Linked Wrecks Data Pilot

Project Report

Submitted to

MEDIN

by

OceanWise Ltd

July 2015

Main Office:
Dovedale House
16 Butts Road
Alton
Hampshire GU34 1NB
United Kingdom
T: +44 (0)1420 768262
E: info@oceanwise.eu
W: www.oceanwise.eu

Revision and Approval for Use

Revision Number	Revision Date	Reason for Revision	Author	Checked	Approved
0.1	10/07/2015	Draft Report	RF		
0.2	20/07/2015	Included Use Case	ES	RF	
1.0	24/07/2015	Finalise Draft	RF	RF	

Distribution

This document has been distributed to the persons named below and no wider distribution is permitted without the written approval of the document owner.

Name	Organisation	Title
Ed Salter	Marine Space	
Clare Postlethwaite	BODC / MEDIN	
Mike Osborne	OceanWise	
Richard Farren	OceanWise	

Copyright

Disclaimer

The information contained in this publication has been compiled from sources believed to be proven in practice and reliable but no warranty, expressed or implied, is given that the information is complete or accurate nor that it is fit for any particular purpose. All such warranties are expressly disclaimed and excluded and users are therefore recommended to seek professional advice before committing to investments that incorporate any materials described in this document.

Contents

Glossary

AIS	Automatic Identification System
APN	Access Point Name – Gateway between a GPRS mobile network and another computer network e.g. the public Internet
EA	Environment Agency
ETL	Extract, Transform and Load
FTP	File Transport Protocol
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
HTTP	Hypertext Transport Protocol
LAN	Local Area Network
MCSE	Microsoft Certified Systems Engineer
MS	Microsoft
NOC	National Oceanography Centre
ODBC	Open Database Connectivity
PABX	Private Automatic Branch Exchange
PLA	Port of London Authority
PPU	Portable Pilot Unit
PSTN	Public Service Telephone Network
RDBMS	Relational Database Management System e.g. MS SQL Server
SAR	Search and Rescue
SCADA	System Control and Data Acquisition
SD	Secure Digital [Card]
SEPA	Scottish Environment Protection Agency
SIS	[Cadcorp] Spatial Information System
SNMP	Simple Network Management Protocol
SQL	Simple Query Language
UHF	Ultra High Frequency (0.3 to 3 GHz)
UPS	Uninterruptable Power Supply
VHF	Very High Frequency (30 to 300 MHz)
VPN	Virtual Private Network

1 Objectives

This study aims to assess and highlight issues that arise from integrating multiple valid sources of spatial data and serving them together as a consistent dataset. Additionally it seeks to provide recommendations/solutions where possible so that improvements can be made going forward.

This pilot study will comprise of a specific set of inputs (the UKHO wrecks database and English Heritage's NRHE dataset from the AMIE database) limited to a defined spatial location and tested against a defined use case scenario. The full objectives and deliverables of the pilot project are:

- Ingest UKHO wrecks dataset and store the data in a relational database where it can be updated efficiently and in an open format and where it can be accessed simultaneously by different systems.
- Publish the data to a WMS in a way that access can be limited to a specific data provider, a specific area or the whole dataset. Access will include the ability to query and display selected attributes.
- Extend the above capabilities to consume English Heritage (EH) historic wrecks and archaeological features dataset (in pilot form initially) in the following three methods:
    - By taking a copy of the EH data, processing it alongside and then serving it with the other data sources
    - By accessing the EH wrecks data remotely through either a Web Feature Service (WFS) or a web database connection.
    - By accessing the EH wrecks data as a published WMS only.

  The first of these methods will allow more flexibility to assess the integration potential between the UKHO and NRHE datasets.
- Integrate the EH wrecks data with the UKHO wrecks database to determine how well the two can be merged/joined. A WMS of the individual datasets as well as the combined dataset must be made available. Note: this will only be possible with option 1 or 2 above and not option 3.
- Investigate ways to publish all resultant web services through discoverable and Linked Data methods.
- Document user feedback for each of the above and assess the advantages and disadvantages of the above methods.
- Apply the pre-defined use-case test scenarios to the resultant web services and report the findings along with any recommendations. The use case is included later in this document.

2 Methodology

The items below are designed to standardise the workflow of the project and enable more informed decisions to be made from any outcomes.

- OceanWise to store source data in a relational database maintaining data structure/content as delivered from the source provider. Depending on the format in which the data are received, it may be necessary to source/write functionality to load the data. Each source record must contain a unique identifier which will allow updates to be applied appropriately and efficiently.
- OceanWise to produce a single dataset from the source data that must have a consistent structure, spatial reference system and terminology. This will require the source data structure to be mapped to a single structure whereby all source information is maintained. The purpose of this is to allow a report to be produced highlighting how achievable this process is and to make aware any issues that are discovered.
- OceanWise to identify and remove duplicates from the resultant single dataset. Duplicates to be assessed across source datasets only (not within source datasets) and by utilising the geographic position, wreck name and date information. Where the single resultant dataset contains duplicates due to supply from multiple sources, these may be removed by two methods:
    - Each source provider must provide information on the valid geographic extent of their data. Data from the source provider outside of this area that exists within another source providers designated extent will be removed from the first source provider (in the single consistent dataset output only).
    - Where duplicates exist within the valid extent of both source providers or outside the valid extent of both source providers, the duplicate which originates from the least reliable source or with the least supporting information will be removed.

  It should be noted that only the single combined dataset will undergo duplicate removal – source data tables will not be affected. Care must be taken when removing duplicates to ensure all relevant attribute data is still accessible to users. Important as one source may record precise location details, whereas another dataset eg NHRE may contain additional historical information of particular interest to users. This means that during duplicate removal – the attribution from both feature records must be assessed and merged where appropriate so that all useful information is maintained on the surviving feature record.

  It must be noted that the idea to produce a single consistent dataset from these two sources is not a new one. Duplicate assessment and identification work was undertaken between NRHE data and the UKHO in the "Enhancing the National Monuments Record", completed under the ALSF by the Maritime Archaeology Trust and SeaZone: http://archaeologydataservice.ac.uk/archives/view/marinenmr_eh_2011/ & http://www.maritimearchaeologytrust.org/nrhe

- OceanWise to publish a WMS of the wrecks dataset which shall be served from a dedicated cloud server. This will be supported on a subscription basis. The WMS endpoint will provide access to the single consistent dataset or individual provider datasets with the harmonised data structure, therefore allowing users to assess and compare the use of all datasets in their needs.
- OceanWise to ingest EH (NHRE) data to a separate location in the relational database and assess how well the data can be joined with the UKHO wrecks dataset. It is understood that there should be a matching unique identifier to join the same wreck feature in both datasets (UKHO WreckID). Additionally OceanWise will assess the attribute structure and content of the EH wrecks dataset and attempt to map the data to the "single consistent dataset" explained above. The original "unmapped" dataset will also remain available within the WMS service so that users can access and compare the original and mapped data in use cases.
- Implement different ways of holding the above datasets including as 'linked data triple stores' to assess what additional functionality or other advantages may result. This activity will address the need to synchronise data between the triple store and the underlying non-RDF database, potentially using a form of D2R connector.
- Develop methods to address situations where a data record exists in one dataset but not the other dataset and return an appropriate response to the end user.
- OceanWise will setup the EH wrecks dataset on a secondary web server and test the functionality available by exposing a secure remote database connection, a secure WFS and a secure WMS of the EH wrecks dataset. These resources will be consumed by the cloud server and an assessment made of whether the dataset processing required to produce a single consistent dataset and a merged EH/UKHO wrecks dataset is still possible with each resource.
- OceanWise will create Linked Data and publish this online at http://data.oceanwise.eu to aid discoverability and data provenance, and to demonstrate good data product practice.
- Parallel to the above, Marine Space to document a use case for the above service, identifying the information and functionality required for a typical screening or EIA. Each of the options above will then be tested against this 'user requirement specification' and advantages and disadvantages assessed and reported on. Functionality that is not possible using existing technologies will be set aside for future reference. As above – I'm envisaging a theoretical case study within inshore waters to maximise overlap of data and likely numbers of wrecks and obstructions in each dataset.
- Generate service metadata for the above pilot and submit to the MEDIN portal.
- Prepare a document reporting on the 'Linked Wrecks Pilot' with clear recommendations on how such a service could be implemented in full either by the primary data custodians (EH and UKHO) or one or more third parties having access to the data. Note the report and recommendations will be applicable to the Devolved Administrations (Scotland, Wales and NI).
- Host the above service for at least 12 months (to Mar 2016).

3 Use Case Scenario

Introduction

In order to test the functionality of the web application interface and WMS being developed by OceanWise, MarineSpace Ltd were commissioned to prepare and test a use case scenario. The chosen scenario comprises the production of an archaeological desk based assessment (DBA) for a theoretical EIA level marine development located in the English Channel, within the 12 nautical mile limit (the Historic England area of responsibility under the National Heritage Act 2002).

This scenario has been specifically chosen to ensure the maximum availability of data from the UKHO and Historic England wreck datasets, and in order to fully test and document the system functionality for specialist end users. The waters off the south coast of England have a long history of maritime and military activity, and have been a major focus of activity since at least the medieval period, with a number of shipping channels and historic ports located along the coast. The area has a significant number of known wrecks and documented shipping losses, and high potential for the recovery of remains associated with all periods of maritime history.

The south coast region also has potential for the recovery of crashed aircraft associated with military conflict during World War I and World War II. The skies above the channel saw significant airborne conflict, including organised and pre-planned attacks on military convoys and strategic locations, airborne defence and numerous other battles and skirmishes. There are a number of important routes to strategic locations and targets located all along the south coast, including Southampton, Portsmouth, Hastings and Dover.

The overarching aim of a standard archaeological DBA for an EIA level project is to document the historic environment resources that may be affected by the proposed project, and to assess the likelihood and significance of any impacts. Therefore, to inform the work, searches of records held in the UKHO wrecks database, and the maritime records of the National Record of the Historic Environment (NRHE) are usually conducted.

To achieve this, at present, wrecks and obstructions data must be acquired from the UKHO database under licence, for which a fee is required. Access to the data held by the NRHE is currently free; users must send in a shapefile of an 'Area Of Interest' (AOI) to a member of the heritage data management team at the NRHE, specifying what type of data records they wish to be extracted eg. maritime records, polygons, find spots etc. Data are usually provided by Historic England as area shapefiles and associated data records in excel, word or pdf format.

UKHO data records include details on wrecks and seabed obstructions, and are primarily collated to ensure the safety of navigation at sea. As such, information on the size, position and nature of features on the seabed is the primary focus, although in a number of cases specific historical detail may be provided to establish the identity and nature of loss of a wreck or obstruction.

In contrast, data from the NRHE have been collected with the sole aim of providing information on the marine historic environment and archaeological interest of sites and features at sea. As such NRHE records may be quite different to UKHO records in terms of the detail they contain. NRHE data constitute records of Known Wrecks where a specific wreck location is known, Recorded Losses linked to casualty records (reports of ships or aircraft seen in distress or lost at sea rather than specific sites on the seabed), isolated Find Spots and Named Locations, that is records where only approximate or no location data exist. Many of these records are broadly indicative of areas of maritime archaeological potential rather than specific records of wrecks on the seabed. The number of sites in the NRHE is far smaller than the number recorded by the UKHO, and records are often also duplicated.

Data Uses and Analysis

The primary use of wrecks and obstructions datasets in the production of an archaeological DBA is to identify sites and monuments within the application area, and assess their presence, stability and significance. The key aim is to establish the character and importance of the historic environment assets within the development area, and therefore to determine significance of any impacts and enable implementation of appropriate mitigation and future management where required.

Data are used to develop a picture of the historic character of an area, both in terms of any wrecks and obstructions that are present within the development area, but also to determine the potential for unexpected discoveries or undocumented sites to be discovered. Such records will generally form part of a wider study into the historic environment resources within a development area utilising documentary research, technical assessment of geophysical and geotechnical survey data (for example sidescan sonar, multibeam bathymetry, boreholes, vibrocores, grab samples etc.), and other supporting studies and data.

Wrecks and obstructions data are specifically used to corroborate, compare and contrast with sites identified in geophysical survey data, either clearly identifiable shipwrecks with associated wreck records, or anomalies of uncertain archaeological interest and potential that may represent previously unknown sites, or features of low archaeological interest.

Amongst the types of outputs that will be produced as part of the DBA are: text based descriptions of wrecks including detail on their key characteristics, form and historic significance; figures including spatial depictions of wreck records or specific assets identified in records or geophysical data; tables; recommendations for exclusion zones or other appropriate mitigation, where required. In addition, gazetteers of all the recorded maritime and prehistoric archaeological sites, features and materials will be produced. The gazetteers include any sites and features identified in the geophysical assessment, the review of wreck records and any other documentary records.

Typical Data Requirements

In order to complete a standard archaeological desk based assessment, wrecks and obstructions data containing specific information and data attribute fields are usually required. Commonly, spatial GIS layers and shapefiles are extracted from searches of the UKHO and wreck databases. These data are then interrogated in a project GIS in order to characterise the nature and significance of any wrecks and obstructions that may be present within the proposed development area.

Some of the typical questions relating to wrecks and obstructions data that might be posed by archaeologists engaged in desk based research for an EIA level marine development project are:

- What wrecks and obstructions are present in the development area and appropriate buffer zone?
- Can these data be used to corroborate features or anomalies identified in geophysical survey data?
- Have any previously unknown sites been identified in these survey data?
- What is the historic significance of any identified wrecks, obstructions or features based on criteria such as age, construction type, manner of loss, cargo, condition, etc?
- Do the data provide further information to indicate the potential of the development for the presence of previously undiscovered archaeological sites, features and artefacts?

With the above in mind, the key wrecks and obstructions data attributes required to complete a fit for purpose marine archaeological DBA are considered to be as follows:

- Wreck/obstruction Name;
- Location – Lat/Long, WGS84 UTM 30/31N;
- Vessel type;
- Nationality;
- Date/period of construction and loss;
- Manner of Loss;
- Propulsion;
- Detailed descriptions;
- Depth;
- Dimensions and size, l x w x h;
- Unique identifier(s) – accounting for discrepancies or duplications between datasets;
- Status (LIVE/DEAD – UKHO records);
- Associated/Related Monuments;
- Data source (eg. UKHO, NRHE, s57, BMAPA protocol etc.); and,
- Related events or date of last data update – ie date of most recent survey, or record creation/modification date.

Data Functionality

To facilitate effective assessment, analysis and the production of high quality outputs for the DBA, the following data functionality is usually required as standard:

- Ability to call up and view datasets simultaneously to facilitate comparison and analysis;
- Clear identification and linkages between duplicate records from different datasets for eg. through matching unique identifiers or clear display in attribute data of original data source;
- Visible monument points/extents, with ability to view data at a variety of scales

- Ability to call up key attribute data in a summary window with additional capability to select, query and display select attributes, including data source;
- Ability to select and manipulate data in such a way that it can be limited to a specific data provider, a specific geographical area, by date, by period, or as a whole dataset;
- Compatibility with other shapefiles and data for comparative analysis;
- Ability to select and export data in order to create figures, images and tables. Key for production of comprehensive and fit for purpose DBA to include spatial depictions of data, creating gazetteers in DBA reports, tables of known wrecks etc.
- Functionality to provide regular updates of datasets to ensure the most recent and up to date records and data are available;
- Measuring and selection tools to facilitate analysis; and
- Base mapping and clearly identified Coordinate Reference System (CRS).

Use Case Methodology

With the above information and data requirements in mind, the use case was tested against the WMS and web application interface developed by OceanWise in order to provide feedback and comment on the service and its potential future application. The functionality of the system was tested against the use case requirements and functionality identified above and the merits, advantages and disadvantages of using the system are reported below.

The web application interface was tested initially in order to provide commentary and feedback on the design, functionality, data linkages and system tools designed by OceanWise. A WMS and WFS feed were also supplied by OceanWise and available data layers comprising NRHE wrecks, MTF (UKHO) wrecks, Bathymetry Digital Elevation Model (EMODnet), contour data, and traffic separation were loaded into QGIS. These layers were tested both individually and via the use case study to further assess functionality and in order to provide commentary on future application and delivery of wrecks datasets.

It is anticipated that the reporting and feedback on these service elements provided by MarineSpace will be directly used to provide further recommendations as to how such a service could be implemented and delivered by either the primary data custodians (HE and UKHO), or by one or more third parties.

Key Findings - Web Map Interface

The first impression of the web map interface is positive; the design of the system is simple and striking, and the use of the interface is smooth. The controls are straightforward and intuitive, and it took very little time to become familiar with operating the system. Panning around the map interface is provided by clicking and holding the left mouse button, and zooming in and out by using the plus and minus buttons in the top left of the screen.

The measuring tool is simple and effective and can be used to take a variety of useful measurements including distances between wrecks, from the coast, or from specific geographic or man-made features – Eg. navigation channels, buoys etc. An area selection tool, or alternative functionality to allow users to select multiple records at the same time would have been a useful addition for the purpose of the use case study. This may create an issue in terms of how to display multiple returns in the 'Item properties' window, however a collapse/drop down option could perhaps provide a simple solution?

The Basemap layers provided are appropriate, and provide important context for assessing specific wrecks and obstructions as part of the use case study. Whilst the layers give the website a good look and feel, if all layers are switched on at the same time the system can look 'busy'. However, once zoomed in to an appropriate scale then the display becomes quite user friendly. The addition of the EMODnet Digital Elevation Model (DEM) bathymetry data and contour layers was found to be particularly useful and the addition of a toggle option for the DEM opacity could be used to make the display more visually appealing. For the purpose of the use case study, the traffic separation layer was not utilised, and as such this layer may be unnecessary in any future versions of the interface.

The available data layers are provided on a 'Layer Controls' bar which can be collapsed or hidden on the left hand side of the screen. Layers could be displayed simultaneously by switching on or off with a simple click on the relevant bar, which was particularly useful for conducting comparative analysis of wrecks records. In addition, the two Wrecks layers (NRHE wrecks and MTF wrecks) are selected using a check box feature, which further allows specific features on the map to be identified by simply clicking on the feature of interest.

On the wrecks layers, as the symbols used for both MTF wrecks and NRHE are broadly similar, just different in colour, it was at times hard to discriminate between features unless zoomed in tightly. This could be mitigated by changing the symbology for one of the layers, but was not of sufficient concern to cause a major problem to the analysis. In addition, for both layers feature name labels and unique identifiers appeared to display randomly on screen, often appearing differently dependent on scale. At times only a few labels were visible at wider scale, but greater numbers appear when the map is zoomed in. To mitigate this display issue, an option for turning labels on or off on the display should be considered in any future versions of the interface.

Once a specific feature is selected, all information and attribute data for that wreck/obstruction is displayed in the right hand side of the screen in an 'Item Properties' bar. The attribute data provided are condensed from the full attribute data provided in the raw UKHO and NRHE datasets. For the purpose of the use case the attributes provided were found to be broadly sufficient, with the majority of the required information as set out in the typical data requirements identified above, being supplied.

The description/information fields were found to be of particular use, often containing the majority of the most important information of particular interest for specialist users. Whilst this field can often appear verbose, this is a result of the data records from which they are created rather than the web interface, and attempting to shorten/summarise these descriptions would be a major task and is not advised. This may also have a further negative impact of removing important information valued by specific user groups. Where descriptions were lengthy a scroll option has been provided.

The 'Linked Wreck' and 'Related Wreck' fields are extremely useful and unique additions, likely to be of interest for both general and specialist users. Linked wrecks provides information on duplicate entries in different datasets (ie NRHE and UKHO) which is particularly useful for the use case study and understanding where wrecks may not be physically present on the seabed for a particular development. Related wrecks show records that are associated with another based on NRHE monument descriptions. Associations may be on the basis of a variety of factors including technical details, place of constructions, ships from the same fleet, or those lost in similar circumstances, e.g. wrecks lost at sea during collisions, battles or storm events.

The linked wrecks layer is presently a work in progress, with a relatively low number of examples currently available compared with the number of duplicate entries that are known to exist across the two wreck datasets. However, these layers can be usefully interrogated, and provide important detail enabling users to understand duplicate entries or other associations between features. In any future versions of the interface it will be important to provide text descriptions or a user guide describing what these fields actually show users. The chosen method of displaying linkages between features, using a line to link associated wrecks may also not be feasible for wrecks located in very close proximity or at the same location. An alternative or additional method of display could provide a 'Linked wreck' field in the item properties bar, allowing users to click on a link which makes the map system 'snap' to the linked entry, or highlights both features simultaneously.

The main area where difficulty in using the web map interface was experienced was using the filter option and ECQL commands. Not having prior working knowledge of ECQL language and commands prior to use of this system, it took some time to firstly become familiar with ECQL and then to start running successful filter entries that returned useful results.

One issue that further hindered progress was that it was found filter results only displayed results correctly at certain scales. Often a correct command had been entered in the filter field, but the display failed to return the filtered results on the map display, instead removing all features, which gave the impression that the filter was either incorrect or had returned a null response. It is however suspected that this issue may have been a result of issues with web connection rather than an inherent issue with the web interface. Upon retesting all filters appeared to be returning results effectively at a variety of scales.

In the main, the issues therefore resulted as a lack of familiarity with the ECQL language. That said, once a level of familiarity with ECQL had been achieved, a number of spatial, temporal and other searches could be conducted. It is certainly seen as a positive and beneficial aspect that ECQL can enable users to conduct searches on a whole variety of attributes, rather than a specific set list.

Amongst the searches conducted were the kinds of questions that might be asked of data for the use case study;

- Cargo type e.g. COAL;
- Vessel Nationality;
- Propulsion e.g. STEA;M
- Depth e.g. > 50; and
- Vessel Length.

Despite this, certain filter search topics did not return successful results, such as vessel age, year of construction/loss, engine type. It is presently unclear if this is a result of not entering the correct ECQL command, or the way the data is structured not allowing this type of search to be undertaken on the datasets.

Based on the testing of the filter undertaken to date, ECQL has great potential to provide users with an incredible amount of flexibility whilst running fulters and searches of data. However, it requires a level of initial input to become familiar with ECQL language commands that may act as a barrier to future use. Will end users of this web interface, likely to be general public and researchers, rather than specialist users, have the patience required to do this?

The disadvantages of ECQL may be mitigated to some extent by providing an introduction to the ECQL language and a number of examples of possible filter commands in a user guide available in the website. However, it is may be considered more appropriate to find an alternative method to provide filter searches, given that the website is designed primarily for non-specialists.

Key Findings – WMS

A Web Map Service feed was supplied to MarineSpace in order to test how the layers developed by OceanWise and provided through this service could be loaded to, and utilised in a GIS system. This approach is more akin to the current method undertaken by specialist users engaged in analysing wrecks data to be included in a typical archaeological DBA report. For the purpose of this test, Quantum GIS (QGIS) was utilised, and therefore all comments should be considered as specific to the loading and operating of WMS and associated layers in this system. The findings will have general relevance to use of WMS in other GIS.

QGIS provides an 'add WMS layer' tool, where the WMS URL supplied by OceanWise could be copied and pasted into the system. This establishes a connection with the WMS service and access to a list of available data layers. The connection can be saved and stored, making re-accessing simple.

The following layers were accessed and loaded into QGIS:

- World Basemap;
- EMODNET contours;
- MTF wrecks; and
- NRHE wrecks.

The main advantage of loading layers into a GIS system as opposed to using the web map interface is that it provides users with the ability to compare and contrast with other datasets within a system with which they are already familiar. For example, in the use case, a shapefile of the proposed development area could be loaded into the GIS allowing users to visually 'filter' results to those lying within the relevant area of interest. Geophysical survey data can be further added, and viewed in combination with the wrecks layers. For the purpose of a marine development project this is key in understanding what wrecks and obstructions remain present in an area, and whether any new sites or anomalies of potential archaeological interest not identified in the wrecks data have been identified, and require further assessment. The system is therefore an excellent option for providing an enhanced display and facilitating in-depth comparative analysis of multiple datasets.

The major drawback of WMS is that as it essentially produces an image (in jpeg or png format) of the dataset in a particular area, and as such selection and export of multiple features within a given area is not feasible. This is exactly the type of functionality that is key to the production of high quality outputs for a DBA, for example creating dbf tables or shapefiles and figures for inclusion with the final DBA report.

In addition, the nature of WMS layers means that other functionality, such as the ability to change the symbology, and switch layer labels on and off is not possible. In working through the use case study it was found that the MTF wrecks layer has been supplied with wreck names labelled; the NRHE wrecks layer displays names of wrecks where known. For the purpose of display this can make the screen look very busy, with layer labels often overlapping and conflicting. Any future version of the WMS should remove these labels from the original source file to enhance the display of data, assuming that an 'on/off' functionality cannot be suppled.

The functionality and information within layers in terms of the ability to review key data and the types of information available was found to be much the same as in the web map interface. When individual features are selected, users are given the option of viewing an html display, a feature display, or a text display. The HTML display provides a user friendly display of the key 'condensed' attribute data. The 'feature' display provides the original source data, and the 'text' display provides the same data in a simple unedited text format. The options therefore all provide the same key information, and selection is purely a matter of user preference.

As with the interface the layers made available provide the majority of the required information likely to be necessary for completing a marine archaeological DBA. The Basemap layers provided were not the same as in the web map interface, and this is presumably as these layers would need to be supplied to individual users under licence or subscription. However, the World Basemap and contour data still provided useful context.

As with the web map interface, the symbols used for both wrecks layers (MTF wrecks and NRHE wrecks) are broadly similar, and even in QGIS it was still found at times hard to discriminate between features in the different datasets. As recommended above this could be mitigated by changing the symbology for one of the layers.

Key Findings – WFS

A Web Feature Service feed was supplied to MarineSpace; the feed supplies direct download functionality, and unlike WMS, this WFS servers support shapefiles and other formats. The main difference between WMS and WFS is that WFS sends a file or stream of features in that area and the GIS does the actual drawing of the data as a map (Richard Farren, pers comm).

In order to test the functionality of the WFS layers were loaded into QGIS. Much like with WMS layers, QGIS provides an 'add WFS layer' tool, where the WFS URL supplied by OceanWise could be copied and pasted into the system. This establishes a connection with the WFS service and access to the list of available data layers. Again, the WFS connection can be saved and stored, making re-accessing at another time simple.

The following layers were accessed and loaded into QGIS:

- World Basemap;
- EMODNET contours;
- MTF wrecks; and
- NRHE wrecks.

Once layers were loaded into QGIS, feature symbologies could be selected and altered depending on user preference. This mitigated the minor issue with display of the two wreck layers identified above and added the additional benefit for users of being able to more readily identify wrecks with statutory exclusion zones, or findspots.

A significant advantage is gained by accessing the data via WFS, as multiple records can be selected and viewed at the same time using the standard QGIS selection tools, or defined polygon/rectangle, shapefiles and attribute tables. These data can also be exported as shapefiles or dbf files, particularly relevant for the production of high quality outputs for an archaeological DBA as identified above. IT is also possible to edit records via WFS, although caution should be applied in doing so, as data or copyright may become invalidated, or inaccurate information from unidentified sources may be included in future outputs.

The WFS, provides the full attribute data much in the same way a shapefile added to GIS would. Therefore, whilst the attribute fields made available for the interface and WMS were seen as sufficient, the full range of data could be accessed via the WFS, which is certainly useful. The main drawback of this is that only the information 'tree' is therefore available in a summary window, and the user-friendly html option is no longer available.

Use Case Conclusions

The pilot project has proved immensely successful in terms of demonstrating the various ways in which wrecks datasets held by the UKHO and Historic England, or other statutory bodies, can be made available to users in the future. The use case study has allowed the various advantages and disadvantages of each system to be assessed and future enhancements, where appropriate have been suggested.

The web interface, whilst designed towards general users rather than for specialist marine development work, provides a user friendly portal containing a wealth of useful data and information. The functionality of the system is certainly appropriate for general users, the main target audience of the website. The system tools and display giving the site an excellent look and feel, whilst at the same time allowing wrecks datasets to be easily accessed and interrogated. The information layers and attribute fields are appropriate for specialist end users. The Linked Wreck and Related wreck fields, a major focus of the work by OceanWise were seen as particularly important additions, enabling users to more readily understand duplications and associations in records and across datasets

Issues were experienced in using ECQL filter searches, and this may be mitigated by using a different approach, or by providing a detailed user guide introducing the ECQL language and a number of search example commands. Such a guide is also seen as being key for any future iterations of the interface to describe the tools and data layers available – eg. What do the 'Linked Wreck' and 'Related Wreck' layers show users? The limited ability to export, extract and fully manipulate these data, and add other shapefiles and data for the purpose of comparison, limit the usefulness of the system to the specialist end user, but it should be noted that the web interface is not aimed at this user group.

The WMS provides a similar level of functionality and information to the web interface, but with the additional benefit of allowing data layers to be imported and loaded into GIS. Data can therefore be compared and analysed in combination with other shapefiles and datasets. The system is therefore more appropriate for specialist users and provides much of the information and functionality identified in the use case. The main drawback is that data cannot be fully manipulated, and multiple features cannot be selected, and exported. Therefore, for the purpose of a DBA level project, whilst incredibly useful, especially for comparative analysis, the approach still lacks certain functionality.

Based on the testing undertaken for the use case study, the system that therefore provided the closest level of functionality to the requirements identified in the use case study was WFS. Whilst the levels of information made available in the web interface and WMS were seen as sufficient for the purpose of the use study, only the WFS could allow users to fully manipulate, select and export data in the way that would be required by specialist end users engaged in EIA level project work.

WFS is clearly a very powerful option, and whilst not the original focus of this study, should be considered as part of the overall package that could be made available to users in the future, alongside the web interface for general users and WMS for certain specialist users.

4 Report

Source Data acquisition

For the pilot study two datasets were considered; The UKHO Wrecks and Obstructions dataset (WODB) and the National Record of the Historic environment dataset of wrecks (NRHE). These two datasets represent definitive wreck data within the UK (England for the NRHE dataset) and also cross reference each other with a linked wrecks ID so that duplicate wreck records can be identified. In both cases – a subset of the dataset within an area chosen for the use-case scenario was extracted. The area for the pilot study is confined within the following Lat/Long bounds:

W 2.73 – E 1.39, N 50.20 – N 51.13

This can be visualised approximately by the area with wreck data coverage in the image below:

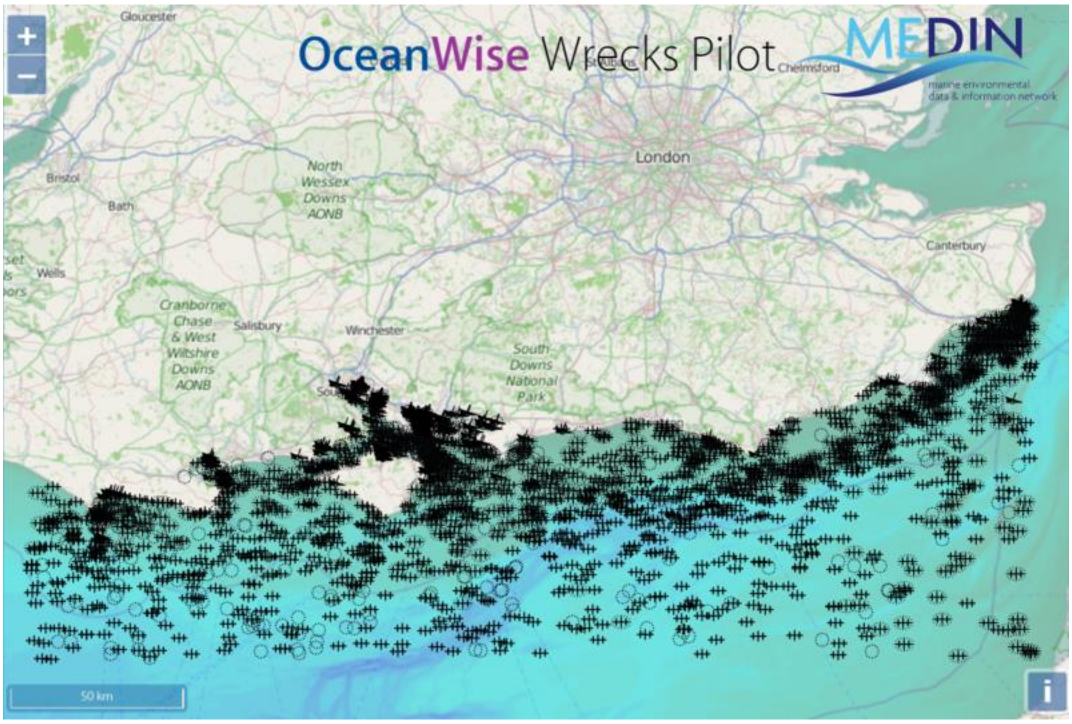


Figure 1: Wrecks Pilot area shown with Open Street Map and EMODNet bathymetry base layers.

The UKHO wrecks dataset is provided to OceanWise under the VAR agreement it has in place. The dataset is provided quarterly in a text based transfer format which OceanWise load to their database. The source dataset contains the following attribute fields in conjunction with a point geometry element:

WRECK_CATEGORY
STATUS
CLASSIFICATION
LATITUDE

LONGITUDE
HORIZONTAL_DATUM
LIMITS
POSITION_METHOD
DEPTH
HEIGHT
DEPTH_METHOD
DEPTH_QUALITY
DEPTH_ACCURACY
WATER_DEPTH
WATER_LEVEL_EFFECT
VERTICAL_DATUM
NAME
TYPE
FLAG
LENGTH
WIDTH
DRAUGHT
SONAR_LENGTH
SONAR_WIDTH
SHADOW_HEIGHT
ORIENTATION
TONNAGE
TONNAGE_TYPE
CARGO
CONSPIC_VISUAL
CONSPIC_RADAR
DATE_SUNK
NON_SUB_CONTACT
BOTTOM_TEXTURE
MAGNETIC_ANOMOLY
MAD_SIGNATURE
SONAR_SIGNAL_STRENGTH
MAGNETIC_INTENSITY
SCOUR_DIMENSIONS
DEBRIS_FIELD
ORIGINAL_SENSOR
LAST_SENSOR
MARKERS
CIRCUMSTANCES_OF_LOSS
SURVEYING_DETAILS
GENERAL_COMMENTS
LAST_AMMENDED_DATE
OBSTRN_CATEGORY
DATE_REPORTED
HO_WRECK_ID

The variability of information contain within these fields is extensive and deemed too onerous to represent fully within the pilot study. In addition to this “raw” data product, OceanWise produce a consistent “processed” model of the dataset which it uses for its Marine Themes data product. This contains a subset of attribute variability where essential information from a selection of the above attribute model is mapped into a more condensed model. The attribute model is as follows:

NAME
DESCRIPTION
CATEGORY
FEATURE_CODE
SOURCE_DATE
SOURCE_IDENTIFIER (SOURCE_INDICATION)
INFORMATION
CONDITION
STATUS
RESTRICTIONS
NATIONALITY
ACCURACY
MATERIAL
WATER_LEVEL
DEPTH
RECORD_DATE

This allows the data to be used more conveniently in its condensed form especially within web applications similar to how these data will be used. This dataset is known as the Marine Themes Features, Wrecks and Obstructions layer (MTF Wrecks for short). The decision to include the full dataset rather than the condensed dataset will be made based on the results of the use case study. Both versions of the dataset are referenced to WGS84 Latitude / Longitude as supplied and in product form.

The National Record of the Historic Environment (NRHE) dataset was supplied by English Heritage (now Historic England). The dataset is not yet published online via data.gov.uk but there is an online front end search tool which allows the dataset to be queried at http://www.pastscape.org. In this case a separate data request was made directly to English Heritage for records within the NRHE dataset where the category of monument is "Wreck" and only records within the defined project area.

English Heritage supplied the dataset to OceanWise in three files;

- A PDF document listing all records in a descriptive format including position information recorded in both Lat Long and OSGB British National Grid. This dataset was not used as part of the pilot study because it is not in a machine readable format.
- A set of ESRI shape files containing point line and polygon representations of records where appropriate and additionally a set of attributes:

  HOB_UID
  NAME
  DESCRIPTIO (Limited to 255 characters)
  MON_PRECIS
  CAPTURE_SC
  Easting
  Northing
  AREA_HA

  The attribute names are truncated because of shape file format specification which limits them to 10 characters. Additionally the content of the attributes are limited to 255 alphanumeric characters which means that any data within the DESCRIOPTIO attribute that exceeds this will be lost. This is why shape files should not be used in GIS systems where an attribute name is greater than 10 characters or attribute values are greater than 255 characters. A shape file should not have been used for this dataset. The dataset contains a geometry element referenced to OSGB British National grid and additionally attributes representing the grid coordinates as Eastings and Northings.
- An XML dataset containing a detailed xml element tree of record data. Due to the nature of the XML dataset, a ridged data model does not exist inherently. That is to say, each monument record within the xml document may have more than one example for each of the child elements. It is valid for there to be multiple wrecks finds, actors (compilers), evidence references, identifiers and temporal events associated with a single monument record. Whilst this allows significant detail and variability within the feature – it does cause difficulty in producing a normalized dataset with a rigid data structure for the purpose of web publishing. Attribution that can be stored within the xml tree are as follows (indents show child elements which may have multiple parent records);

  RECORD CREATION DATE
  RECORD CREATION NAME
  PRIMARY IDENTIFIER

OTHER IDENTIFIERS (INCLUDING IDENTIFIER TYPE)
DESCRIPTION (FULL)
DESCRIPTION (SUMMARY)
ACTORS (PEOPLE WHO HAVE ADDED DATA TO THE RECORD)
  ACTOR TYPE
  ACTOR NAME
  ACTOR ORGANISATION
  ACTOR ROLE
  ACTOR DATE (START & END DATE OF RECORD UPDATE)
  ACTOR ADDRESS
  ACTOR LOCATION (BNG GRID REFERENCE)
MONUMENT TYPE
MONUMENT MINIMUM DATE
MONUMENT MAXIMUM DATE
CRAFT (THE PYSICAL MONUMENT DETAILS)
  CRAFT TYPE
  CONSTRUCTION METHOD
  PROPULSION
  NATIONALITY
  CARGO
  DATE OF LOSS MINIMUM
  DATE OF LOSS MAXIMUM
  MANNOR OF LOSS
  DIMENSIONS (L x W x H)
RELATED MONUMENTS
  RELATED IDENTIFIER
  RELATED NAME
RELATED RESOURCE REFERENCES
  REFERENCE TITLE
  REFERENCE EXTENT
  REFERENCE DESCRIPTION

The xml dataset contains significant valuable information about the wreck monument records and therefore needs to be processed to be used within the resultant dataset product.

Data Processing

To produce a useable published dataset, work was needed to make the source datasets suitable for the web.

For data management ease it was a preference to utilise a relational database to store the data from both data sources. Maintaining the desire to use Open Source tools and open data formats where possible, the choice to use PostGRES and the excellent PostGIS spatial extensions was an obvious one. The other Open Source alternative "MySQL Community Edition" offers little or no spatial functionality so was not as well suited for this purpose.

The benefits of using a relational database for this project are:

- More efficient data management. Loading, updating and using data can be controlled and monitored more tightly than with open file structures.

- Utilisation of Server resources vs desktop resources. The database server can be used to process data during load and use, freeing up user desktop resources to handle result rendering only.
- Relational / Queryable functionality. Data sources can be linked by creating "views" which join data through common fields and can even create new virtual datasets as a result of these joins.

The data processing journey can be simplified by the following diagram below:

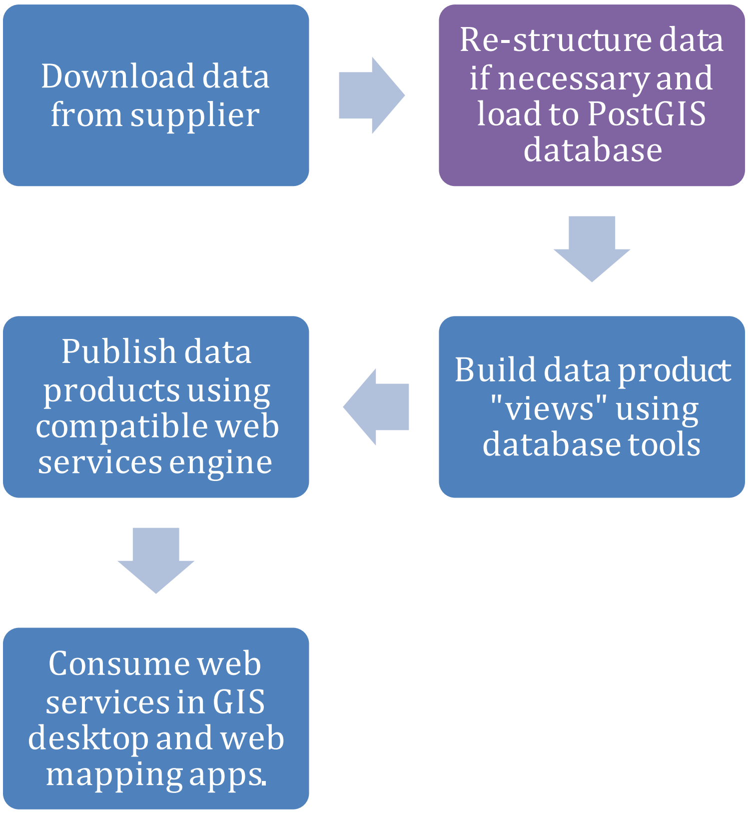


Figure 2 Diagram showing stages of the data processing journey for published web services

The second stage (highlighted in purple) indicates the topic under discussion at this stage of the report. In all cases the data are supplied in Open formats. The geo-spatial datasets (shape file and gml data) needed little configuration to load to the database but the xml file required normalizing before it could be loaded to the database.

The MTF Wrecks data was exported from OceanWise's Oracle database to gml format. The resultant file was then loaded to the PostGIS database on OceanWise's web server using GDAL / OGR tools (http://www.gdal.org/ogr2ogr.html) an Open Source data manipulation toolset. It is feasible that in future a direct database to database load is possible using OGR tools.

The NRHE shape files were also loaded to PostGIS in a similar way. OGR2OGR is able to read the shape format directly as the files were supplied to OceanWise and so these were loaded without modification.

The NRHE XML dataset required normalizing before it could be loaded to a database. The method to do so was as follows:

- Identify the XML elements which are useful to the published data product (listed above)
- Use a Python based XML parser to extract XML attribute and XML element values from the XML file.
- Join multiple ACTOR, CRAFT, RELATED MONUMENTS and RELATED RESOURCES values into single fields and produce an output in an Open Data format – csv.
- Load the csv to PostGRES as a non-spatial data table using the built-in PostGRES copy function.

After this stage the database consisted of four source tables

- MTF_Wrecks and Obstructions
- NRHE_points
- NRHE_lines
- NRHE_polygons
- NRHE_text (loaded from the processed XML file)

Data Product Layers

The next stage was to build useful product layers from these data layers. By employing a PostGRES/PostGIS database it was possible to do this using database views. The main tasks to be completed in this stage were to limit the content of the MTF_Wrecks layer to just the spatial extent of the pilot study and to merge the NRHE points, lines and polygon layers into a single layer that could then be joined to the NRHE_text layer. Essentially we wanted a working product layer for both datasets.

The view “MTF Wrecks and Obstructions MEDIN” was created using a spatial intersect query:

```
CREATE OR REPLACE VIEW "mtf_wodb_MEDIN" AS
SELECT a.the_geog,
    a.name,
    a.description,
    a.category,
    a.feature_code,
    a.source_date,
    a.source_indication,
    a.information,
    a.condition,
    a.status,
    a.restrictions,
    a.nationality,
    a.accuracy,
    a.material,
    a.water_level,
    a.depth,
    a.record_date
   FROM mtf_wodb a
  WHERE (a.feature_code = ANY (ARRAY[20020, 20021, 20022,
20023, 20024, 20025, 20010])) AND
st_intersects(a.the_geog, st_makeenvelope((-2.73)::double
```

```
precision, 50.2::double precision, 1.39::double
precision, 51.13::double precision, 4326)::geography);
```

Note that only a specific list of feature codes were selected too. This eliminates all features that are of type "Foul Ground" as these are not considered wreck objects.

The NRHE dataset firstly required the separate point line and polygon tables to be merged. This is due to a weakness in the shape file format where it can only store one geometry type per file. The database can support multiple geometry types in the proper way so a merged view was created as follows:

```
CREATE OR REPLACE VIEW nrhe AS
         SELECT nrhe_polygon.ogc_fid,
            nrhe_polygon.wkb_geometry AS the_geom,
            nrhe_polygon.hob_uid,
            nrhe_polygon.name,
            nrhe_polygon.descriptio,
            nrhe_polygon.mon_precis,
            nrhe_polygon.capture_sc,
            nrhe_polygon.easting,
            nrhe_polygon.northing,
            nrhe_polygon.area_ha
           FROM nrhe_polygon
UNION
         SELECT nrhe_line.ogc_fid,
            nrhe_line.wkb_geometry AS the_geom,
            nrhe_line.hob_uid,
            nrhe_line.name,
            nrhe_line.descriptio,
            nrhe_line.mon_precis,
            nrhe_line.capture_sc,
            nrhe_line.easting,
            nrhe_line.northing,
            NULL::double precision AS area_ha
           FROM nrhe_line
UNION
         SELECT nrhe_point.ogc_fid,
            nrhe_point.wkb_geometry AS the_geom,
            nrhe_point.hob_uid,
            nrhe_point.name,
            nrhe_point.descriptio,
            nrhe_point.mon_precis,
            nrhe_point.capture_sc,
            nrhe_point.easting,
            nrhe_point.northing,
            NULL::double precision AS area_ha
           FROM nrhe_point;
```

Finally the resultant view of all geometry items was joined onto the xml upload table to create a single complete dataset "nrhe_full":

```
CREATE OR REPLACE VIEW nrhe_full AS
 SELECT g.wkb_geometry AS the_geom,
    t.hob_uid,
    t.ukho_id,
```

```
          t.name,
          t.date_text,
          t.description,
          t.craft,
          t."monument type",
          t.construction,
          t.propulsion,
          t.nationality,
          t."manner of loss",
          t.cargo,
          t.length,
          t.breadth,
          t.height,
          t.displacement,
          t."related to"
         FROM (         SELECT nrhe.wkb_geometry,
                           nrhe.hob_uid
                          FROM nrhe) g
         JOIN nrhe_text t ON g.hob_uid = t.hob_uid;
```

The two primary datasets were then ready to be published as WMS layers.

Within the pilot study extent the total feature count for both layers was:

- NRHE – 992 features
- MTF Wrecks – 2433 features

The discrepancy is partly accounted for by the fact that the NRHE dataset coverage is limited to only the UK part of the English Channel whereas the UKHO wrecks data has full coverage of the bounding box.

Sourcing NRHE Data from WFS & WMS publications

One of the project objectives was to assess the ability to source the NRHE data not just directly from English Heritage via an email request but via pre-published layers such as a WFS or WMS (Web Feature Service or Web Map Service). For this purpose, the nrhe_full layer and the mtf_wodb_MEDIN layer were exposed as WFS endpoints so that this assessment could take place.

A WFS allows the user to request actual feature data (in GIS format) from the server rather than a rendered image of the data which a WMS provides. This means that theoretically the entire dataset can be "downloaded" using WFS commands.

The WFS endpoint was provided at:
http://geoserver1.oceanwise.eu/BASEMAP/nrhe_full/wfs
It is theoretically possible to send filter and join requests to the WFS, allowing it to behave like a database. However, to produce the types of derived layers that are required for this project, the WFS would need to be served alongside the other data layers as WFS also. As this is not possible, the WFS filter and query functions are not appropriate.

Instead, it was possible to download the entire Pilot study area via WFS from within a web browser by issuing the following HTTP GET request (url):
http://geoserver1.oceanwise.eu/BASEMAP/wfs?service=WFS&version=1.0.0&request=GetFeature&typeName=BASEMAP:nrhe_full&outputFormat=application/json

The resultant GeoJSON dataset for the full NRHE layer was produced and delivered to the browser in under two seconds. It would then be possible to load the dataset file to a database in the same way that the source NRHE data was provided. In addition to the GeoJSON format the dataset can be provided through a WFS in the following formats:

- GML 2.0 / 3.2 / 3.3
- CSV
- Shapefile
- Google KML

The result of this assessment was that it would be possible to consume the source data through WFS (read-only required) without significant difficulty, if it were possible for the source data provider to utilize this technology.

A practical assessment of the ability to ingest the NRHE dataset via WMS was not made. This was because the WMS standard will only provide a rendered image of a dataset and not a data stream of the features or a file of the features that can be consumed. The only possible support that a WMS could offer is through the getFeatureInfo request which responds with a data table of features at a selected point. Usually, however a map server will be restricted to returning between 1 and 10 features only for a getFeatureInfo request which would not be suitable when used to harvest a dataset containing around 2000 records. Based on this assessment an attempt was not made to consume a dataset as a source layer through a WMS.

Derived Data Layers

Along with the primary layers – it was possible to exploit the scripting ability of the database to build other views that provide useful information to the user. The two significant uses of this are creating links to show where a feature from one dataset can be identified as matching a feature in the other via a common attribute. We call this "Duplicate Wrecks" or "Linked Wrecks". The second use is in creating links to show where wreck features are associated (or related) to other wreck features. We call this "Related Wrecks".

It was possible to create Linked Wrecks because the NRHE dataset contains information on the UKHO wreck identifier in addition to the primary NRHE identifier. The work to create these links is still in progress so the links do not represent a fully matched dataset at this time. There are also other inhibitors to this process including "within dataset" discrepancies where multiple wreck features exist in different positions for one wreck (due to uncertainty). Barriers to the process are discussed later on in the report.

Using PostGIS database functionality, a spatial view was created using the point geometries of both matched wrecks to create a virtual line geometry (a physical line connecting the two items). The links can be used as if they were a separate layer but exist only as a virtual feature. The view was created as follows:

```
CREATE OR REPLACE VIEW wodb_nrhe_duplicates AS
 SELECT n.hob_uid,
    n.ukho_id,
    st_makeline(st_centroid(st_transform(n.the_geom,
4326)), geometry(u.the_geog)) AS the_geom,

round((st_length(geography(st_makeline(st_centroid(st_tra
```

```
nsform(n.the_geom, 4326)), geometry(u.the_geog)))) /
1000::double precision)::numeric, 3) AS length
   FROM nrhe_full n
   JOIN "mtf_wodb_MEDIN" u ON n.ukho_id =
NULLIF("substring"(u.source_indication::text,
"position"(u.source_indication::text, '-'::text) + 1,
12), ''::text)::integer;
```

Note that line and polygon features will be linked from their centroids (natural center points).

Again, PostGIS was able to create a virtual spatial dataset of Related Wrecks by querying the "related to" column which contains a comma separated list of identifiers to which the wreck item is related. PostGIS was able to create virtual links from the host feature to the related features in the same way as the Linked Wrecks. However, due to the possibility of there being more than one related wreck for any given wreck feature it was necessary to create a function script to return a normalized pivot table of related wreck Id's…

essentially going from this:

Wreck_ID	Related To
108723	17334
108724	197324, 56788, 2348

To this:

Wreck ID	Related To
108723	17334
108724	197324
108724	56788
108724	2348

The function:

```
CREATE OR REPLACE FUNCTION getRelatedWrecks() RETURNS SETOF
nrhe_full AS
$BODY$
DECLARE
    r nrhe_full%rowtype;
    start_id int;
    related_ids text[];
    id text;
BEGIN
    FOR r IN SELECT * FROM nrhe_full
    WHERE "related to" is not null
    LOOP
```

```
        -- split "related to" into parts
        related_ids := string_to_array(r."related to",',');
        FOREACH id in ARRAY related_ids
        LOOP
           r."related to" := CAST(nullif(id, '') AS integer);
           RETURN NEXT r;
        END LOOP;
    END LOOP;
    RETURN;
END
$BODY$
LANGUAGE 'plpgsql' IMMUTABLE;
```

The resultant function was designed to then be passed directly into the spatial view to create the spatial links as a virtual layer. Slightly different from before, this view required the NRHE dataset with geometries to be joined twice to the related wrecks table, once for the source identifier and once for the related identifier:

```
CREATE OR REPLACE VIEW nrhe_related AS
 SELECT related.source_id,
    related.related_id,
    st_makeline(st_centroid(st_transform(related.source_geom,
4326)), st_centroid(st_transform(related.related_geom, 4326)))
AS the_geom
   FROM ( SELECT n1.hob_uid AS source_id,
            n1.the_geom AS source_geom,
            n2.hob_uid AS related_id,
            n2.the_geom AS related_geom
           FROM nrhe_full n1
      JOIN ( SELECT getrelatedwrecks.hob_uid AS source_id,
          getrelatedwrecks."related to"::integer AS related_id
          FROM getrelatedwrecks(the_geom, hob_uid, ukho_id,
name, date_text, description, craft, "monument type",
construction, propulsion, nationality, "manner of loss",
cargo, length, breadth, height, displacement, "related to")) r
ON n1.hob_uid = r.source_id
   JOIN nrhe_full n2 ON r.related_id = n2.hob_uid) related;
```

The double join accounts for instances where the source_id and the related_id exist in the reverse sense and will in this case only create one link rather than two coterminous links. I.e. If Wreck A is related to wreck B it is possible that the record for wreck B contains a relation to wreck A also (but not in all cases). If this happens the link is only created once instead of being duplicated. This is a benefit of using database joins with relational data where this type of join can be easily configured.

At this stage the database was configured ready to be consumed by web service software to publish to WMS and potentially other web service formats.

To summarise; The PostGRES / PostGIS database contained the following layers:

Source Data

- NRHE Point, Line and Polygon geometry data layers
- NRHE text data from the XML source data

- MTF Wrecks & Obstructions Geometry Layer (product view of the UKHO Wrecks and obstructions dataset)

Product Views

- NRHE Full – The combined view of the multiple NRHE source layers
- MTF Wrecks & Obstructions MEDIN – A filtered view of the source UKHO wrecks layer

Additional Views

- Linked wrecks (match between UKHO and NRHE wrecks using info recorded in the NRHE text dataset)
- Related Wrecks (NRHE wrecks which relate to other NRHE wrecks).

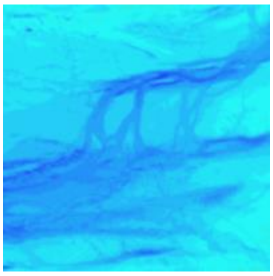
The **Source Data** items are physical data layers which are stored in the database. The **Product Views** and **Additional Views** contain virtual layers created from database views or database function views. In this sense they occupy no physical space on the database and are also dynamic in the sense that they respond to changes in the **Source Data** items without intervention.

Additional base layers:

Using the multiple wreck layers and views in isolation would make a desk based assessment difficult due to the lack of context surrounding the features. To provide minimal support to these layers OceanWise opted to provide supportive base mapping that could be used along with the wreck data for the purpose of the assessment.

This consisted of:

A bathymetry DEM (Digital Elevation Model) produced using the EMODNet European 1/8th arc minute gridded open bathymetric data.



A contour model built from the above DEM. This was produced using Open Source tools (gdal_contour)

A land/terrain dataset that could provide users with a spatial context around their assessments. It needed to include labelling of villages, towns and cities and road information as a minimum. The Open Source “OpenStreetMap” project dataset provided the obvious choice for a web service solution as a web service already existed that could be used for the pilot project.

Official IMO Traffic Separation Schemes. OceanWise licence these annually for use in their data products. The traffic separation schemes layer indicates areas of high marine traffic which may be useful in an assessment of historical wreck information.

Provision of Web Services

OceanWise already use the OpenSource software package GeoServer to manage and serve web services. Apart from the obvious cost benefit, GeoServer presents a package that can be configured and used from an entirely GUI (graphical user interface) or by scripting or otherwise manipulating xml configurations files. In addition to the support of most common web service protocols it also supports and manages data access and styling information. More information can be found at: http://geoserver.org/

Due to familiarity and suitability GeoServer was used in this project to produce the web service layers and serve them online.

The process was made simple because the PostGIS database connection was made available to GeoServer as a comprehensive “Data Store”. The layers within the database could then be published in a single step to all supported web service formats. For this project, it was only required to serve the data via the WMS protocol (OGC WMS 1.1.1 and 1.3.0) so the other web service protocols were disabled for these layers. In addition the layers were served to WGS84 Lat / Long and WGS 84 Spherical Mercator (web Mercator) only. Other coordinate reference systems are supported by GeoServer but it did not seem necessary to add the complexity during the pilot study stage.

For maximum performance and security, GeoServer was setup on OceanWise’s dedicated cloud server where it would provide the most computational power for the rendering of WMS layers and lowest latency within the web communications protocol (HTTP transactions).

For the pilot study it was decided not to restrict access to the WMS layers or hide them from web browsers although they were published in a read-only format. Authentication & authorization is available in a user/group/role structure within GeoServer providing maximum flexibility to a multi-user environment.

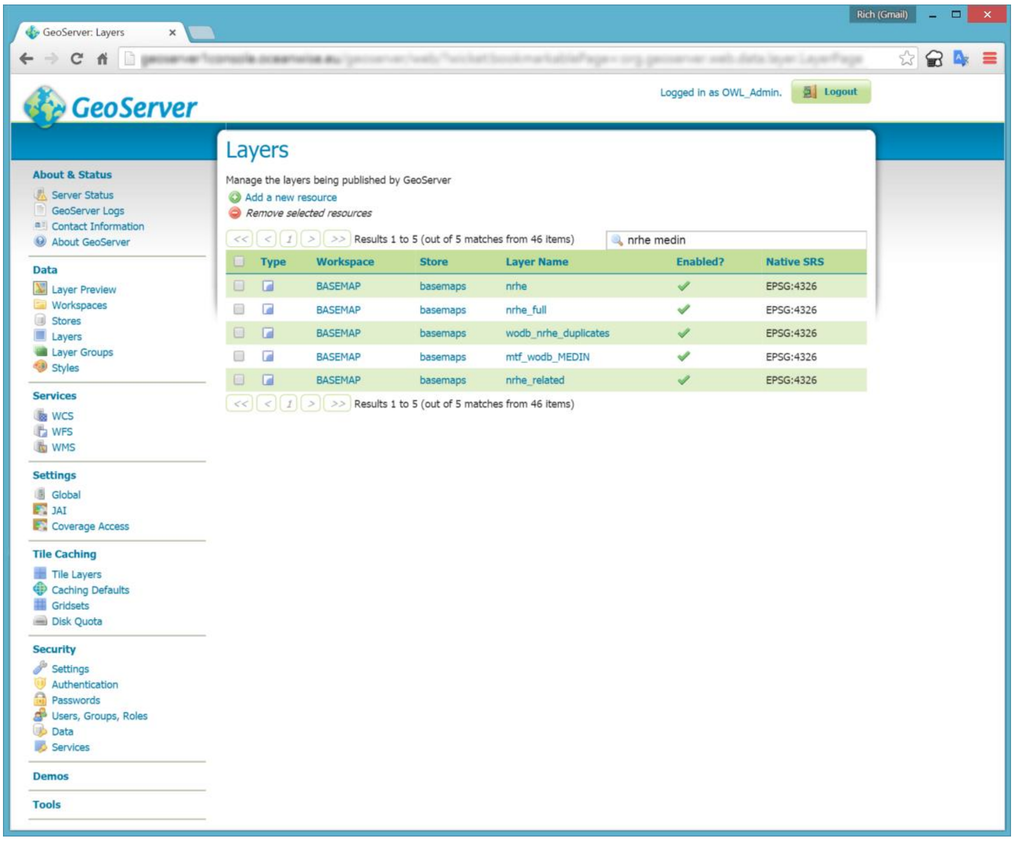


Figure 3: GeoServer configured to serve PostGIS data layers to web services.

Once the WMS layers were configured – they required styling. GeoServer supports SLD style format (Stylized Layer Descriptor) which is an xml format file and therefore an open data format. The SLD file describes the colours and styles of features within a dataset, and the images/graphics used to symbolise point features as well as being responsible for labelling / scaling and handling of overlaps within dataset features.

A useful feature of the SLD format is that it supports rules and filters. This means that rather than just a file containing style items, an SLD can also be used to define the decision making process by which a layer is styled in a certain way, for example according to information within the dataset (attribute filters) and the map configuration of the client software (zoom settings).

Styles were created to show wreck items using IHO S52 symbols for obstruction, non-dangerous wreck, dangerous wreck and wreck showing portion of masts or hull (superstructure). The symbols were created as png images using an OpenSource image editing suite (Paint.NET) and applied to the MTF Wrecks and Obstructions dataset and NRHE dataset using an SLD file.
For the MTF Wrecks & Obstructions layer the attribute “Category” was used to denote which symbol to use. For the NRHE dataset, there is currently no definition of wreck category that matches those described above so there could be no definition in the style descriptor. For the NRHE dataset, the non-dangerous wreck symbol was used for all features, shaded red to differentiate from the MTF Wrecks and Obstructions layer. When zoomed in beyond a scale of 1:200,000 a label is produced showing the contents of the “name” attribute if it is populated. When zoomed out beyond 1:5,000,000 all wreck features are drawn as a small circle to avoid cluttering the display.

To summarise, the wreck symbols are as follows:

MTF Wrecks & Obstructions:

- Wreck or Obstruction Small Scale
- Wreck - Undefined
- Wreck - Non Dangerous
- Wreck - Dangerous
- Wreck - Distrbuted Remains
- Wreck - Showing Masts
- Wreck - Showing Hull
- Obstruction

NRHE Wrecks

- Wreck or Obstruction Small Scale
- Unclassified Wreck

The WMS was configured to support two additional features; wmsGetFeatureInfo and ecqlFiltering.

A brief explanation of each of these features is useful.

wmsGetFeatureInfo
This is different from the normal wmsGetFeature request in that the server will respond not with a rendered image of the map area but with an attribute table. This provides record details for those features at a certain selected point on the map canvas. The client, whether a GIS or a web map application can request the format of the response as an xml data table or a marked up data table in html.

The WMS server (GeoServer) was configured to respond with a single feature representing the closest feature to the selected coordinates, providing the selected coordinates were within 10 pixels of a feature at the rendered zoom scale. The functionality of wmsGetFeatureInfo can be compared to the item properties feature found within a GIS, and commonly within a GIS it is this feature that is used to send wmsGetFeatureInfo requests to the wms server.

ecqlFiltering
Not an OGC WMS standard, but a property supported by GeoServer as a "vendor specific" parameter is the CQL Filter or ECQL Filter. This allows (as part of the standard wmsGetFeature request) an optional CQL/ECQL statement to be provided to filter the data that is returned. The purpose of this was to deliver the functionality akin to a GIS "Query Filter" within the published datasets.

ECQL filters can support attribute filters such as: 'cargo LIKE "coal"' (which shows only features that have the specified attribute value) or spatial queries like 'DWITHIN('selected wreck', 'all wrecks', distance)' which shows all wrecks within a specified distance of a specific wreck feature.

ECQL filtering allows users freedom to query the datasets in a more verbose way than can be offered by simple preset layer filters, providing they are familiar with the ECQL language and concepts. For more information on ecql filtering within GeoServer follow the link: http://docs.geoserver.org/latest/en/user/filter/ecql_reference.html

The urls to access the wrecks pilot layers are:

NRHE Full:
http://geoserver1.oceanwise.eu/BASEMAP/nrhe_full/wms
Layer name: "nrhe_full"

MTF Wrecks & Obstructions:
http://geoserver1.oceanwise.eu/BASEMAP/mtf_wodb_MEDIN/wms
Layer name: "mtf_wodb_MEDIN"

The Web Application Interface

For non GIS users an interactive web map interface provides basic functionality to anyone with a web browser and an internet connection.

The web application was created and hosted on OceanWise's cloud server and utilized only Open Source technologies:

- **HTML:** To build the structure of the system as a web page
- **CSS:** To provide consistent and reactive styling and responsiveness to the system
- **Javascript:** The scripting language which provided the functionality in the system, including the Open Source mapping library "OpenLayers 3" which was used to create the interactive map itself.

The web application provides discovery based functionality for non specialist users. The map itself was designed to fill as much of the screen as possible with minimum interference from control buttons and other paraphernalia. The image below shows the web map display in its full map view with all layers activated:

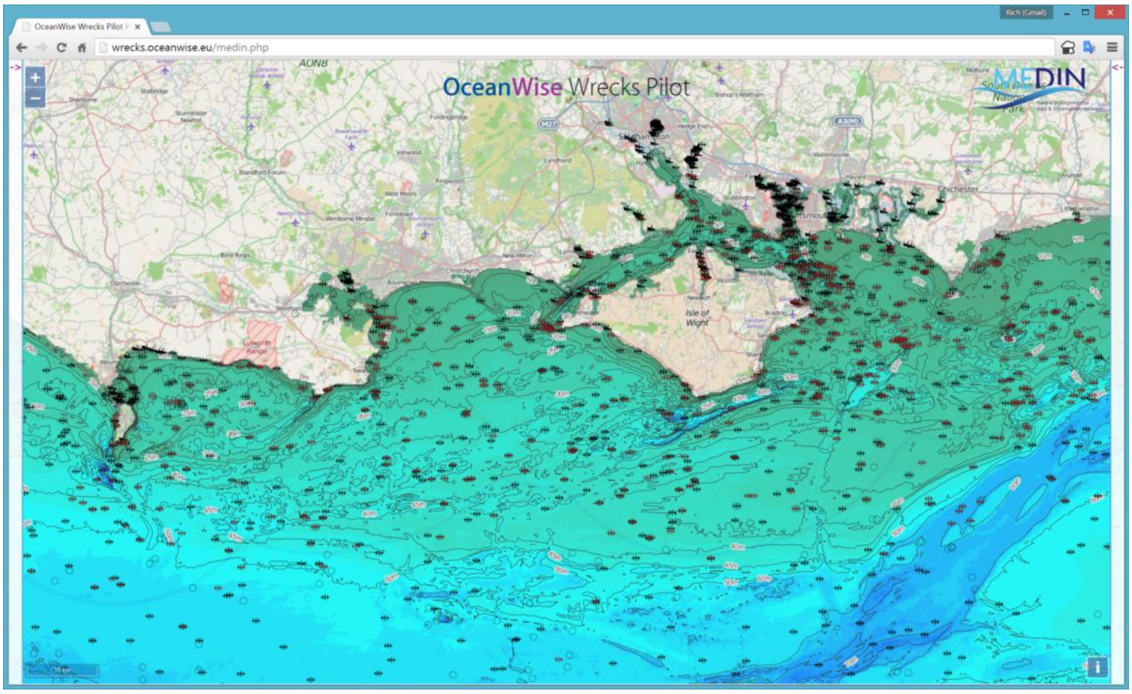


Figure 4: Web map application in full screen mode showing basic control items

The map can be panned (moved) by dragging with the left mouse button and zoomed using the mouse scroll wheel or by clicking the plus/minus buttons at the top left of the window. In the bottom left and right corners respectively there lies a scale bar and mouse pointer coordinates in Decimal Degrees.

To the sides of the map screen are two control bars;

To the left is a "Layer Controls" bar. This can be expanded by clicking the arrow at the top of the window, and collapsed in the same way. The control bar allows layers to be switched on or off, and houses controls for the selection, filtering and measurement tools. Options are active when coloured blue, and inactive when coloured white.

To the right is an "Item Properties" bar. This can be expanded or collapsed using the arrow button at the top of the window. The control bar is empty in normal use – but when a feature is selected the feature attribute table is drawn here and the control bar is automatically expanded if it is not already. If the attribute table is too large to fit within the screen it can be scrolled to show the information that is hidden. A feature is selected by firstly activating the selection tool. A checkbox shows which layer is currently under

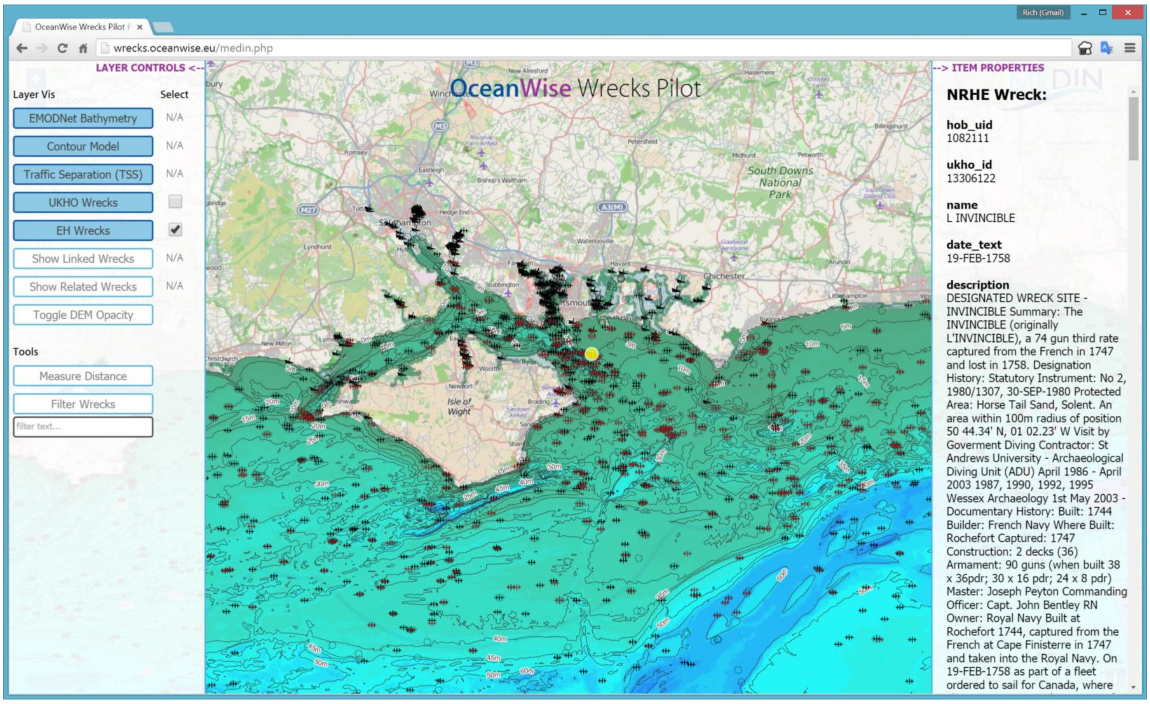


Figure 5: Web map application with advanced controls expanded.

Basic tooling provided in the web application allows users to measure distances on the map screen. Measurements are calculated over great circle (geodesic) routes although connecting lines are drawn using a loxodrome (rhumb line) for simplicity. This means that distances calculated between wreck features are represented as the real-world distance to travel between the two features.
The filtering feature of the web application supports the same ecql filtering that was discussed earlier, using a simple web form input and toggle button.

5 Conclusions & Recommendations

Source Data & Data Processing

The nature of the supply of source data presented interesting challenges. Partially this was due to limitations within the chosen supply formats. For the NRHE dataset, the GIS data in shape file format provided issues in the fact that a different file was supplied per geometry type. Point, Polygon and Polyline. The attribution on these datasets was also a problem. Both the field names and the field contents were potentially truncated meaning that data could be lost (cut short).
The provision of an xml file alongside the GIS dataset to include full textual information mitigated the loss of information form the shapefiles but presented problems in that the dataset needed to be passed through an xml passer to extract a data table and then combined with the GIS dataset.

The UKHO wrecks and obstructions dataset was supplied in a single open data file (CSV format) which meant it was easier to process. However, the data in CSV format are not inherently GIS compatible as the geometry data is not standardised. Formatting the dataset to a standard open data format would allow a GIS system to read the data directly but it could also be loaded into database systems using standard tooling.

Recommendation: *The NRHE and UKHO datasets to be provided in an open GIS compatible data format where all content can be accommodated. Suggest GML (Geographic Markup Language) or GeoJSON (Geographic JavaScript Object Notation). Both of these formats are supported in all major GIS systems.*

In loading the data for the pilot project – tools were developed to ingest the source data to a database in their native state. By following the recommendation above, the process to load the data to a database and create a useable spatial layer is greatly simplified.

Utilising a relational database with spatial support provided great flexibility in manipulating the source data. It was possible to support the different coordinate reference systems in use by the source providers and to provide multiple end layers/datasets by using database views (virtual datasets). This reduced the volume of data stored in the database and the work involved to re-produce and update content as changes are made. The NRHE dataset was provided in the British National Grid coordinate reference system. So far this has not presented a problem as the database was able to manage and convert this as required. The potential issue is in the creation of features in these coordinates, as British National Grid is not "realised" offshore, therefore transformation parameters do not have a defined accuracy. A further issue is that offshore positions may fall outside of the British National grid extent and will be rejected by GIS systems. A solution to this problem is to use a coordinate reference system that is valid within a larger extent.

Recommendation: *Continue to use a relational database to store the source data and to manage and produce derived datasets/layers.*

Recommendation: *The NRHE dataset to be provided in a Coordinate reference system that is valid across the entire UK waters. Suggest WGS84 Geodetic – EPSG:4326*

It was noted that the UKHO wrecks and Obstructions layer contained a great number of attributes, and for the pilot project OceanWise's Marine Themes product version of the

dataset (MTF Wrecks and Obstructions) provided a more concise dataset. To produce a more functional end product and consistency between the datasets some degree of attribute mapping between the NRHE dataset and the UKHO dataset would be necessary so that information from both datasets can be compared. Some of these mappings are simply 1 to 1 such as UKHO:CARGO → NRHE:CARGO but others would require multiple fields to match the contents from one dataset to the other,
UKHO:CIRCUMSTANCES_OF_LOSS, UKHO:SURVEYING_DETAILS,
UKHO:GENERAL_COMMENTS → NRHE:DECRIPTION (FULL).

Recommendation: *Utilise the full UKHO Wrecks and Obstructions dataset so that greater variability in attribute data is available to users. Attempt to "map" matched fields within the database so that a common, comparable attribute set exists for the user to interrogate in all end products. Mappings must be available to end users within metadata for purposes of lineage and context.*

Provision of WMS & WFS endpoints

The creation of the WMS & WFS endpoints was necessary for the web map application to function. However they are also relevant in achieving other objectives in the project.

Firstly, analytical capabilities were a requirement of the service. Whilst this would have involved a great deal of time and expense to integrate fully within a web map application (effectively an online GIS), it was much simpler to utilise existing GIS desktop packages by providing web services that would be compatible with them.

The WMS endpoint can be consumed in any desktop GIS that supports the WMS specification (versions 1.1.0 – 1.3.0). Once loaded as a layer the full functionality of the desktop GIS is at the disposal of the user. One comment in the use case focused upon the style types used within the WMS service. The WMS service can support the use of multiple style types that can be selected by the end user. Whilst this is not as flexible as defining custom styles, it means that a selection of possible styles can be provided for the user to choose from, including options for the use of labelling rather than this being an automatic decision of the service, a factor only of scale (zoom level).

Recommendation: *Further style options to be created to support user requirements of the WMS. This should include variations in symbol colour and options to include labels or not as a minimum. Further options could include theming on different common attributes such as "cargo type", "nationality", "construction", "length/tonnage".*

The WFS endpoint provided a solution to the objectives to provide data download compatibility and goes some way towards the concept of a linked data URI. Effectively a wreck feature can exist at a specific address on the web, the response of that address not being a website but the feature itself in a GIS compliant format. The WFS was not configured to allow feature editing although it is a possibility within the WFS specification.

Within a GIS the WFS layer behaves in an identical fashion to a regular GIS overlay. The features can be selected, interrogated, filtered, queried and styled exactly how the user requires. This makes the WFS method most suitable to advanced users. It also allows the data to be stored into a file on the users local system, I.e. it allows the data to be downloaded. The WFS functionality performed well in the use case scenario and required no additional time to configure above that of the WMS using GeoServer so there are no further recommendations for this item.

The Web service Online Application

The purpose of the web application was to provide a discovery portal for the wrecks datasets. In this sense it did not need to provide advanced GIS functionality but it did need to provide a concise and user-friendly interface to explore the wreck datasets with appropriate base data.

The web application did this successfully by using the WMS end points along with other WMS feeds for basemap and contextual datasets.

As tested in the use case scenario, the web map application provided functionality to pan and zoom a map of the wreck datasets and allow the user to switch layers on and off. Other tools available allow for features to be selected (1 by 1), distances to be measured and filters to be applied using a CQL filter function. Missing from the web interface is the ability to download data to a file. This is due mainly to the exclusive use of WMS services in the system which does not support data download. Using WFS layers would allow more flexibility within the web application than is currently provided.

Recommendation: *Add a "download data" function. This would need to allow the user to download either or both wreck layers to a GIS file format by exploiting WFS functionality. For simplicity, the current view extent could be used as the spatial predicate so that the user has the option to download specific areas only.*

The two major critiques of the web application were the complexity of the CQL / ECQL filter system, and the lack of user information or guidance to explain layer definitions and how to use the tools.

Recommendation: *Add a "help" / "user guide" button which details what the layers contain and how to use the tools in the system.*

Recommendation: *Limit / guide the use of the CQL filter function by providing a set of drop down menus to select the query to use and the values to query against. Fully user controlled querying is provided by using the WMS/WFS layers in GIS desktop so there is no overall loss of functionality in doing this.*

The web application is currently configured to allow any user access to the system. By exploiting the web server or web map server security functionality restrictions could be placed on the system to allow authenticated users to use the system, if this should be required.

The use case scenario indicated that the "Linked Wrecks" and "Related Wrecks" views were useful in the desk based assessment and that they helped the user to assess matched wrecks more easily than the non-spatial equivalent comparisons. That said, there is variation in the preferred way that the relationships are shown on the map. Perhaps larger scale testing will allow a better assessment of how to best show these features on map going forward.

Linking Wreck Data across the UKHO and NRHE datasets

Producing matched wreck information was a challenge in this project. This was due to a number of unforeseen challenges, most of which are inherent in the underlying data. Most recommendations in this section would need to be undertaken by the data custodians to provide improvements.

Two issues presented the major barriers to linking wreck data.

Firstly, the NRHE dataset contains the matched field for “linked” wrecks from other datasets. This field within the NRHE xml text dataset potentially contained multiple values per wreck. It was a challenge to extract just the identifier that would present a match to the UKHO wreck ID. This was further hindered as the UKHO ID sometimes contained an ID from the older now discontinued UKHO database as well as the newer current database so the process had to account for this and select only the newer wreck ID to find a match.

The second issue was that the UKHO wrecks and Obstructions dataset contained in some cases, multiple records for the same wreck. Due to the wreck ID’s being different it was only through the Name and description fields that this effect could be identified. For example one record may have the vessel name “INTERCEPTOR (POSSIBLY)” and another wreck record in a close but not identical spatial location may have the name “INTERCEPTOR (PROBABLY)”. It is the assessment of the definition of “probably” and “possibly” in this case which relate to the confidence of the record. It is worth pointing out here that wrecks within the UKHO wrecks and Obstructions dataset have not been assessed for duplicate records. The NRHE dataset has been through this assessment resulting in the production of “Related Wrecks” information.

Recommendation: *The UKHO wrecks and Obstructions dataset to undergo self deconfliction, I.e. remove or link duplicate wrecks to each other with a matched features field in the dataset.*

Recommendation: *The NRHE dataset and the UKHO dataset require further assessment to match wreck records between them. It would not be the purpose of the published wrecks data layers to attempt automated matching, because there are various record attributes which can provide information to indicate that two wrecks are the same. This would be a manual, or human controlled process. Currently around 40% of the wrecks within the pilot study area were matched between the datasets.*