Towards Data Centricity

Executive Summary:
This paper outlines the need for HOs to think in terms of data, rather than products. Such thinking will affect the internal management of data within HOs, leading to greater possibilities for effective exploitation of valuable data assets, and improved interoperability.

Background

1. All HOs depend on data, but mostly think in terms of products. The creation of ENCs requires chart information to be decomposed into features and attributes, but the output is still a product, albeit data sets. The delivery of these data sets is also on a cell-by-cell basis, characterized by scale. They are therefore restricted in their use, by their design.

2. Most HOs focus on supplying products to a narrow group of users. The driving force is navigational safety, with any additional use, being an opportunistic spin-off.

From Data to Knowledge

3. The principle of data being the foundation of knowledge is well known (figure 1). In essence, knowledge is only of value if conveyed, and as the pyramid shows, a broad base of data is required to extract a smaller volume of knowledge. What is less often articulated is the amount of redundancy in many knowledge systems, where far more data is held than actually converted into conveyed knowledge (figure 2). The conversion of detailed bathymetric surveys into charts with sparse soundings, is a good example of this. A large amount of data is collected, and although it presents a lot of information to the compiler, only a small amount of the knowledge is passed on to the recipient of the product. Thus, the knowledge transfer is only a small part of the potential of the original data.

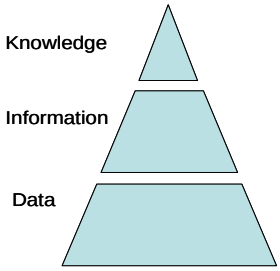


Figure 1

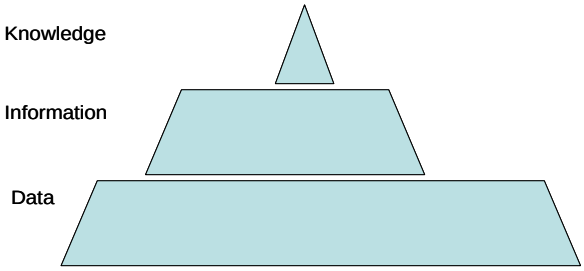


Figure 2

4. Most data sets have the potential of conveying a range of categories of knowledge. Continuing with the bathymetric survey example, in addition to providing knowledge relevant to a navigational chart (where the knowledge conveyed is shoal biased to ensure a suitable depth of water safety margin) the data could also be used for the following:

    - Full 3D seafloor modeling for engineering purposes (e.g. underwater construction; pipeline/cable laying; dumping and dredging operations) for which shoal bias is inappropriate.
    - Seafloor type definition from backscatter data for sedimentary studies; engineering planning;
    - Sound velocity data for oceanographic studies;
    - Wreck and obstruction data in more detail than portrayed on charts (e.g. for historical studies, defence applications, recreational diving interest);
    - Geodetic and tidal information for datum studies.

5. In addition to the bathymetric survey case described, hydrographic offices have a wealth of other data including navigational marks, traffic schemes, boundaries and limits. The ideal of exploiting as much of the data as possible, for numerous applications, is depicted in figure 3.

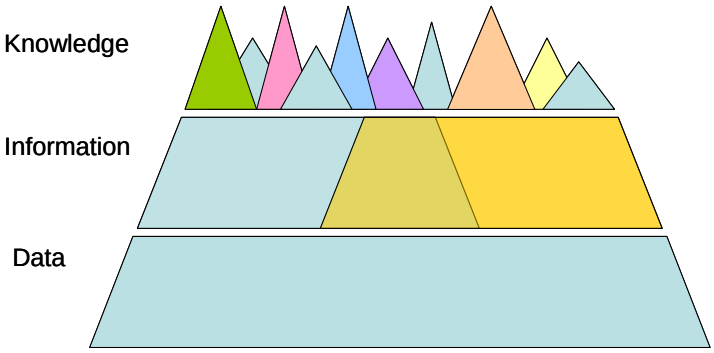
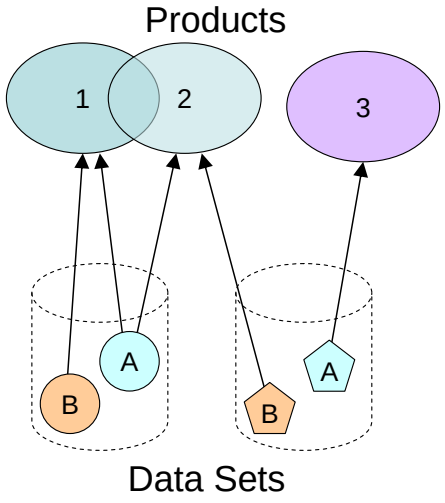


Figure 3

Data duplication and data conflict

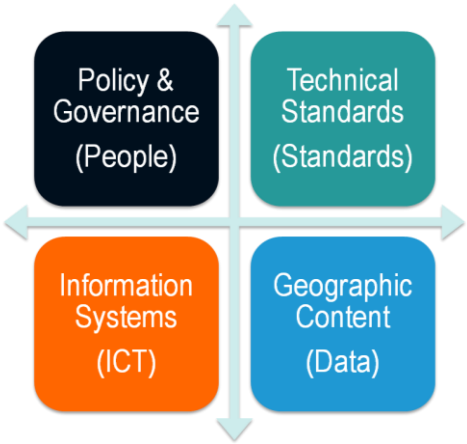
6. An organization focused on the delivery of discrete products (such as paper charts, ENCs and nautical publications) may have separate data holdings that contribute to different product lines. This can result in the same data being held more than once (e.g. light information shown on charts being stored in a separate system to the light information in a List of Lights). Figure 4 demonstrates the same feature (A) being held in different formats in different data sets, used in different products. Product 2 uses features from two different datasets. This is not only inefficient in terms of the volume of data held, but can also lead to differences between the data held for the same feature. This has



become particularly evident where some ENCs and paper charts have discrepancies between them. This puts into question the value of the knowledge portrayed, as the conflict demonstrates doubt in what is correct. Such fragmentation of data, together with proprietary or product specific formats limits interoperability.

Interoperability principles

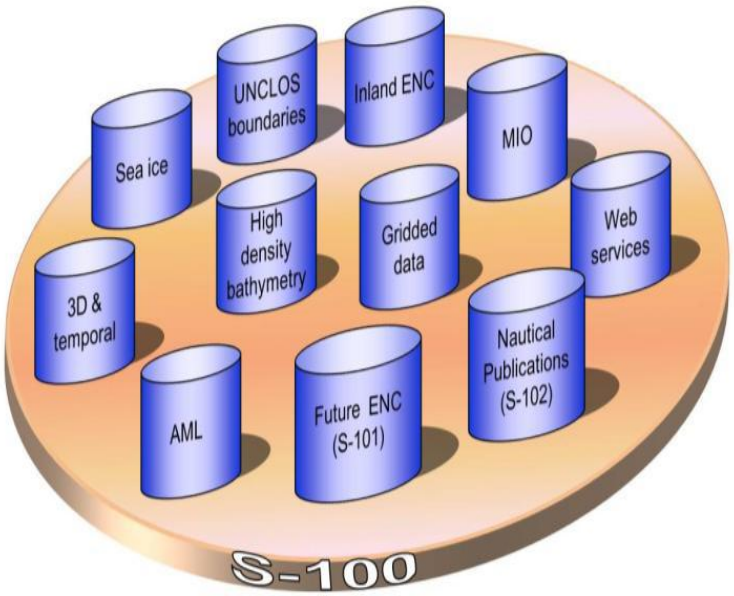
7. The concept of Marine Spatial Data Infrastructures (MSDI) is gaining wider appreciation. MSDI is the combination of a variety of data types, for efficient analysis by a wide range of disciplines, such as maritime spatial planning, environmental management and emergency response. This requires the data to be held in a generic way, rather than for a particular product for a limited user group or for a specific purpose.



8. MSDI require data, technology, standards, and policies to work together for the common good. There is a tendency to focus on ICT and data, without sufficient focus on standards and people issues. It is also a common mistake to think that existing products fulfill data requirements. But for effective exploitation of the true value of the data, it cannot be restricted by the intentions of the product compiler or editor. MSDI are not collections of hydrographic products, but infrastructures that promote interoperability.

9. Many countries have already established National Spatial Data Infrastructures, or have initiatives to do so. The marine element is often less well developed and there is a tendency to start building SDI on spatial products rather than spatial data. This is a good start, but there is considerable scope for improvement.

10. The development of the IHO Universal Hydrographic Data Model (S-100) is a strong enabler of enhanced data sharing across multi-disciplinary groups. A declared intention of S-100 is that it will make the use of hydrographic data easier. It will extend the use of hydrographic data beyond the focus on navigation, going beyond ENCs to imagery, dynamic data, and high-density bathymetry. Being based on ISO19100, it has international and multidisciplinary recognition, such as with the Open Geospatial Consortium (OGC), which has the declared intention of “geo-enabling the web”.



11. The IMO e-navigation initiative also has a vision beyond current navigational products. The Strategy Implementation Plan (SIP) states that *'as shipping moves into the digital world, e-navigation is expected to provide digital information and infrastructure for the benefit of maritime safety, security and protection of the environment, reducing administrative burden and increasing the efficiency of maritime trade and transport.*' E-navigation relies on S-100 as an enabler, but also on data not currently held by HOs. Thus in sympathy with MSDI, e-navigation requires interoperability of data.

12. E-maritime is defined by the European Commission as something bigger than e-navigation. It '*aims to foster the use of advanced information technologies for working and doing business in the maritime transport sector',* and *'envisages promoting interoperability in its broader sense'.*

13. Projects such as EfficienSeas, Mona Lisa, and AccSeas demonstrate the potential of better ways of information management. They have attracted EU funding because the principles above (MSDI, e-navigation and e-maritime) make good economic sense for the good of the wider community. These projects demonstrate the value and potential of more accessible and interoperable data, but what impact will they have on the way HOs work outside of the project environment?

External Drivers

14. Within recent years there have been major incidents necessitating the coordination of spatial data, such as:

    - BP oil spill in Gulf of Mexico in Sep 2010
    - Japan earthquake and tsunami in March 2011
    - Korean ferry disaster in April 2014

15. Reaction to each of these, and numerous other events, requires a multi-disciplinary approach including emergency response, environmental protection and longer term regional planning. Although precompiled products are currently essential, the knowledge they convey is often trapped in a form (whether paper or digital) that is not easily compatible with the tools and systems used by non-marine agencies.

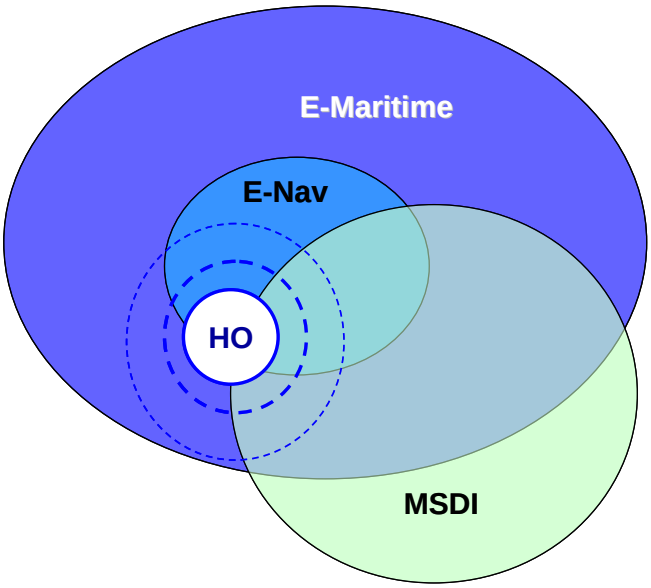
16. Climate change has the potential to raise concern for the coastal zones both in terms of sea level rising and increasing occurrence of extreme weather. A growth in the use of cross-polar routes may put environmental pressure on new sea routes as well as increasing challenges for disaster response. These will all need interoperable spatial data.

17. Population growth and the drive for additional energy generation, food production and other resource exploitation, increasingly puts pressure on both the marine environment and seaborne trade. This in turn puts pressure on HOs to provide suitable support to marine spatial planning.

18. All of the above demonstrate the need for better utilization of marine data, such that more informed decision making can lead to effective solutions.

Challenges

19. Given the scope of MSDI, e-navigation and e-maritime, and no doubt other initiatives, HOs need to consider the extent of their domain and influence, and how this might need to change to address future expectation. At present most HOs work in a relatively restricted domain, mostly due to their government status, tightly defined responsibilities and funding arrangements. This limits their opportunities to reach their full potential as data custodians rather than product producers. Authorities who define the role of HOs need to be challenged to encourage them to support the wider potential of hydrographic data.



20. To make data accessible to users outside of the HO community, it needs to be held in a universally recognized format. S-57 is an established format, but limited to ENCs. S-100 provides the universal data model for holding a wide range of data in a widely recognized format. It is understandable that S-100 data is not yet widely available, as the standard is not yet mature, but HOs need to consider whether they will simply export their existing data into appropriate S-100 specifications, or convert their internal data holdings according to the principles of the S-100 model.

21. Moving to an internal data holding based on S-100, will also provide the opportunity to remove duplication and ambiguity. The aim must be to only hold each feature once such that it is known to be authoritative. If scale dependent portrayal is required, this should be an attribute of the feature and not an excuse to hold the feature more than once.

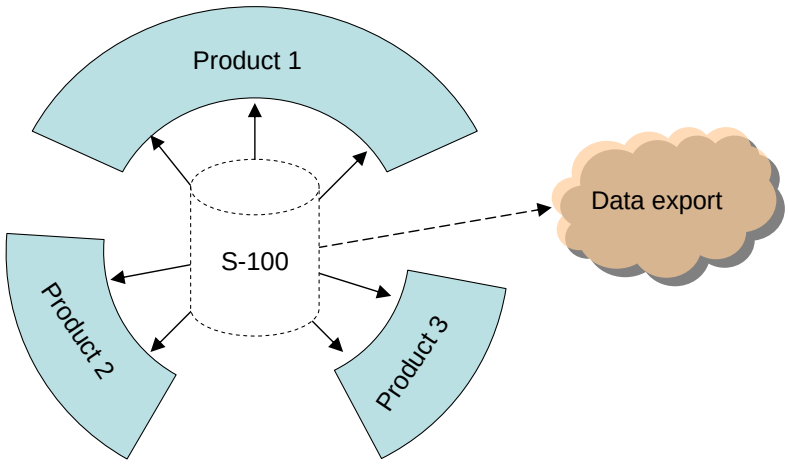
Conclusions

22. Much of what is written in this paper could be considered to be an academic view of a future that is not of immediate concern. But interest in S-100 is growing, as demonstrated by the soon to be published IALA S-201 standard for the exchange of Aids to Navigation data. The delivery of the IMO Strategy Implementation Plan to NCSR1 in July this year may not instantly change the world, but it does give encouragement to various organizations to strive for more integrated data and systems.

23. The potential for HOs to contribute to national and regional spatial data infrastructures is becoming more realistic. This requires serious consideration in terms of the consequences to how data is managed.

24. It is therefore important to:

- Use a data centric approach, holding unique features such that they are stored once but used many times;
- Use S-100 internally;
- Consider and promote wider use of data.



25. These matters could be advanced through sharing of best practice, promotion of case studies, and given a higher level of importance, through engagement with local and national government to seek support for maintaining MSDI as an enabler for better environmental management, faster response to disasters, and promotion of more efficient shipping.

Proposals

26. NSHC is invited to:

- Note this report
- Encourage a more data centric approach
- Engage in discussions about how HOs can contribute to e-navigation, MSDI and other geospatial initiatives
- Request HOs to offer support to the MSDI Working Group, through promotion of best practice and sharing of issues and concerns
- Consider relevance to the North Sea Region
- Take any other actions as it considers appropriate.