GIS FOR PROCESSING MULTIDIMENSIONAL MARINE DATA IN SAAS MODEL

KRZYSZTOF DRYPCZEWSKI, BARTOSZ WIŚNIEWSKI, MARCIN KULAWIAK, KRZYSZTOF BRUNIECKI

Gdansk University of Technology
Gabriela Narutowicza 11/12, 80-233 Gdansk, Poland
krzysztof.drypczewski@gmail.com

Geographic Information Systems (GIS) have always been a useful tool for visualization and processing of geospatial data. However, their capabilities of analysis non-standard information such as hydroacoustic soundings has thus far been very limited. This paper proposes a general-purpose GIS which uses techniques such as OLAP, WCS and WCPS for processing of multidimensional spatio-temporal data. The versatility of the GIS is exemplified by performing a set of analysis operations on a time series of soundings collected via a multibeam echosounder.

INTRODUCTION

In the era of information society, data processing and visualization are key aspects of everyday life. Software supporting cognition and perception of data by humans enables Visual Analytics in domains such as science [1], defense [9], commerce [10] or social aspects [11].

A universal solution to these problems is provided by Geographic Information Systems (GIS), as they not only provide spatial data processing but also allow for visualization of results. Those systems can either be general purpose solutions [12] or dedicated ones [13][18][19].

GIS supports processing of various types of spatial data, including raster and vector elements [14]. Typically these types of data have two or three dimensions, depending on the purpose of the system. In the case of advanced analyses, spatial information alone is not enough. Such systems operate on four dimensions in the spatio-temporal domain.

Using time as the fourth dimension provides a variety of new possibilities in Geographic Information Systems. In the spatio-temporal domain a GIS may not only perform processing and visualization of geographical data, but with using this additional dimension it is possible to track object history, analyze feature evolution and perform prediction analysis. Such predictions, in particular, are very popular in commerce and business applications.

This being said, implementing a four-dimensional GIS is very problematic. Aspects that need to be taken into account in the design phase include identifying different versions of spatio-temporal objects, formats of their database storage or methods of processing time. Many of these issues may be solved by using a structure known as an OLAP cube [17][15]. A proper implementation of a multidimensional GIS creates a multi-purpose system for geovisual analysis of spatio-temporal data. This paper presents a concept of such a system and exemplifies its application to processing of hydroacoustic data.

1. STORAGE AND ACCESS OF MULTIDIMENSIONAL DATA

Systems processing spatio-temporal data require additional resources for activities such as multi-dimensional related operations or more efficient data processing. Some of those additional requirements - especially those which concern efficient processing of multidimensional data - can be satisfied by structure called the OnLine Analytical Processing (OLAP) cube [15]. Generally OLAP is a multidimensional array in which every dimension can store data. This data can either be simple or hierarchical (i.e. stored with a parent-child relation between single objects). Typically an OLAP cube is stored in a dedicated database using a star schema [16].

Thanks to its construction, an OLAP cube provides a variety of operations which can support geovisual analysis. Aside from the possibility to show more or less details (Fig.1a) for every single dimension, an OLAP cube enables operations like slicing (Fig.1b), dicing (Fig.1c) and pivoting (Fig.1d).

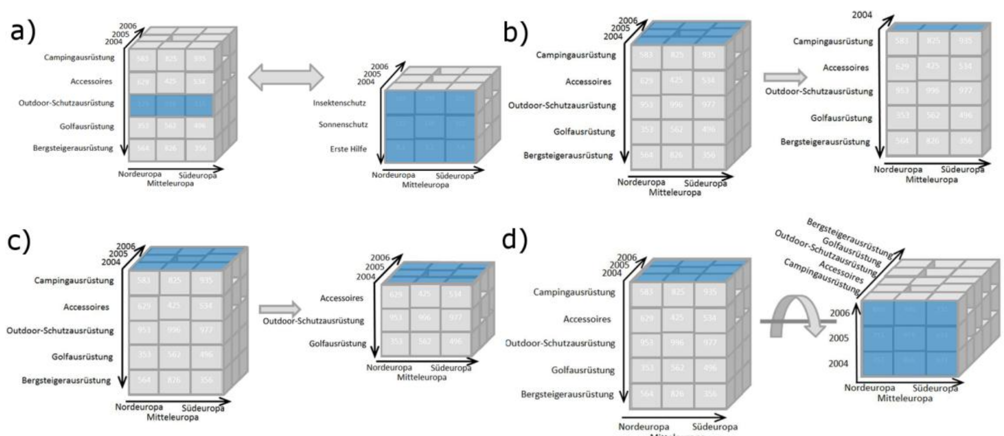

Fig.1. OLAP cube operations: a) roll-up/roll-down, b) slice, c) dice, d) pivot [16].

In the presented work, an OLAP cube was constructed from hydroacoustic data collected by a multibeam echosounder. Consecutive water column records form a multi-dimensional collection of geographic data which describes underwater environment. Each sounding is represented in the form of a two-dimensional array, in which each cell stores a numerical value. Those data sets are referred to as Multidimensional Discrete Data (MDD) and share common characteristics like potentially unlimited number of dimensions, spatial extent or cell semantics.

Once it has been stored in a geodatabase, such information needs to be remotely accessed and retrieved. Because the presented system works in a SaaS model, it should employ common standards of spatial data exchange. An open standard developed for the exchange of digital multidimensional geospatial data is known as Web Coverage Service (WCS). WCS, as the name suggests, operates on data in the form of coverages, as defined by ISO 19123.

The ISO 19123:2005 norm defines a conceptual schema for the spatial characteristics of geographic coverages. Coverages support the mapping of spatial, temporal or spatio-temporal domain to feature attributes, where attribute types are common to all geographical locations in the domain. A coverage consists of a set of direct locations expressed in a coordinate system which can be defined for up to three spatial dimensions and a temporal dimension. Examples include rasters, Triangulated Irregular Networks as well as collections of points and vector polygons. As data structures, coverages are dominant in many application areas such as remote sensing, meteorology and mapping of bathymetry, altitude, soil and vegetation. ISO 19123:2005 specifies the relationship between the coverage domain and the associated range of attributes. The standard defines the characteristics of the spatial domain, but omits the description of the range of mapped attributes.

WCS provides access to potentially detailed and rich sets of geospatial information in forms that can be used for client-side rendering, creation of multi-value coverages, or as input data for other software, such as simulation models. Similarly to the OGC Web Map Service (WMS) and Web Feature Service (WFS), WCS allows the clients to download selected elements of information available on the server side based on spatial constraints as well as other criteria.

However, in contrast to WMS, which filters and processes spatial data into the form of static maps (rendered by the server as bitmap images), the Web Coverage Service provides access to data along with its detailed descriptions, allows the formulation of complex queries regarding this data, and returns the data retaining its original semantics. This allows the client to not only visualize the data, but also process it.

In comparison to WFS, which returns geospatial data in the form of discrete objects, Web Coverage Service differs in that it provides representations of spatially variable phenomena whose spatio-temporal domain is mapped by a range of attributes.

The main advantage of WCS over the latest version of the WMS standard (which also supports multi-dimensional data) is a greater range of supported data formats. As in the case of WMS, it can return both 3D and 4D data, but they can reach the client in raster as well as tabular formats such GML or HDF. This enables further analysis and processing of the data on the client side.

While the WCS allows for retrieval of spatio-temporal geographical data, it does not provide the means of specifying OLAP-style queries on this data. This issue, however, may be alleviated by employing a companion standard known as Web Coverage Processing Service (WCPS).

WCPS is an OGC standard which defines a language for filtering and processing of multidimensional coverages, in particular sensor data, simulation results and statistics. Similarly to the WCS syntax, this query language allows clients to retrieve original coverage data as well as the results of its processing over the Web in a manner independent of software and hardware platforms.

The WCPS standard supports 1D data streams, 2D objects, spatial 3D cubes (x = y = z) and 4D spatio-temporal cubes (x = y = z = t). It allows the client to formulate queries with multiple levels of complexity, including inter alia data processing.

Similarly to the case of standard WCS syntax, in response to queries WCPS may generate both raster and tabular data. However, in contrast to the standard WCS syntax, WCPS syntax and semantics allow the formulation of generated queries and automatic chaining of services.

As a language, WCPS is not tied to any particular transmission protocol, which allows queries created in it to be embedded within any service. In particular, existing standards such as the OGC Web Coverage Service and OGC Web Processing Service are commonly used for this purpose.

At the time of writing the current version of the standard was WCPS 1.0, which allows for sending queries in the form of character strings to the server and retrieving sets of coverages in response. The query may be expressed in abstract syntax as well as in the form of an XML document.

The level of complexity represented by WCPS allows for expression of classification algorithms, kernel filters, convolution filters, histogram operations and Discrete Fourier Transforms. It also allows the construction of OLAP-style queries using operations such as slicing, dicing, drilling and pivoting.

Because the WCPS standard is relatively young, the number of its implementation is currently quite limited. Practically the only server which fully supports WCPS is rasdaman, which is maintained by rasdaman GmbH. rasdaman is a reference implementation of the WCPS standard which also supports WCS, WCS-T and WPS.

Because typical, popular database management systems (DBMS) do not provide sufficient methods of storing MDD, and most of the available solutions focus on two-dimensional (2D) instead of n-D data [2][4][5][6], a new tool dedicated to storing, processing and transfer of spatio-temporal data was needed.

2. SYSTEM ARCHITECTURE

Nowadays processing of hydroacoustic data is most often conducted with use of dedicated software which offers a narrow and specialized set of operations Because such solutions are proprietary, their interoperability is limited. In addition, since the dedicated software is distributed in the form of standalone applications, the end user needs to assure the availability of appropriate hardware for performing advanced computations. In this context, such shortcomings may be alleviated by implementing a Software as a Service (SaaS) model for distribution of software for processing of multidimensional data. The architecture of the proposed system including the SaaS model of distribution is presented in Fig. 2.

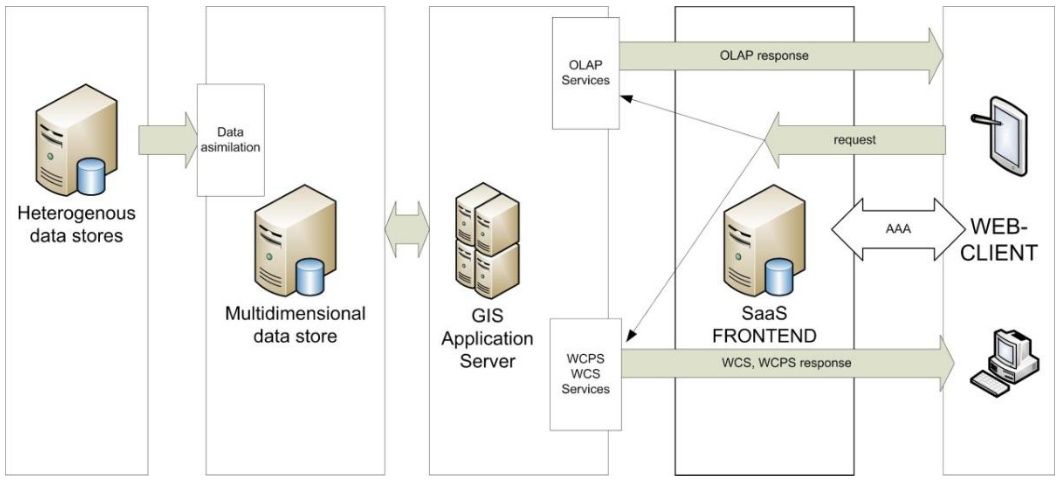

Fig.2. System architecture.

Because the proposed system operates in the SaaS model, it is delivered to the end-user through a web browser. Such a solution is assures that the system is accessible on many different software and hardware platforms. The thin client is only responsible for non time-consuming tasks such as providing a Graphical User Interface (GUI). The GUI allows the end user to choose data to be analysed as well as the desired algorithm of its processing. This includes regular operations like cuts and selections as well as dedicated operations encoded in the WCPS, WPS and OLAP.

After the user requests some processing operations to be performed, a query is issued to the SaaS Frontend. The Frontend is responsible for user authentication and authorisation as well as accepting and tunneling the user requests. Depending on the type of the service, derived from the request (WCPS, WMS, WPS, OLAP services) the processing is forwarded to the particular working modules of the system.

The GIS Application Server is realized by means of Rasdaman [7], which provides a database service for MDD structures in a domain independent way. Rasdaman consists of several components that manage software features. Definition of the MMD is handled by Rasdaman definition language, RasDL, tool for maintain data schema. User can define data types that will describe an unrestricted number of instances; no other information is required to access database [3]. Schema, defined by RasdDL, consists of three elements: cell type definitions (similar to C++ primitive types can be used), MDD type definitions (array over base type and with spatial domain can be created) and collection type definitions (which describe a set of arrays of the same type). User MDD data is processed by Rasdaman and stored in the underlying database, called base database. Community version of Rasdaman can be easily embedded in PostgreSQL while commercial also in MySQL, Oracle, IBM Informix and DB2 base databases. Alphanumerical information that are not arrays are stored outside Rasdaman, but reference between alphanumeric and MDD data can be established [3].

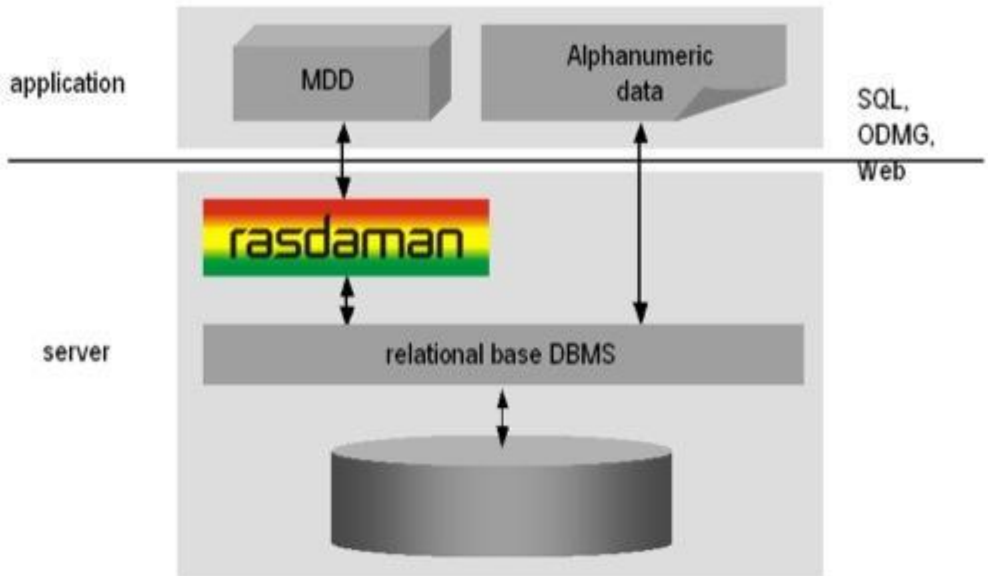

Fig.3. Embedding of Rasdaman in IT infrastructure [3].

The hydroacoustics data is stored in one original, proprietary form (data files) inside the heterogenous store. At regular intervals and on request the data is synchronized and replicated from the heterogenous store to the common multidimensional database. Rasdaman interprets multidimensional data as spatio-temporal cubes, where the first two dimensions are X and a Y pixel position, third dimension is a time parameter. In this context a section operation can be defined (Fig.4). A section allows to extract lower- dimension layers (also called slices) from n-D array. Operation can be performed on any of the dimensions and is accomplished by specifying a slicing position. One reduces dimension of data received from this operation.

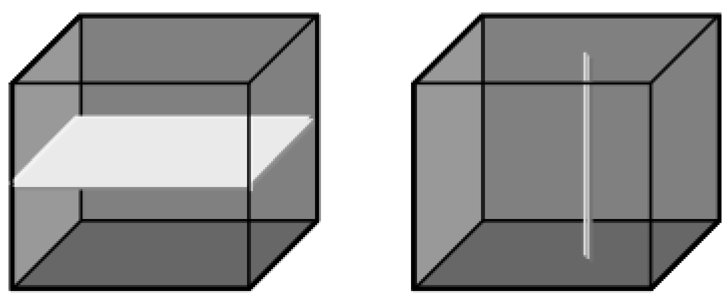

Fig.4. Data Cube section [3].

After a request for processing is received, the working module process the MDD data and returns the results in WCS form. Due to efficiency reasons, some modules may partly replicate the data to ensure the higher responsiveness of the system. For example, the presentation layers are also available in the form of WMS services.

3. RESULTS

The following section contains sample results of processing hydroacoustic data by means of the proposed general-purpose system. After the spatio-temporal cube was constructed, the Graphical User Interface (GUI) was used for constructing sample queries [8]. Fig. 5 presents the main window of the query editor, which also allows for establishing database connection.

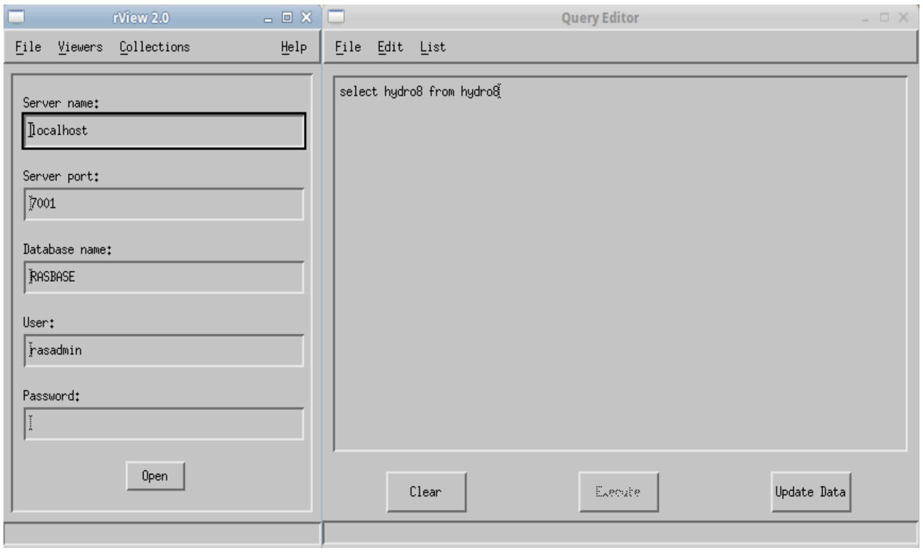

Fig.5. GUI connection window and Query Editor.

After a query has been executed, its results are returned to the user. A specialized tool for their preview is presented in Fig. 6. The results window presents information about obtained data such as its type, resolution, size and spatial domain.

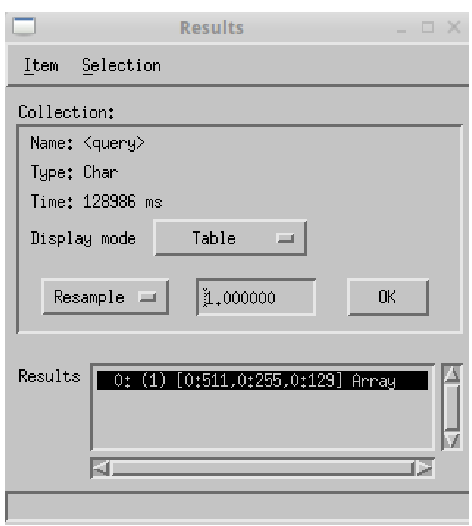

Fig.6. GUI Results window.

The GUI is equipped with several specialized tools for data visualization, which allow the user to choose from following activities: flat image, volumetric, orthosection, height field, chart, table, sound, string. Possible visualizations are presented in the Fig. 7.

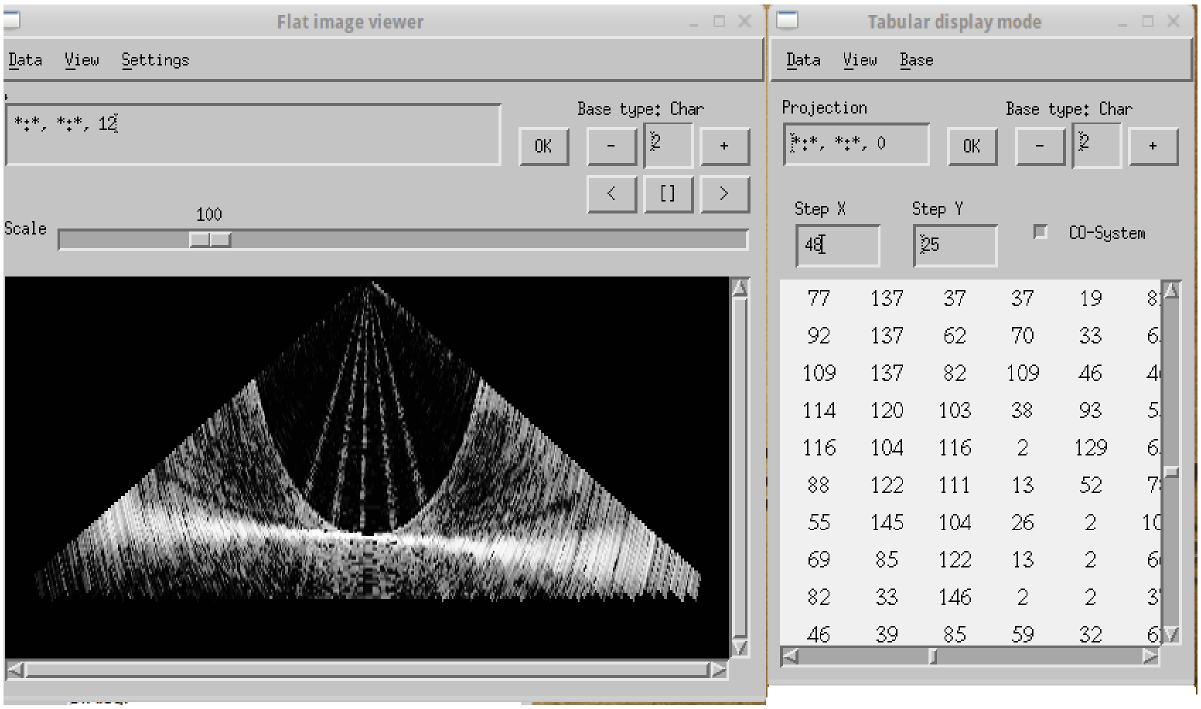

Fig.7. Flat image viewer and Tabular display mode windows.

The utilized sample set of hydroacoustic data contains 130 consecutive swaths, each sized 512 x 266 pixels. Assuming that time series data is located in "img" directory, a 3D data cube called "hydro8" can be created with following formula:

rasimport -d img -coll hydro8 -3D top

Spatial domain of the created cube can be obtained by following formula:

rasql -q "select sdom(m) from hydro8 as m" --out string
rasql: rasdaman query tool v1.0, rasdaman osgeolive.5 -- generated on 21.02.2013 02:03:59.
opening database RASBASE at localhost:7001...ok
Executing retrieval query...ok
Query result collection has 1 element(s):
Result element 1: [0:511,0:255,0:129]

When the hydroacoustic information has been properly converted into the form a data cube, its analysis can be performed. In the presented examples, the hydroacoustic cube has been sliced in every dimension. Fig. 8 presents a section operation performed on the time dimension, which creates one of the time series images representing a single sounding. The command for producing the image from the data cube is as follows:

rasql -q "select png(m[:*, *:* ,10]) from hydro8 as m" --out file*

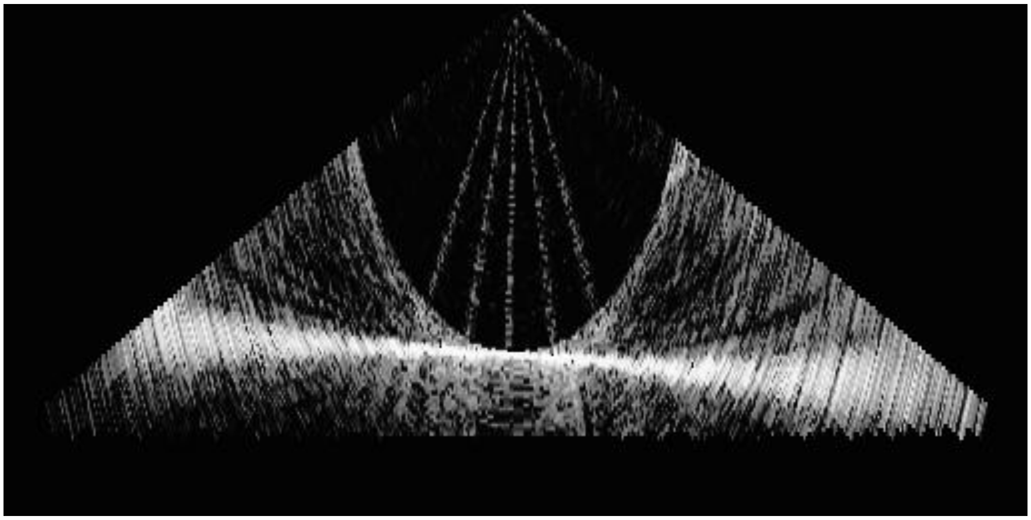

Fig.8. A section operation performed on the time dimension allows to view a single sounding.

More interesting results are received after slicing the cube in the vertical dimension, represented by coordinate X. Fig. 9 represents a section performed for a parameter value corresponding to the position of the ship, which shows the seabed and water column along the ship's course. This information may be acquired using the following formula:

*rasql -q "select png(m[256, *:* ,*:*]) from hydro8 as m" --out file*

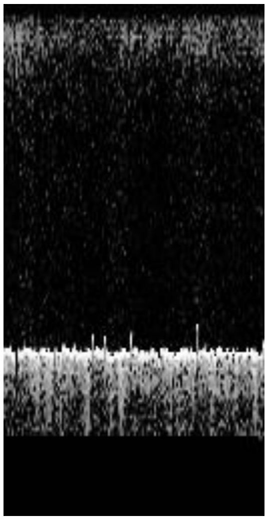

Fig.9. A vertical slice of the data cube shows the seabed along with water column data.

In a similar manner, the following command:

rasql -q "select png(m[:*, 127 ,*:*]) from hydro8 as m" --out file*

Produces a horizontal slice of the cube data, which allows the user to view e.g. the seabed along the ship's course, as shown in Fig. 10.

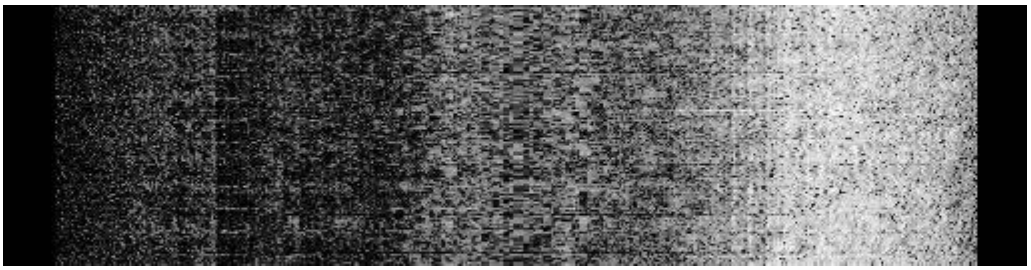

Fig.10. A horizontal slice of the data cube showing the seabed.

The utilization of OLAP queries over WCPS allows for acquiring the query results in various popular formats. For example, numerical data in a coma-separated values format (CSV) (Fig. 11), can be obtained from cube with the following formula:

rasql -q "select csv(m[200:400,160:200,12]) from hydro8 as m"--out string

```
user@osgeolive:~$ rasql -q "select csv(m[200:400,160:200,12] ) from hydro8 as m"
 --out string
rasql: rasdaman query tool v1.0, rasdaman osgeolive.5 -- generated on 21.02.2013
 02:03:59.
opening database RASBASE at localhost:7001...ok
Executing retrieval query...ok
Query result collection has 1 element(s):
  Result object 1: {135,125,117,111,104,90,77,93,109,197,230,238,244,243,245,233
,212,181,144,141,135,79,20,28,33,23,16,69,137,136,135,51,2,18,37,18,2,2,2,29,77}
,{45,19,2,36,104,119,135,67,13,183,249,249,249,247,245,243,240,211,170,177,178,1
59,135,154,166,154,128,150,160,152,137,104,54,22,2,2,2,32,82,32,2},{137,104,54,1
12,160,74,17,74,160,179,201,233,249,231,190,190,198,197,201,145,75,121,166,159,1
44,159,166,121,82,113,135,123,109,83,54,68,82,35,9,8,2},{54,22,2,19,45,98,135,12
5,116,214,247,247,245,220,190,203,211,195,174,174,174,122,77,122,164,122,73,112,
143,114,81,32,2,29,77,38,13,11,9,11,13},{117,111,104,124,152,147,135,168,197,231
,247,247,247,229,188,193,201,197,190,182,174,122,104,47,16,37,73,59,45,65,77,76,
73,76,77,29,2,2,2,2,2},{135,107,65,29,9,18,27,70,164,206,233,246,250,247,245,221
```

Fig.11. Data in CVS format.

4. CONCLUSIONS

The article shows that a general-purpose Geographic Information System working in a System as a Service architecture is a versatile tool for processing multidimensional data. The proposed system uses open standards such as OLAP, WCS, WMS and WCPS, which enables it to perform a variety of operations on spatial data, including a wide array of queries and format transformations.

The results of performed analyses have shown that such a general-purpose system may be successfully applied to the analysis of hydroacoustic data. A time series of data acquired via a multibeam echosounder was transformed into an OLAP cube, and was used to perform a variety of slicing operations. The analysis performed on a multidimensional cube has shown to produce results similar to those acquired via the use of dedicated software. An added value of the presented system is that the data is stored within a GIS, which enables its immediate visualization in a geographical context as well as further processing by any GIS software which supports open standards. Although the presented results are preliminary, they show great promise regarding the future of general-purpose GIS software.

ACKNOWLEDGEMENTS

The contribution was sponsored by The National Centre for Research and Development as project for the purposes of national defence and security.

REFERENCES

[1] J. Thomas, K. Cook, Illuminating the Path: The Research and Development Agenda for Visual Analytics, National Visualization and Analytics Ctr, 2005.

[2] P. Baumann, A. Dehmel, M. Höfner, N. Widmann, Exploring Sensor and Simulation Raw Data with RasDaMan, Extending Database Technology (EDBT) 2000, 2000.

[3] P. Baumann, Query Language Guide, http://rasdaman.eecs.jacobs-university.de/trac/rasdaman/browser/manuals_and_examples/manuals/pdf/ql-guide.pdf [Accessed 25.03.2013]

[4] L. Libkin, R. Machlin, L. Wong, A query language for multidimensional arrays: Design, implementation, and optimization techniques, Proc. ACM SIGMOD'96, 228 – 239, 1996.

[5] A.P. Marathe, K. Salem, Query Processing Techniques for Arrays, Proc. ACM SIGMOD'99, 323-334, 1999.

[6] S. Sarawagi, M. Stonebraker, Efficient Organization of Large Multidimensional Arrays, Proc. ICDE'94, 328-336, 1994.

[7] http://www.rasdaman.org/ [Accessed 20.03.2013]

[8] P. Baumann, rView the rasdaman DBMS Visual Frontend, http://rasdaman.eecs.jacobs-university.de/trac/rasdaman/browser/manuals_and_examples/manuals/pdf/rview-guide.pdf [Accessed 28.03.2013]

[9] http://www.esri.com/software/arcgis/defense-solutions [Accessed 04.04.2013]

[10] M. Zhang, C. He, G. Liu, Application of GIS in electronic commerce, Second IITA International Conference on Geoscience and Remote Sensing, 483-486, 2010.

[11] http://www.nrccwdt.org/2011/10/using-gis-for-policy-and-planning/ [Accessed 06.04.2013]

[12] http://www.tatukgis.com/ [Accessed 12.04.2013]

[13] M. Kulawiak, M. Moszynski, Integration of Geographic Information Systems for Monitoring and Dissemination of Marine Environment Data, Geospatial Techniques for Managing Environmental Resources, 33-52, 2011.

[14] P. A. Burrough, R. McDonnell, Principles of geographical information systems (Vol. 333), Oxford university press, 1998.

[15] http://www.olapcouncil.org/research/resrchly.htm [Accessed 05.04.2013]
[16] http://en.wikipedia.org/wiki/OLAP_cube [Accessed 10.04.2013]
[17] W. Hümmer, A. Bauer, G. Harde, XCube: XML for data warehouses, DOLAP '03 Proceedings of the 6th ACM international workshop on Data warehousing and OLAP, 33-40, 2003.
[18] M. Kulawiak, A. Prospathopoulos, L. Perivoliotis, M. Luba, S. Kioroglou, A. Stepnowski, Interactive visualization of marine pollution monitoring and forecasting data via a Web-based GIS, Computers & Geosiciences Vol. 36(8), 1069-1080, 2010.
[19] M. Kulawiak, A. Chybicki, M. Moszyński, Web-based GIS as a tool for supporting marine research, Marine Geodesy, Vol. 33(2,3), 135-153, 2010.