A Specification for Developing a New Method of Visualising Data Quality in Electronic Navigational Charts

Sam HARPER, United Kingdom

SUMMARY

With any physical measurement comes a level of uncertainty as to its accuracy. In the context of hydrographic surveying for navigational purposes, this uncertainty propagates through from data acquisition to data processing and on to chart compilation, increasing and becoming harder to quantify as it goes. It is therefore imperative that we have meaningful, useful and intuitive methods of representing this uncertainty so that the end user, the mariner, understands the limitations of the data by which he navigates.

The representation of geospatial data quality in a GIS environment is well researched with many different methods employed. The same is not true for the representation of data quality in Electronic Navigational Charts (ENCs). There is a concern amongst the international hydrographic community that the current methods of representing data quality in navigational products are not meeting the needs of the mariner. Instead they rely heavily on the user's ability to understand the relevance of data quality indicators such as 'survey date' and 'acquisition method' or composite quality classifications like CATZOC.

This study discusses the results of a questionnaire on the mariner's current perception of data quality in both paper and digital charts. The questionnaire was distributed internationally amongst a broad range of both professional and leisure mariners. With over 550 responses, 67% of which are from mariners with over 10 years of experience, it has been possible to identify aspects of current data quality representation that are not fulfilling the mariner's needs with respect to safe navigation. In conclusion, a specification is suggested for the development of a new approach to representing data quality in ENCs.

1. INTRODUCTION

Current methods of representing the quality of the source data used to compile a navigational chart fail to provide mariners with the information they need to objectively decide where they can safely navigate. Instead they rely heavily on the mariners' ability to understand the relevance of data quality indicators such as 'survey date' and 'acquisition method' or composite quality classifications like CATZOC (Category of Zone of Confidence).

With the use of electronic navigational charts on the increase, and ECDIS mandation on the horizon, never has it been more necessary that we ensure that the limitations of charted data are fully understood.

As bathymetric data acquisition systems become ever more sophisticated, understanding the implications of combined uncertainty and error become more complex. It is unreasonable to expect the professional mariner to be able to assimilate all of this extra information and draw valid inferences from it. Instead we need to better understand their requirements and expectations, and utilise developments in technology to develop better means of representing data quality.

In 2011, the International Hydrographic Organisation's (IHO) Data Quality Working Group (DQWG) undertook a study into the Mariners' perception of data quality. The principle aim of this project was to develop and recommend a specification for the development of any new means of representing data quality in future ENCs. This specification would take into account why mariners need data quality information, how mariners currently use data quality information, what mariners need from data quality information and the limitations of providing data quality information.

The results of this study and the specification derived from them are detailed in this paper.

2. BACKGROUND

2.1 Data Quality and Uncertainty as a Concept

When we talk about representing data quality we are in essence trying to depict in a contextualised manner the total combined uncertainty of the bathymetric data from acquisition through to compilation. This is highlighted by Pang et al. [1997] who identified three instances in the "visualisation pipeline" where uncertainty is encountered: collection uncertainty due to measurements and models in the acquisition process, derived uncertainty arising from data processing and manipulation (cleaning, gridding etc.), and visualisation uncertainty introduced during the process of chart compilation. In addition to these three instances a fourth can be identified; once charted, data quality will suffer temporal degradation due to changes in seabed topography.

2.2 Challenges in representing data quality and uncertainty

Buttenfield [1993] suggests that there are three problems with the effective representation of uncertainty; Firstly uncertainty itself is an "ill-defined concept" with little distinction being made between similar concepts. It is also the case that the terminology used to describe these concepts is poorly understood and frequently misused. Secondly, it is difficult to measure multiple aspects of uncertainty, such as temporal degradation and random error, in a geospatial environment. Thirdly, it is difficult to represent uncertainty simultaneously with the data it describes.

MacEachren [2005] goes further than Buttenfield [1993] and identifies 7 more challenges:

1. Understanding the components of uncertainty and their relationships to domains, users, and information needs.
2. Understanding how knowledge of information uncertainty influences information analysis, decision making, and decision outcomes.
3. Understanding how (or whether) uncertainty visualisation aids exploratory analysis.
4. Developing methods for capturing and encoding analysts' or decision makers' uncertainty
5. Developing representation methods for depicting multiple kinds of uncertainty
6. Developing methods and tools for interacting with uncertainty depictions
7. Assessing the usability and utility of uncertainty capture, representation, and interaction methods and tools.

2.3 Existing methods of representing data quality in navigational charts

There are many types of navigational products available, each with different methods of representing data quality to the mariner. Generally these products fall into two categories; official government endorsed products and non-official products. The representation methods described below are found in official government endorsed products and as such their use is controlled by international standards.

2.3.1 Source or reliability diagram

Figure 2.3.1 shows an example of a source diagram, as found on a British Admiralty Paper Chart. It shows the individual areas of survey coverage, along with the year of completion, Survey authority, scale and sometimes acquisition method. In order for this to be useful the mariner must be able to infer from these data quality indicators what affect they will have on the quality of the survey.

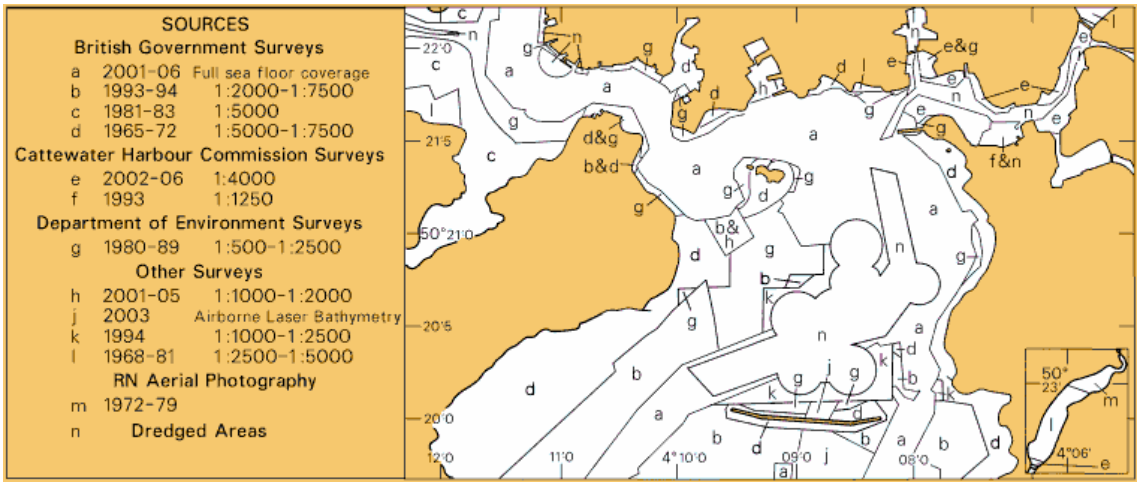

Figure 2.3.1 Source Diagram

2.3.2 Category of Zone of Confidence

Category of Zone of Confidence (CATZOC) is the primary indicator of data quality in ENCs. It is an S-57 attribute that is populated with a composite indication of the quality of the bathymetric data in a specific area. It differs from the source diagram in that it has it gives an overall indication of the quality of the charted data rather than providing individual data quality indicators. The various CATZOC categories are summarised in Table 2.3.1. CATZOC also has its own symbology and it can be toggled on and off depending on the preferences of the user.

Table 2.3.1 CATZOC Descriptions

ZOC	Position Accuracy	Depth Accuracy	Seafloor Coverage
A1	± 5m + 5% depth	0.5m + 1% depth	Full area search undertaken. Significant seafloor features detected and measured.
A2	± 20m	± 1m + 2% depth	Full area search undertaken. Significant seafloor features detected and measured.
B	± 50m	± 1m + 2% depth	Full area search not achieved; uncharted features, hazardous to surface navigation are not expected but may exist.
C	± 500m	2m +5% of depth	Full area search not achieved, depth anomalies may be expected.
D	Worse than ZOC C	Worse than ZOC C	Full area search not achieved, large depth anomalies may be expected.
U	Unassessed – The quality of the bathymetric data has yet to be assessed.		

Designating CATZOC values for charted areas is at least a partially subjective process; no more is this the case than when it comes to assessing legacy data. As a precursor to this research, a study was carried out to establish by what criteria CATZOC is being designated for legacy data by ENC producing National Hydrographic Offices [Harper 2010]. This research showed that there is significant variance in the way in which legacy data is designated with a CATZOC value. A consequence of this is that a mariner navigating across an ocean may be using data with the same CATZOC value, but be unaware that there are differences in the actual data quality.

2.3.3 Zone of Confidence Diagram

The zone of confidence diagram (figure 2.3.3) appears on some paper charts and delimits general areas of data quality in the same way that CATZOC does. As a consequence it suffers the same short comings as CATZOC with the exception of symbology and the ability to toggle it on and off.

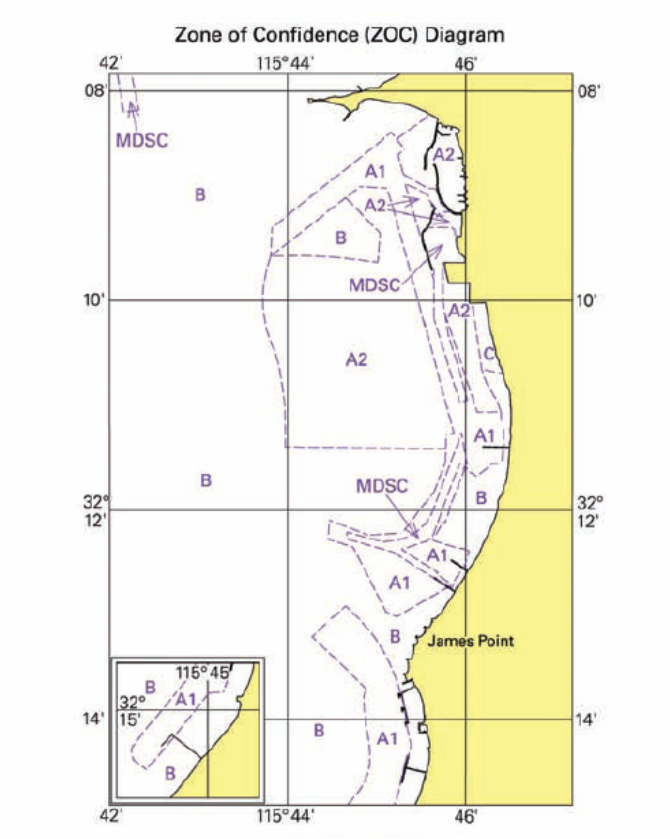

Figure 2.3.3 Zone of Confidence diagram

2.3.4 Data Quality Symbology

There are various symbols, legends and notes that supplement the information found in the source diagram or CATZOC display. These symbols and legends are often used to indicate data quality issues that relate to a specific feature, e.g. a reported depth note. For British admiralty products the mariner can find information on these symbols in BA NP 5011 (UKHO, 2004).

It is unknown how well understood the symbology that relates to data quality is. As it is entirely possible that users of nautical products are unaware of the relevance some symbols have in relation to data quality, this issue was explored in the questionnaire.

3. RESEARCH METHODOLOGY

The main investigative element of the study took the form of a questionnaire. The questionnaire was developed to support the aims and objectives of the project by facilitating the investigation of:

- The mariners' perception and understanding of data quality representation in navigation products
- The mariners' opinion of data quality education and information availability
- The mariners' preferences with regard to future methods of representing data quality in navigational products

The questionnaire was distributed by the IHO to member states, and was available as a PDF and an on-line version via surveymonkey.com. Over 600 responses were received, however due to time constraints the analysis was based on 574 responses.

A 'QUANqual' mixed methods approach was taken with the design of the questionnaire. The qualitative questions can be subdivided into two types:

- Those designed to elaborate on or give context to quantitative questions, e.g. 'other' and 'please explain your answer' free type fields
- Those designed to directly test the respondents' knowledge of data quality issues, e.g. 'what does the PA abbreviation mean?'

The qualitative analysis took the form of the identification of recurring themes and the ranking of these themes by their frequency of occurrence.

4. RESULTS

4.1 Demographics

In terms of the survey sample, the demographic information showed that 74% (421 respondents) had over 10 years navigational experience with 63% (357 respondents) having in excess of 15 years navigational experience. In addition the results showed that a broad range of shipping sectors were represented. As a consequence, it is considered that a strong representative sample has been collected.

4.2 Paper Charts

Respondents who said that they use paper charts were asked whether the charts they use have either a source/reliability diagram or a zone of confidence (ZOC) diagram. The respondents that answered yes to these questions were then asked to indicate whether they use the information in the source/reliability diagram or a ZOC diagram. Figures 4.2.1 and 4.2.2 show that 73% (296 respondents) of respondents use the information in the source/reliability diagram and 75% (82 respondents) of respondents use the information in the ZOC diagram.

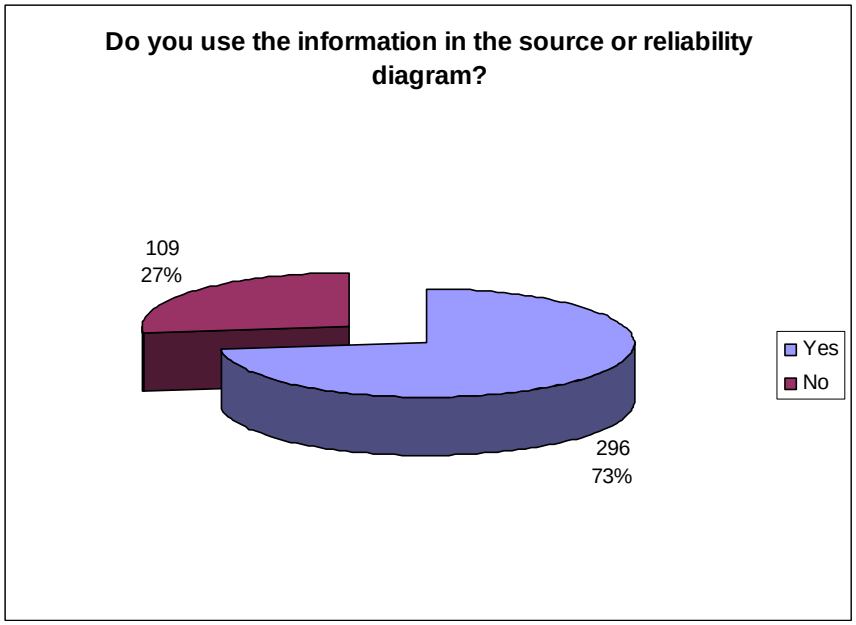

Fig. 4.2.1 Percentage of respondents that use the information in the source/reliability diagram

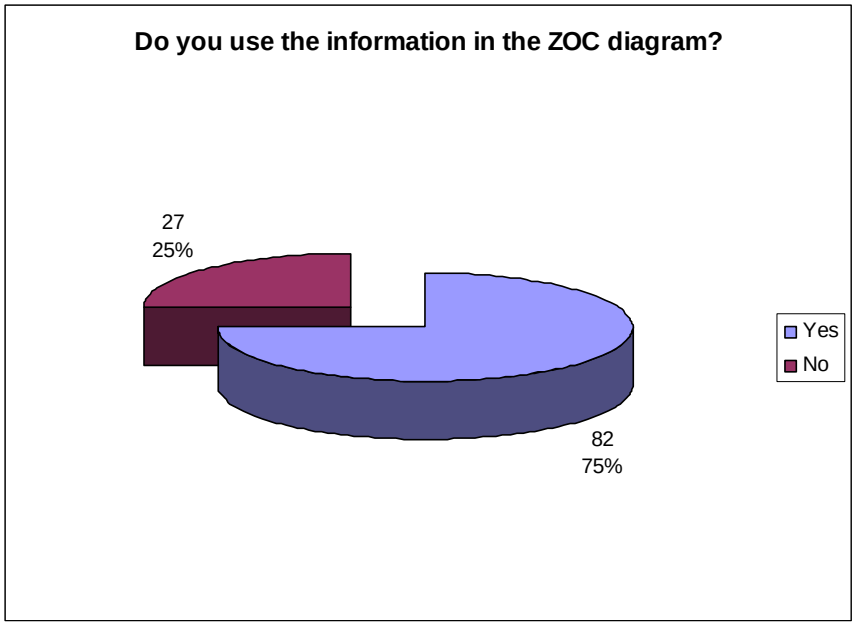

Fig. 4.2.2 Percentage of respondents that use the information in the ZOC diagram

Respondents that indicated that they did not use the information in the source/reliability diagram or ZOC diagram were then asked to explain why not via a multiple choice question. The most common reason chosen by respondents was "because I have travelled the same route many times before". A number of respondents selected the 'other' free type option and the themes arising from these answers are detailed in table 4.2.1. The most common reason cited was that "I trust that the charts are correct".

Table 4.2.1 Themes and ranks for why respondents do not to use the information in the source /reliability diagram

Theme	**Rank**
"I trust that the charts are correct"	1
"We are restricted by the Pilots limited area of operation and bow to their local knowledge"	2
"We rely upon experience and instruments instead"	3

Respondents were presented with a series existing data quality indicators that appear on paper charts and were asked to indicate whether they understood their meaning. Those that said that they did were then asked to give an explanation of the meaning of the respective indicator. These answers were then marked as either correct or incorrect. Table 4.2.2 shows a summary of these results. Those figures coloured red indicate where the percentage of respondents who gave incorrect explanations is greater than 60%. The figures that are coloured amber indicate where the results were between a 41% to 59% split. The figures coloured green indicate that either the

number of respondents who indicated that they understood the data quality indicator or those that gave a correct explanation exceeded 60%.

It should be noted that due to an oversight in the design of the questionnaire, respondents were asked “do you understand the meaning of the Unsurveyed and Depths notes?” This has meant that the values for the first part of the question are the same for both indicators. However, respondents were given the opportunity to explain their meaning individually. Regrettably, the same situation occurred question relating to the PA, PD, ED, SD and Rep’d (1999) notes.

Table 4.2.2 Summary of results to questions relating to mariners’ understanding of existing data quality indicators in paper charts

	Do you understand the meaning of…?		**Of those who answered yes, how many gave a correct explanation?**	
Data Quality Indicator	**Yes (%)**	**No (%)**	**Correct (%)**	**Incorrect (%)**
Broken depth contour symbol	56	44	73	27
Broken coastline symbol	66	34	69	31
Dotted danger line symbol	76	24	44	56
Discontinuity between surveys note	53	47	55	45
Unsurveyed note	88	12	94	6
Depths note	88	12	74	26
PA	62	38	98	2
PD	62	38	90	10
ED	62	38	82	18
SD	62	38	79	21

Rep'd (1999)	62	38	36	64
Sounding in an upright font	44	56	36	64
Discoloured water note	59	41	Corrupted	Corrupted
Sandwave symbol	64	36	91	9
Dredged to… note	98	2	98	2
Potentially dangerous wreck symbol	98	2	76	24
Bar above a dangerous wreck symbol	75	25	57	43
Works in progress legend	93	7	100	0

Generally the understanding of existing paper chart data quality indicators appears to be good, however the understanding of the 'dotted danger line symbol', 'discontinuity between surveys note' and the 'bar above a dangerous wreck symbol' appear to be marginal. Further, the respondents' understanding of the 'Rep'd (1999)' abbreviation and soundings in an upright font could be considered poorly understood.

The poor understanding of the 'Rep'd (1999)' abbreviation is attributed to the fact that answers not including the condition 'but not confirmed' were marked as incorrect. The question of whether a mariner would react to the Rep'd abbreviation in a different way to any other sounding is also raised.

The Sounding in an upright font was commonly misinterpreted as indicating that the value was in a different class of units (imperial or metric) to the rest of the data.

It was noted that the marking of these answers was a subjective process and as a consequence it is plausible that a different marker (from a different area of expertise) may generate slightly different figures.

4.3 ENCs

In contrast to the questions relating to source/reliability and ZOC diagrams, the results show that a large portion of ENC users (77%) do not use S-57 CATZOC (Figure 4.3.1). Further, sector analysis showed that percentage is fairly stable regardless of number of years of experience.

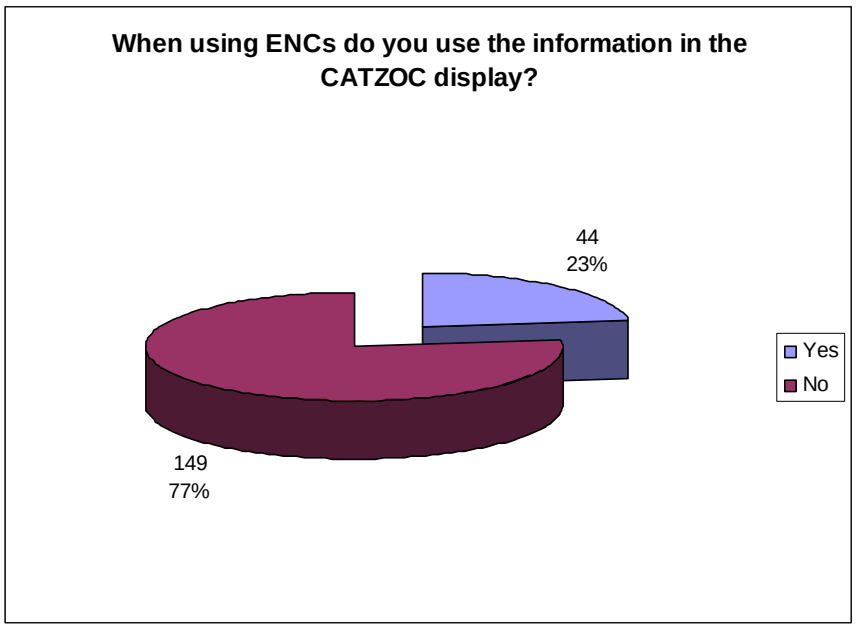

Fig. 4.3.1 Percentage of respondents that use the CATZOC display

As with paper chart DQIs, respondents were asked to indicate whether they understood the meaning of a range of S-57 data quality attributes. Those that said that they did were the asked to give an explanation of the meaning of the respective attribute. The results, detailed in table 4.3.1, show very poor understanding of the S-57 acronyms.

Table 4.3.1 Summary of results to questions relating to mariners' understanding of existing S-57 data quality attributes

	Do you understand the meaning of…?		**Of those who answered yes, how many gave a correct explanation?**	
S-57 Attribute	**Yes (%)**	**No (%)**	**Correct (%)**	**Incorrect (%)**
HORACC	24	76	57	43

POSACC	29	71	60	40
SOUACC	31	69	91	9
VERACC	22	78	78	22
SURATH	42	58	91	9
SURSTA	32	80	94	6
SUREND	21	79	94	6
TECSOU	43	57	96	4
QUASOU	31	69	78	22
QUAPOS	27	73	79	21

3.4 Wider Data Quality Issues and Future Developments

On the issue of training, 66% (183 respondents) indicated that they felt they had received insufficient training on data quality. This was reinforced by 78% (216 respondents) indicating that they would like to receive further training on data quality. The DQWG are currently investigating how training on data quality is delivered and what mechanisms for delivering further training to practicing mariners could be utilised.

Mariners were presented with a variety of conceptual future methods for representing data quality and invited to comment upon the various options. In general respondents seemed to favour an on demand data quality colour overlay.

5.0 CONCLUSIONS AND SPECIFICATION

Analysis of the results has allowed us to draw the following key conclusions:

- Large proportions of ENC users are not using the CATZOC information (77%)
- A large proportion (75%) of mariners that use charts with a ZOC diagram stated that they do use the information contained within it. This suggests that it is the digital application of CATZOC that mariners don't like, and that there is no clear preference for individual quality indicators over composites ones.

- The additional S-57 DQ indicator attributes are not understood and not used
- Whilst the results would suggest that mariners are aware of the relevance of seabed morphology, it is not clear that they understand the variable degradation of data quality with time.
- Majority of mariners state that they have not received enough training on data quality issues, and that they would like to receive more training
- With high percentages of mariners indicating that they use the information in the source and ZOC diagrams versus the low percentage of mariners using the CATZOC display, it is possible that we can identify a preference for the clear delimitation of uniform areas of data quality. This is supported by the results of the future developments section which showed the favourite option to be an 'on-demand colour overlay'.

5.1 Specification for Developing New Methods of Representing Data Quality in ENCs

Using the results and conclusions from the questionnaire, the DQWG has developed the following draft specification for developing future methods of representing data quality in ENCs. These recommendations are meant to bring in new possibilities for implementation into ECDISs.

- As a minimum the constituent elements of S-57 CATZOC (positional uncertainty, sounding uncertainty, features detected and seafloor coverage) must be encoded in S-101 ENCs for depth areas, as separate attributes
- All encoded data quality information must be discoverable
- Temporal degradation of data should be encoded
- New representation methods should be able to accommodate inputs such as dynamic tides, under keel allowance and vessel specific parameters. It is understood that international efforts on standardization of display and mariner training address possible issues with user inputs.
- Where possible ENC attribute names should be more descriptive (eliminate 6 letter acronyms and make use of camelCase)
- Visualisation should take advantage of the mariner's preference for an on demand colour overlay
- Recommend to add ability for mariners to add notes to specific features, that again changes presentation of the feature (as an addition to the mariners' objects)
- Any representation method should be accompanied by an appropriate education strategy

6.0 REFERENCES

[Buttenfield, 1993] Buttenfield, B.P., (1993) *Representing Data Quality*. Cartographica (special content, Mapping Data Quality) **30(**2&3), 1-7

[Harper 2010] Harper. S., (2010). *DQWG4-05A Analysis of Responses to IHO DQWG CL59/2010.* Data Quality Working Group Paper, IHO Data Quality Working Group.

[MacEachren 2005] MacHeachren, A.M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., and Hetzler, E., (2005). *Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know.* Cartography and Geographic Information Science **32**(3), 139-160.

[Pang, 1997] Pang, A.T., Wittenbrink, C.M., and Lodha, S.K., (1997). *Approaches to Uncertainty Visualization.* Visual Computer **13**, 370-390

[UKHO, 2004] United Kingdom Hydrographic Office, (2011). *Symbols and Abbreviations used on Admiralty Charts.* Admiralty Charts and Publications, Chart 5011, 4th Edition

BIOGRAPHICAL NOTES

Sam Harper works as a Senior Bathymetric Appraisal Officer for the United Kingdom Hydrographic Office Seabed Data Centre and is a graduate of Plymouth University's MSc Hydrography programme. He is also the UKHO representative on the IHO Data Quality Working Group, a Director of the Hydrographic Society UK and the Hon. Secretary of the Hydrographic Society UK South West Region. Prior to specialising hydrography, Sam worked in GIS and marine science. His other interests include, flyfishing, snowboarding and spearfishing.

CONTACTS

Mr Sam Harper
United Kingdom Hydrographic Office
Admiralty Way
Taunton
UNITED KINGDOM
Tel. +44 (0)1823 337900 Ext. 3843
Fax +44 (0)1823 284077
Email: samuel.harper@ukho.gov.uk
Web site: www.ukho.gov.uk