# An Evaluation of CARIS Bathy DataBASE as a Bathymetric Data Management Solution for CHS Quebec

Authors: Jean-Guy Nistad[1], Frédéric Lavoie[2], Karen Cove[3]

## Abstract

A bathymetric data management solution has been identified by the Quebec region of the Canadian Hydrographic Service (CHS) as the missing element in the transition from a file-based to a database-driven operational workflow. Responding to this need, a pilot project was initiated by the CHS Quebec Region and the Interdisciplinary Center for the Development of Ocean Mapping (CIDCO) to evaluate the solution provided by CARIS Bathy DataBASE. The pilot project's aims were to determine if Bathy DataBASE could provide efficient loading, management, extraction and distribution operations of CHS Quebec Region bathymetric data. The criteria used for the evaluation were performance, data integrity and robustness.

This paper provides a description of the evaluation and identifies the considerations of integrating Bathy DataBASE into the CHS Quebec Region operational workflow. A better understanding of Bathy DataBASE's current capabilities and limits provides for a smoother system integration and better overall utilization.

## Introduction

Ever increasing size, coverage and spatial resolution of Multibeam Echosounder (MBES) surveys require ever more human, software and hardware resources dedicated to their storage and management. Likewise, increased demand for hydrographic products requires simpler and more efficient distribution mechanisms. At the present time, the operational workflow used at CHS Quebec Region for bathymetric data management cannot provide for these future requirements without affecting product quality and delivery times. The present architecture on which this workflow is based consists of a data bank comprising a locally-owned repository of files and a nationwide database of metadata (CHSDir). All data operations and manipulations such as data compilation, data validation, non-hydrographic product creation (bathymetry requests from clients) are file-based. Replacement of the latter architecture with one based on a database-driven management system is expected to provide the following benefits:

- Integration with existing hydrographic production databases.
- Data traceability from source to final product.
- Minimal loss of information from reduced manipulation.
- Ease of data access from both internal and external parties.

[1] Interdisciplinary Centre for the Development of Ocean Mapping (CIDCO)
[2] Canadian Hydrographic Service (CHS) Quebec Region
[3] CARIS

CHS Quebec Region has been evaluating CARIS Bathy DataBASE as a database-driven replacement to the current file-based architecture. The former anticipates that Bathy DataBASE will play a critical role in managing various bathymetric data and products: surface source bathymetry, new detailed bathymetric products for navigation (potentially distributed in S-100 format) and new bathymetric products intended to respond to needs other than navigation.

## The Bathy DataBASE Solution

CARIS Bathy DataBASE is a client-server application designed as a framework for the manipulation of bathymetric data. In the CARIS 'Ping to Chart' workflow, it sits between the data processing and validation system (HIPS/SIPS) and the hydrographic data management and production system (HPD). Bathy DataBASE's architecture consists of the BASE Manager client application, the Bathy DataBASE server and the suite of administration tools. The architecture provides a three-step operational workflow consisting of 1) analysis, edition and validation; 2) loading, management and extraction; 3) product creation.

The analysis, edition and validation step helps users prepare the data for integration and control the quality of survey data from both primary and less stringent third-party data collection sources.

The loading, management and extraction step provides the user interaction mechanisms with the database content. Bathymetric data is stored as either high-density gridded datasets or lower density vector point datasets. Gridded datasets are stored at surface density, i.e. at a resolution that reflects the spatial accuracy of the acquisition sensor, the surveyed water depth and the location type of the survey (e.g. open-water, maintained dredged channel, harbour, delta). Within the database structure, datasets are grouped within a geographic entity called a survey. The survey is defined by a common spatial extent and a table that describes S-57 type attributes, survey vessel configurations, and surface properties (Figure 1). During the creation of the survey or the loading of bathymetry datasets, attribution/metadata is defined to facilitate the effective management of the data.

Surveys and datasets stored in the database are accessed through the spatial extent defined within a display window, not on a file by file basis. The display window allows the user to select specific datasets or a user-defined area for extraction. When the user-defined area selects multiple, overlapping datasets, Bathy DataBASE provides a de-conflict resolution mechanism based on the user-selected supersede rules presented in Table 1. This mechanism allows for the overlapping datasets to be combined into a seamless representation. Rules 1 and 2 provide a node level attribution mechanism, whereby individual attribute layer nodes are de-conflicted based on their unique value. Rules 3 to 10 provide a dataset level attribution mechanism, whereby entire datasets are de-conflicted based on their specific attribute value.

The product creation step helps create hydrographic products (contours, depth areas and soundings) as properly attributed S-57 objects. The products can be made available to a production system such as HPD for analysis and possible inclusion or updating of existing chart products (e.g. ENC, AML). Product creation can be enhanced with the use of tools for surface generalization and surface de-confliction.

**Attributes**

| Acronym | Name | Value |
|---|---|---|
| SURSTA | Survey date - start | |
| SUREND | Survey date - end | |
| OBJNAM | Object name | BS-2004-583_Ile_Rouge-3m_Interp |
| TECSOU | Technique of sounding measurement | found by multi-beam |
| acqsys | Aquisition system used in survey | EM1000 |
| possys | Positioning system used in survey | DGPS |
| vessel | Name of acquisition vessel | FCG Creed |
| STATUS | Status | permanent |
| modtim | Last modified time | 20070806 024939.000 |
| cretim | Creation time | 20070806 024837.000 |
| srfcat | Category of bathymetric surface | standard |
| srfres | Gridded surface resolution | 3.00 m |
| srftyp | Storage Type | Gridded |
| idprnt | Parent id | 16777279 |

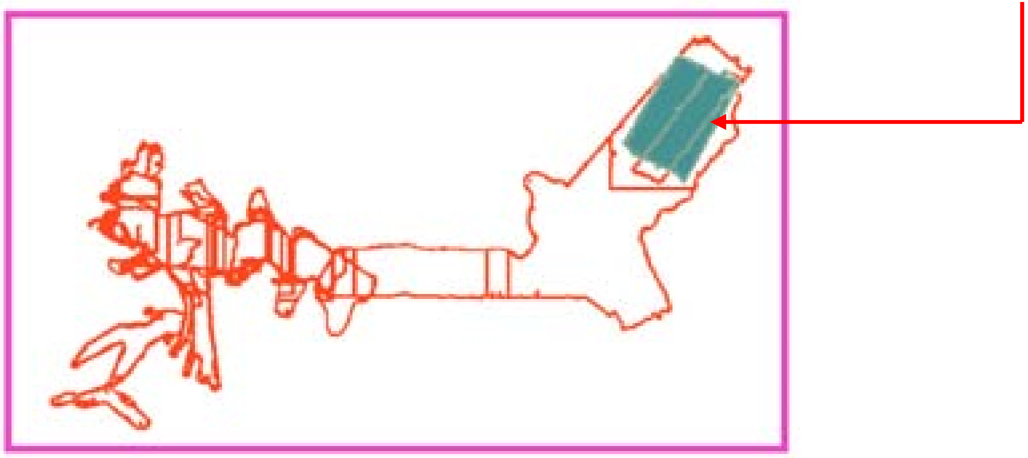

**Figure 1 – Dataset contained within a survey and its associated attribute table**

**Table 1: Supersede rules used for conflict resolution**

| Rule # | Supersede rule name | Supersede rule description |
|---|---|---|
| 1 | *Where the [attribute] is greatest* | $L_n(S1) > L_n(S2)$ |
| 2 | *Where the [attribute] is least* | $L_n(S1) < L_n(S2)$ |
| 3 | *Where the attribute [acronym] is greatest* | $A_n(S1) > A_n(S2)$ |
| 4 | *Where the attribute [acronym] is least* | $A_n(S1) < A_n(S2)$ |
| 5 | *Where the attribute [acronym] is equal to [value]* | $A_n(S1) = X$ |
| 6 | *Where the attribute [acronym] is not equal to [value]* | $A_n(S1) \neq X$ |
| 7 | *Where the attribute [acronym] is less than [value]* | $A_n(S1) < X$ |
| 8 | *Where the attribute [acronym] is greater than [value]* | $A_n(S1) > X$ |
| 9 | *Where the attribute [acronym] is [in a specific range]* | $X_2 > A_n(S1) > X_1$ |
| 10 | *Where the attribute [acronym] is [not in a specific range]* | $X_1 > A_n(S1) > X_2$ |

*Where:* $L_n(S1)$ *is the node value of the* $n^{th}$ *attribute layer of dataset* $S1$.

$A_n(S1)$ *is the value of the* $n^{th}$ *attribute of dataset* $S1$.

$X$ *is a value taken by a specific attribute.*

## Evaluating Bathy DataBASE: a pilot project

The CHS Quebec Region has been evaluating Bathy DataBASE as part of a larger initiative to integrate database-driven technologies. CARIS HPD is already at the core of CHS' operational workflow for hydrographic production and management. The addition of Bathy DataBASE will provide a seamless workflow for all relevant hydrographic data.

Preliminary evaluations were very convincing, yet some results were not as expected. For example, conflict resolution occurring at the node level instead of the expected dataset level (Cove & Lavoie, 2007).

A second phase of the pilot project occurred in the summer 2007. This phase of the pilot project was to determine if Bathy DataBASE could provide efficient loading, management, extraction and distribution. The criteria used for the evaluation were performance, data integrity and robustness. These tests were performed at CHS Quebec Region offices on the following versions:

- Bathy DataBASE Server 2.0 to 2.1 (hotfix 1)
- Bathy DataBASE 2.0 (hotfix 12) to 2.1 (hotfix 1)

Subsequent tests were performed at CIDCO offices on the latest versions at the time of publication:

- Bathy DataBASE Server 2.1 (hotfix 4)
- Bathy DataBASE 2.1 Service Pack 1 (hotfix 4)

## a. Loading

Source of CHS Quebec Region bathymetry is centered on Saint-Lawrence River and Gulf and includes new and historical data from leadline, single-beam, MBES, etc. from both primary surveys and external providers. Low–density data was traditionally stored as vector point layer in the CARIS Interchange (NTX) format. The data was converted to Sounding Set (BPS) file format before being loaded in Bathy DataBASE Server. High-density data was initially stored at surface density as gridded BASE Surface (HNS) file format created by HIPS. These could directly be imported into Bathy DataBASE Server.

All new BASE Surfaces were loaded into the server. Initially, problems with the surface contour tracing algorithm on Bathy DataBASE Server version 2.0 prevented some surfaces from loading properly. Moreover, drawing of the contours at large scale was computationally intensive. Modifications to the algorithm as of version 2.1 solved both of these issues.

11% of new BASE Surfaces were not loaded because of referencing errors between the HNS file and the CHSDir metadata database. These errors, most likely manipulation errors, highlight the importance of improving the current operational workflow used at CHS Quebec Region. Table 2 provides the main statistics of BASE Surface loading. The high difference between mean and median highlights the fact that the majority of surfaces were of relatively small size: less than 50 Mbytes. Not surprisingly, statistical correlation between file size and load time is very high, reflected by Pearson’s r correlation of 0.91. Figure 2 illustrates a quantile analysis where file sizes are grouped in 10 classes (deciles) and show mean load time per decile. Y-error bars represent 95% confidence interval mean load time variability. A greater variability in file size for the last five deciles is clearly highlighted by larger confidence intervals.

**Table 2: Load statistics for CHS BASE Surfaces**

| Min/max file size (MB) | Min/max load time (sec.) | Mean load time (sec.) | Median load time (sec.) | Standard deviation (sec.) |
|---|---|---|---|---|
| 0,017/335 | < 1/356 | 17 | 6 | 30 |

Low-density data stored as NTX files were first imported into the BASE Manager client as Sounding Sets and subsequently loaded into the server. At CHS Quebec Region, each NTX file sounding possesses a validity attribute. The sounding is said to be 'active' if it is accepted as legitimate or 'suppressed' if it is superseded by another dataset. This attribute is retained through the load process.

Datasets stored as Sounding Sets were not extensively tested in this evaluation project. Load times are not so much of an issue since Sounding Sets are much lower density formats than BASE Surfaces.

A few recommendations were put forth from the observations made during the evaluation project. Among these were the necessity to further investigate the behavior of Bathy DataBASE when loading very large BASE Surfaces (100 Mbytes – 2 Gbytes) and to optimize Bathy DataBASE to take advantage of hardware capacities such as Duo-Core processors and hyper-threading support.

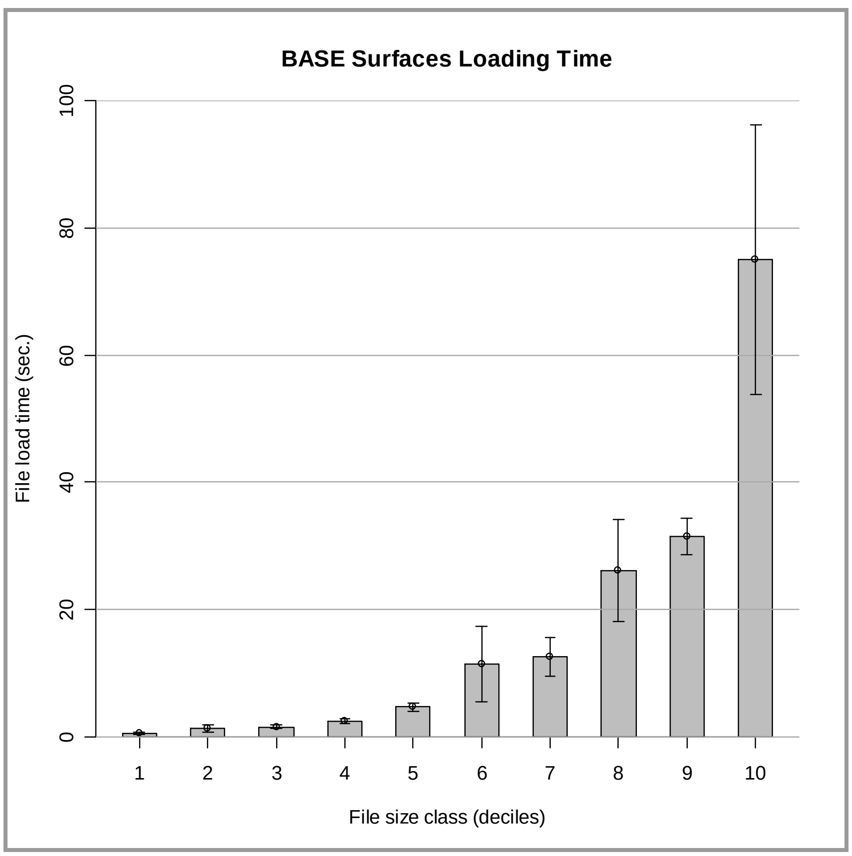


**Figure 2 – Quantile analysis of BASE Surface load time**

### b. Management

Bathy DataBASE achieves successful management by proper data integrity, of which there are three types: entity, attribute domain and referential.

Entity integrity ensures that each object in the database is represented by a unique identifier. The physical data storage structure ensures that this rule is always valid as each object is represented by a unique identifier that indicates its category and its chronological order. In the context of CHS operational workflow, care will have to be taken to prevent surface duplication. Indeed, the filename which presently uniquely identifies each bathymetry file is not part of the identifier of the dataset in Bathy DataBASE. Data duplication during the loading process is thus possible if based solely on the filename.

Attribute domain integrity is ensured by a pair of user-definable and extendable XML files. Their purpose is to describe each possible integer value that an attribute can take and to define the permitted value for the attribute. They are known respectively as the *Pool* and *Profile* files. Again, in the context of CHS' future operational workflow, care will have to be taken when entering dataset attributes. Although each attribute's integrity is ensured, inter-attribute integrity is not. It is thus possible to enter, within a single dataset, attributes that are not logically related. For example, a surface could have its *acquisition system* attribute set to 'Simrad EM3002 multibeam' while its *survey type* attribute is set to 'single-beam echo sounder'.

In a database, referential integrity ensures that an existing relationship rule is applied. Referential integrity makes it possible, among other things, to manage the objects in a database efficiently. Figure 3 shows Bathy DataBASE's entity-relationship diagram. Datasets are grouped logically within a unique survey, which itself can contain 0 to N datasets. A Product Dataset inherits its properties from its parent dataset and forms a unique pair with its Product Profile, the equivalent of a survey for Product Datasets. The only difference is that a Product Profile can have one and only one Product Dataset. The Product Profile encapsulates the result of the conflict resolution based on the selected hierarchy of supersede rules. Each supersede rule is associated to a unique pair of datasets.

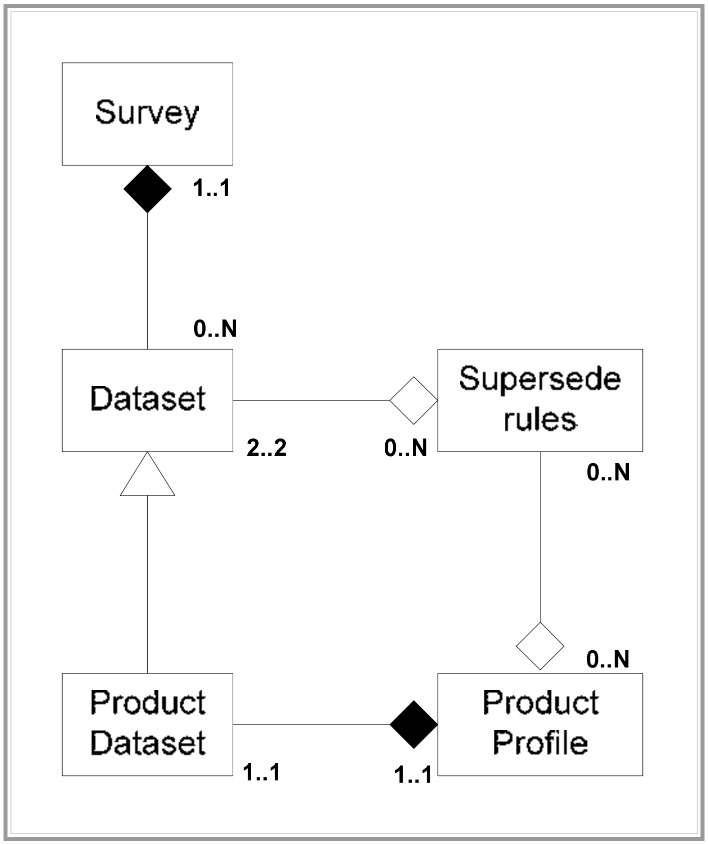


**Figure 3 – Entity-relationship diagram for Bathy DataBASE**

A few recommendations were put forth concerning the management of bathymetric data. Among these was the fact that the objects (datasets, surveys, Product Datasets and Product Profiles) could not be physically removed from the database. The delete function present at the time of the pilot project only removed the object from the client side display window, but it remained on the physical storage device. In the latest Bathy DataBASE release, the delete function marks the object as deleted on the server side. Objects flagged as deleted are purged and removed from the physical storage device at a scheduled interval. An exception to this procedure is if the object in question has contributed to the creation of a Product Dataset. To ensure integrity and traceability, the object will not be deleted as long as the Product Dataset has not been deleted beforehand.

Another important consideration is that the content of the database that is viewable depends on the extent of the display window at the instant the link was established between the client and the database. If one tries to pan, no new data corresponding to the new extent will be loaded. This issue has been addressed and a solution will be part of future releases.

## c. Extraction

Extraction tests involved the creation of Product Datasets using different dataset combinations. In general, three types of combinations were possible, as illustrated in Figure 4:

- Two BASE Surfaces creating a new BASE Surface
- A BASE Surface and a Sounding Set creating a new BASE Surface
- Two Sounding Sets creating a new Sounding Set

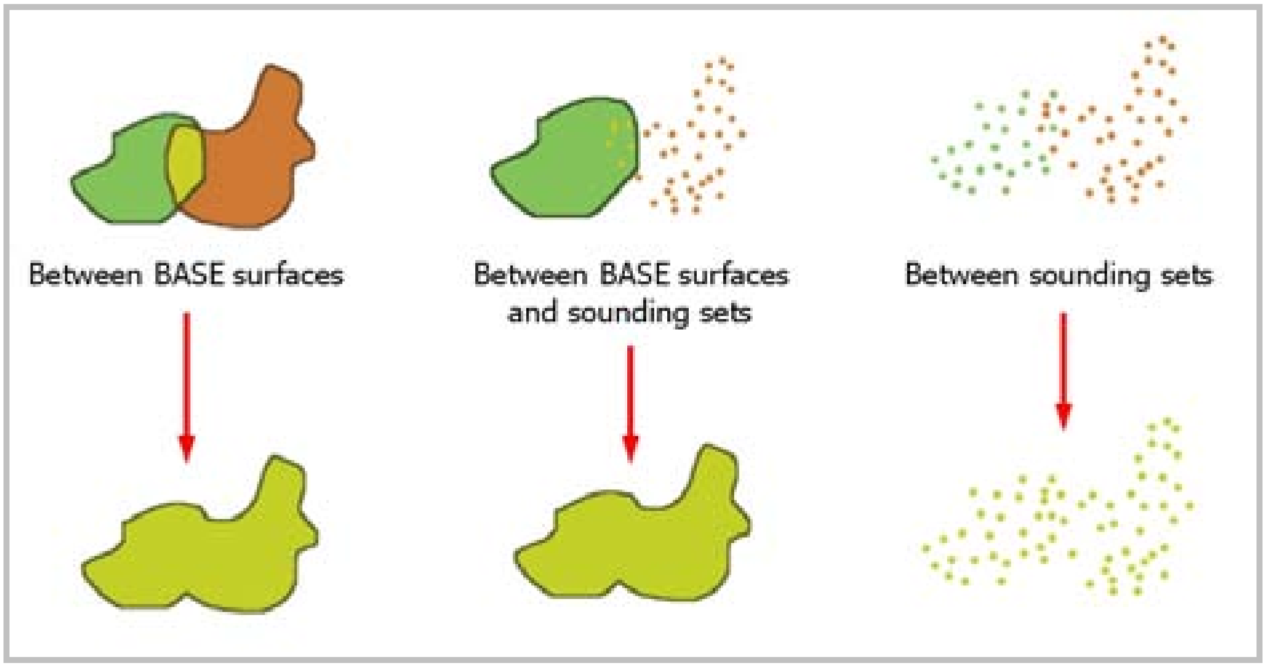


**Figure 4 – Possible types of dataset combinations**

The Product Dataset is defined by an appropriate choice of resolution, spatial extent, projection and hierarchy of supersede rules. The resolution applies only for BASE Surfaces and can be set to any value. However, a too high resolution will generate holes in the resulting BASE Surface as no oversampling is performed. The spatial extent defines the area of the Product Dataset. It may be a subset of a dataset or encompass one or many datasets, contiguous or not. The selected projection must be chosen appropriately as drastic changes in projections may create holes in the resulting BASE Surface. Interpolation may be required to fill them. Initial tests also revealed a dependency of results on the destination projection. The Modified Transverse Mercator (MTM) projection used at CHS Quebec Region was one of the problematic projections. User-selected supersede rules would give erroneous results. This problem has been fixed on the latest release. A hierarchy of rules allows the generation of a Product Dataset with a low possibility of remaining ambiguities. Still, if none of the conditions of the supersede rules are met, a default Product Dataset is currently generated. Future releases of Bathy DataBASE will be more stringent with this issue by warning the user if no valid result can be achieved with the selected set of supersede rules.

When de-conflicting two BASE Surfaces to generate a new lower resolution BASE Surface, we were expecting the conflict resolution to be applied only on the overlapping area for node level attribution (rules 1, 2 of Table 1). However, tests revealed that the rules were being applied on the non-overlapping area as well. In that case, the rules were used for decimation from a high resolution source surface to a lower resolution product surface. For dataset level attribution (rules 3-10 of Table 1), the decimation process simply selects the first node it encounters as the rules do not apply at the node level. The current release of Bathy DataBASE however now allows the user to define a rule to de-conflict the nodes within each dataset. This provides the user with the control to determine which values from the dataset in areas of non-overlap will be carried forward when going from a high resolution to low resolution dataset.

When conflict resolution occurs between a BASE Surface and a Sounding Set, the Product Dataset is always a BASE Surface. If the BASE Surface has precedence over that Sounding Set at the dataset level, the Product Dataset will show holes if gridded at a resolution higher than the average distance between soundings in a non-overlapping area of the Sounding Set (Figure 5). On the other hand, if the Sounding Set has precedence over the BASE Surface at the dataset level, the nodes of the Product Dataset in the overlapping zone with no associated sounding will the filled with nodes from the BASE Surface (Figure 6). This reveals that conflict resolution based on dataset level attribution still occurs at the node level. Results have also shown that the Sounding Set validity status is not considered in the conflict resolution. All soundings are considered, regardless if their status is 'active' or 'suppressed'.

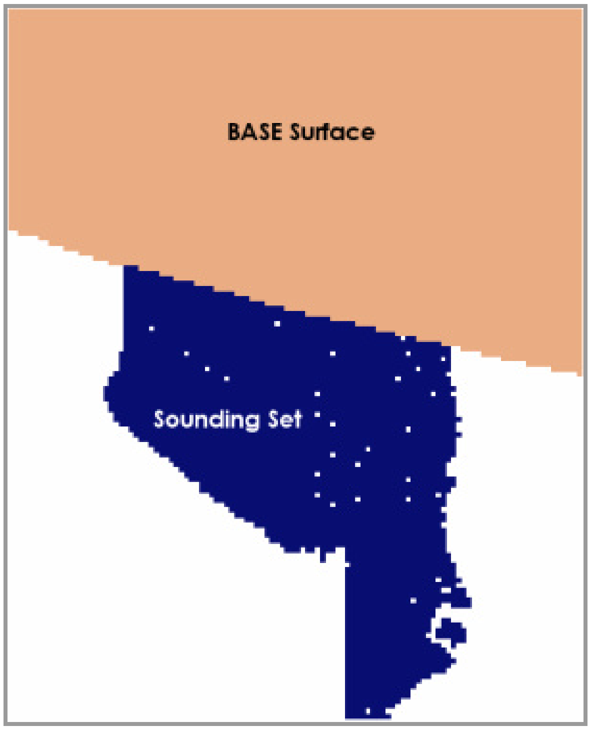


**Figure 5 – Contribution layer of a Product Dataset generated from a BASE Surface and a Sounding Set when the dataset level attribution selects the BASE Surface. Spatial extent is 925m x 1140m gridded at 8m.**


**Figure 6 - Contribution layer of a Product Dataset generated from a BASE Surface and a Sounding Set when the dataset level attribution selects the Sounding Set. Spatial extent is 925m x 1140m gridded at 8m.**

Conflict resolution between Sounding Sets is trivial. It occurs at the sounding level for soundings that have the exact same location. Identical locations depend on the precision and scale used to store the projected coordinates. More often than not, conflict resolution between two Sounding Sets will create a new Sounding Set which is the union of the source datasets, regardless of the supersede rules used. Future releases will provide an optional *zone of influence* for each node, allowing for a more realistic de-confliction to occur.

### d. Distribution

Data distribution was not evaluated in this pilot project. The CARIS solution for the distribution of spatial data on the web is Spatial Fusion Enterprise (SFE) and CARIS is currently undertaking the initiative to provide internet access to bathymetry data from Bathy DataBASE using SFE and standard web browsers. Through this web service users will have the ability to display, select, and query BASE Surfaces, Sounding Sets and their associated metadata. Data download will also be addressed through future implementations.The other intention of SFE is to maintain interoperability with external data providers by providing Bathy DataBASE content using Open Geospatial Consortium (OGC) compliant services (Figure 7). This will include the use of Web Map Service (WMS) and Web Feature Service (WFS). A separate pilot project on its integration at CHS Quebec Region is being considered.

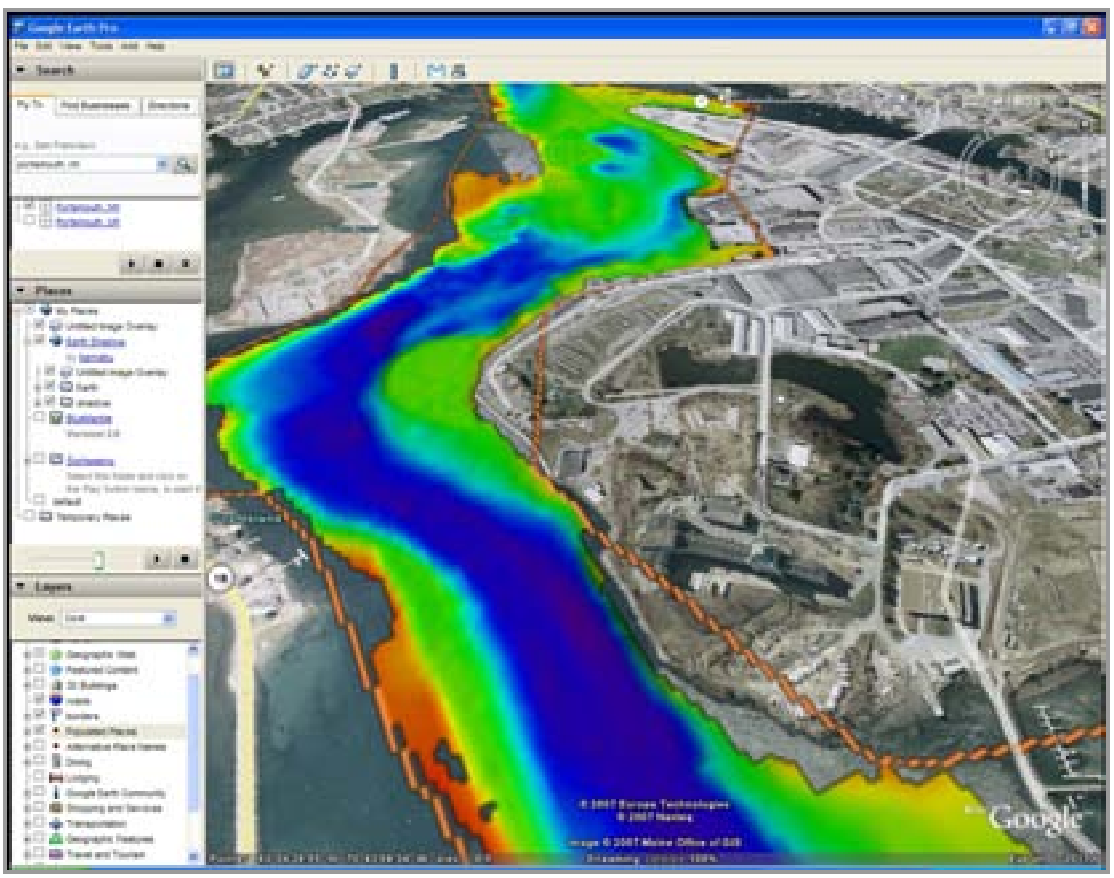


**Figure 7 – Prototype of interoperability between bathymetric data and Google Earth**

## CHS Integration Considerations

Successful integration between CHS operational workflow and the CARIS Bathy DataBASE architecture will require adaptability from both parties. Indeed, current practices must be adapted to the design philosophy of the software, and, likewise, the software must adapt to the necessities of current practices.

Results of the pilot project provided CHS Quebec Region the possibility to identify remaining considerations before full integration of Bathy DataBASE. Some considerations are described herein. The first concerns the standardization of storage formats. Should all datasets be gridded and stored as BASE Surfaces at surface density or should both BASE Surfaces and Sounding Sets be used? Chart bathymetry stored as Sounding Sets refer to measured depth, whereas BASE Surfaces attribute a calculated node value based on neighboring soundings. Acceptance of BASE Surfaces as source bathymetry data in the hydrographic community is still a subject of discussion. A second consideration concerns data duplication between CHSDir and Bathy DataBASE. Ideally an automated link could be developed between both databases. However, CHSDir is a nationwide system and cannot be modeled exclusively to the needs of CHS Quebec Region. A third consideration is data integrity and traceability along the ‘Ping to Chart’ workflow in order to minimize data manipulation and consequently errors being introduced. Providing a traceable single project number for each dataset from the moment it is generated in HIPS could be a step in that direction. Yet another important consideration is that of datum adjustment. When product creation involves source bathymetry referenced to various chart datum, a common vertical datum must be used. Two solutions are possible: reference all datasets to a single chart datum the moment they are loaded in the database or let the user select the output chart datum when a conflict occurs between datasets. Finally, training of CHS personnel is primordial and will determine if overall efficiency improves and if a new operational workflow is adopted.

Most likely, modifications to the current operational workflow will be similar to those implemented when HPD became part of CHS Quebec Region current practices. Working divisions will be geographically based, not activity based. Database interaction will be bi-directional as opposed to the current unidirectional practice.

## Bathy Database Future Developments

Evolution of Bathy DataBASE is constant and will continue in the direction of improved efficiency. Among expected improvements, the BASE Surface and Sounding Set formats will both be redesigned to offer enhanced support for the ever increasing volume of individual bathymetric datasets. These new formats will also offer support for the storage of coordinate information in geographic coordinates, which will eliminate the projection transformation issues. Common functions will be updated to operate on the new formats and visualization will be further optimized in both 2D and 3D environments. All these changes are involved in the progression of Bathy DataBASE from a DataBASE of individual datasets to an enhanced, master surface/seamless coverage solution planned for a subsequent release. Other new features will include the totally redesigned Plot Composer, version 5.0, for the intuitive creation and plotting of hard copy plots and posters. Plot Composer will be utilized by BDB, HIPS and other CARIS applications.

## Conclusion

The evaluation of CARIS Bathy DataBASE as a bathymetric data management solution involved three partners (CIDCO, CARIS and CHS) working collaboratively with software engineering best practices in mind. Loading, management and extraction tests revealed issues that were progressively dealt with to ensure improvement and adaptability. Remaining considerations have been identified and will be addressed before full integration into a more efficient operational workflow. CHS Quebec region customers should notice the benefits of this new approach with a simpler and faster access to bathymetry data while maintaining the same high standard expected for hydrographic products.

## References

Cove, K. & Lavoie, F., 2007. Addressing the Need for a Bathymetric Data Management System. U.S. Hydro. 14-17 May 2007.

## Authors Biographies

Jean-Guy Nistad,
Interdisciplinary Center for the Development of Ocean Mapping (CIDCO)
Rimouski, Quebec, Canada
e-mail : jean-guy.nistad@cidco.ca
Jean-Guy Nistad completed a graduate degree (DESS) in Geomatics from Université du Québec à Montréal in 2006. He joined CIDCO in 2007 as Geomatician.

Frédéric Lavoie, Canadian Hydrographic Service (CHS)
Mont-Joli, Quebec, Canada
e-mail: lavoief@dfo-mpo.gc.ca
Frédéric Lavoie graduated with a bachelor’s degree in Geomatics from Laval University in Quebec in 1998. He began his career at CHS in 1999 as a Hydrographer and he has been in the Technical Support and Development Division of CHS Quebec region for five years.

Karen Cove, CARIS
Fredericton, New Brunswick, Canada
e-mail: karen.cove@caris.com
Karen Cove graduated from the University of New Brunswick in 2005 with a Master of Science in Engineering from the Department of Geodesy and Geomatics. She joined Caris as a Product Co-ordinator in the Canadian office in 2005.