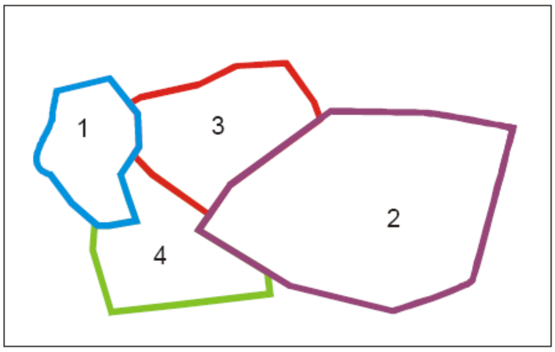

Figure 6. VCM.

The rule set used for the VCM in Figure 6 excludes IM 5 and give IM 4 the lowest priority. IM 1 and 2 have an identical priority and IM 3 has a higher priority than IM 4. Using a different rule set on the same archive could result in the VCM shown in Figure 7.

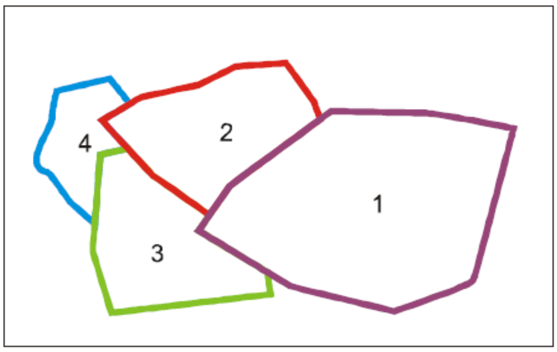

Figure 7. Alternative VCM.

The rules applied to the VCM definition can be very simple (e.g. most recent survey on top) or could be based on complex queries taking many meta data aspects into account. Although intersecting survey hulls is relatively straightforward in most GIS systems, care must be taken of the geodetic system used for the intersection. When using a non-projected coordinate system based on latitudes and longitudes, intersections must be based on geodetic lines if high accuracies are expected.

3.4 Performance

In order to make the VCM concept workable it is important to store the archive in a very efficient way. Fast data access to the data points is extremely important when the VCM is used in 2D or 3D visualisations or when output products (e.g. gridded bathymetry sets, contour lines or spot soundings) need to be generated.

ATLIS has implemented the VCM technology in SENS Bathymetry. The storage solution of SENS is based on Oracle spatial and utilises the latest in spatial technology that the Oracle 11g platform is offering today. Using the highly efficient spatial indexing technique combined with Oracle partitioning ensures a good performance of the VCM concept.

Performance is further enhanced by using the helical hyperspace (HHCode) concept for point data as originally designed for bathymetric applications by Mr. Herman Varma of the Canadian Hydrographic Service. The HHCode is a multidimensional construct based on the Riemannian Hypercube Structure that has an inherent z-order pattern also known as the Peano Curve [7]. The HHCode expresses dimensions in interleaved binary integer form to achieve high levels of compaction as well as rapid indexing and retrieval. Alternately stated, the HHCode is a recursive decomposition of an n-dimensional space. The entire planet Earth can be used as an example to illustrate the concept of the HHCode construction in 2D for a given spatial position (see Figure 8).

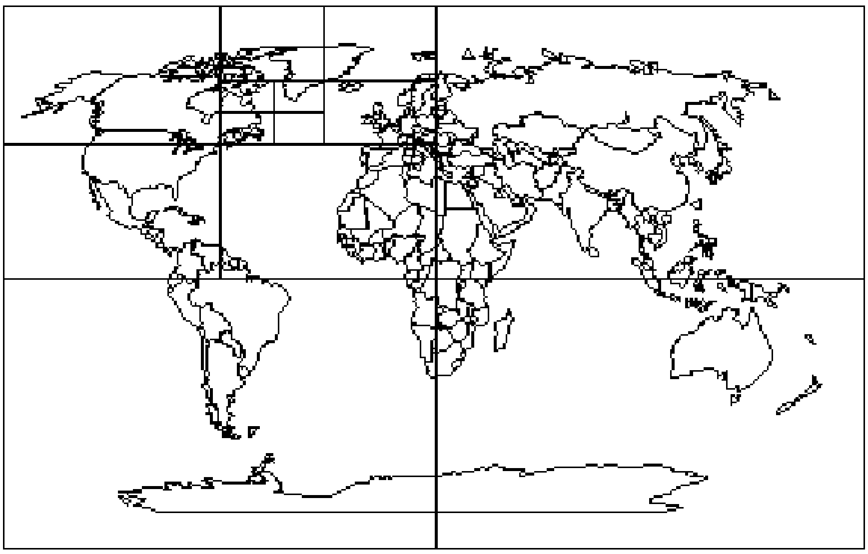

Figure 8. Recursive decomposition of space.

4. Future requirements

Even though the Oracle 11g platform provides enough power today to support the VCM concept, it is expected that the increase in data volumes of modern sensors system will continue to stress the limits of the storage capabilities. This requires on going research in further enhancing data access speed, indexing and partitioning technology. Especially the latest point cloud data structures introduced by Oracle are promising in that aspect.

5. Conclusions

Today's and future multi beam echo sounders and LiDAR systems generate a staggering amount of data. Although processing of huge individual survey files is still possible using advanced software packages like Fledermaus, managing the volumes of data in a seamless and unambiguous quality controlled way is becoming more and more a problem. Performance of today's system is no longer adequate given the data volumes and the increasing number of users that access data simultaneous. End users have different requirements and demand better data quicker. They are no longer satisfied with a fixed dataset built for general use. This requires the management and maintenance of multiple high resolution digital terrain models (DTMs). Given the huge volumes involved this will eventually lead to problems.

The Virtual Continuous Model technology, introduced by ATLIS provides a workable solution for this problem by no longer storing the data points redundant for each different DTM. The VCM only stores the definition of the required DTM by specifying the rules for combining and de-conflicting the individual surveys that are stored as individual models in the survey archive. Using the VCM concept gives a hydrographic organisation full control and maximum flexibility when it comes to managing bathymetric data. Participating in national and international spatial data infrastructures is no longer a problem but becomes an interesting show case for organisations using the VCM technology.

References

1. http://ec.europa.eu/maritimeaffairs/emodnet_en.html (May 2010).
2. Calder, B.R, "Automated Statistical Processing of Multibeam Echosounder Data", Submitted to Int. Hydro Review, January 2003.
3. Smith, W. H.F., and D.T. Sandwell, Global seafloor topography from satellite altimetry and ship depth soundings, Science, v. 277, p. 1957-1962, 26 Sept., 1997.
4. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).
5. IHO Standards for Hydrographic Surveys (5th edition), www.iho.org. (See also L.L. Dorst and C. Howlett, Evolutions in the IHO standards for hydrographic surveys, S44, Proceedings of the International Hydrographic Conference 'Evolutions in Hydro-graphy', Hydro06, Antwerp, Belgium, 6-9 November, 2006).
6. http://ubicomp.algoritmi.uminho.pt/local/concavehull.html (May 2010).
7. Varma et al., "A Data Structue for Spatio-Temporal Databases", International Hydrographic Review, Monaco, LXVII(1), Jan. 1990.