Seasearch data report, May 2020.

Seasearch Wales: Data Evaluation and Cleansing.

Dr A.C. Jackson

Date: 5th May 2020

Recommended citation:

Jackson, A.C., 2020.

Seasearch Wales: Data evaluation and cleansing 2020. Report to Natural Resources Wales. Seasearch Report, 5th May 2020. Marine Conservation Society: Ross-on-Wye, UK.

Executive Summary

This report summarizes the work done by Seasearch staff in 2020 to identify, evaluate and correct errors and inconsistencies in positional information for Seasearch data for Wales, held by Seasearch and the Marine Conservation Society. This has been partly supported by funds from NRW, because of their need to use high quality data to meet obligations for understanding and monitoring the marine environment. Approximately four percent of records were incorrectly placed above the high tide level. These have been moved seaward so that they are no longer on land. Just over six percent of records were incorrectly placed in relation to the intertidal area. These records can be incorrect for multiple reasons. These reasons have been described and records enumerated for each reason; recommendations are made about how the inaccuracies should be remedied. They have not yet been corrected, but the information and processes required have been described. New data validation processes have been implemented that drastically reduce the possibility of 'dry' or incorrect intertidal records entering the dataset in future. Such improvements will boost the correctness, credibility and usability of Seasearch data.

Contents

1 INTRODUCTION

1.1 Seasearch

Seasearch is a volunteer underwater survey project for recreational divers to record observations of marine habitats and the life they support. The information gathered is used to increase our knowledge of the marine environment and contribute towards its conservation. Seasearch is coordinated by a Steering Group led by the Marine Conservation Society and including representatives from the UK statutory conservation bodies (CCW, EHS(NI), JNCC, NE, SNH), the Environment Agency, The Wildlife Trusts, the Marine Biological Association, the diver training agencies (BSAC, PADI, SAA, SSAC), Nautical Archaeology Society and independent marine life experts. Seasearch is supported financially by all of the UK statutory conservation agencies and the Environment Agency. Volunteer divers can participate in training courses and this is one of many surveys organized during the diving season. For more information www.seasearch.org.uk

The objectives of the Seasearch programme are to:

• Gather information on seabed habitats and associated wildlife throughout Britain and Ireland, by the participation of recreational SCUBA divers,

• Provide standardized training to enable volunteer divers to participate in Seasearch surveys,

• Ensure the quality of the data gathered,

• Make the data available through websites and reports,

• Raise awareness of the diversity of marine life in Britain and Ireland and its environment through participation of volunteer divers and dissemination of information.

1.2 Marine Conservation Society

The Marine Conservation Society (MCS) is the UK Charity dedicated to the protection of the marine environment and its wildlife. Since its formation in 1983, MCS has become a recognized authority on marine and coastal conservation and produces the annual Good Beach Guide, as well as promoting public participation in volunteer projects and surveys such as Adopt-a-Beach, Seasearch and Basking Shark Watch.

2 BACKGROUND

2.1 Marine Recorder terminology

This is a report to NRW of activity by Seasearch staff to identify, evaluate and eliminate positional errors in Welsh records held by Seasearch, which has been partly supported by funds from NRW, because of their need to use high quality data to meet obligations for

monitoring and condition assessement of marine protected areas (MPA). Data are entered to and saved within an Access-based database called Marine Recorder (MR). To allow ready comprehension of the issues being addressed in this report, some relevant terms are defined here.

Observer records – records from an 'observer form' collected by divers or snorkellers qualified to observer or surveyor level. All data are linked to a single sample.

Surveyor records – records from a 'surveyor form' collected by divers qualified to surveyor level. Data may be linked to one or more samples.

Survey – collection of dives for a stated location or area over a stated time period (often a year)

Survey-event – falls within a survey and is usually a single dive of a stated duration.

Sample – data from a distinct habitat, within a single survey-event. Multiple samples (habitats) per survey-event may be recorded by Seasearch surveyors using a surveyor form.

Location – an area of seabed that can contain one or more survey-events.

Position – The latitude and longitude of a single survey event (and or sample) using the WGS84 coordinate system.

2.2 Seasearch dataset

The Seasearch dataset and its extensive records of species and habitat have many potential and valuable applications. These all require the data to be robust and credible. Seasearch emphasises the importance of accurate position fixing in its training, and the forms completed by the divers go through a number of stages of validation (Appendix 1) before the data are circulated to partners and stakeholders. Unfortunately, due to the primacy of human evaluation and input during this process, errors can enter the dataset. One mechanism that can detract from the credibility of Seasearch records is when the positional information for a record is clearly not correct; for example, the position for a marine species is located on land (creating so-called 'dry' records).

2.3 Dry records

Survey events could be 'dry for several reasons.

1) The survey event was recorded as a line, with a start and end point. In such instances, Marine Recorder interpolates linearly between these two points to find

the mid-point, which becomes the location for that survey event. Where the line follows a curving coastline, the mid-point can be on land.

2) Dives occur and are correctly recorded in habitats such as in tidal lagoons or under piers that can fall inside the MHWS polygon (i.e. the polygon is not always a perfect delimiter of where surveys can occur).
3) Locations are broadly correct, but identified using an imprecise system, e.g. location is incorrectly estimated from a chart or map with coarse resolution. When converted to WGS84 in degrees, decimal minutes, these locations plot above MHWS.
4) The location was recorded incorrectly when completing the survey form, e.g. through incorrectly set GPS, careless writing, use of different land boundaries (e.g. extent of the realm, which approximates low tide level), incorrect identification of survey location using map, chart or web-service.
5) The location was recorded correctly on the form, but subsequent errors were made during digitisation and data entry.

It may not be possible to distinguish between the last two of these.

2.4 Intertidal records

Another example of incorrect positioning, is where the record is shown as being intertidal (below MHWS, but above mean low water), but the minimum depth of the dive is below chart datum (CD). This is a less serious error than being 'dry', but different species and particularly biotopes exist, between intertidal and subtidal conditions. It would be desirable to correct these where appropriate or possible. Survey-events that actually occur at subtidal depths, but show intertidal positions in the database should be repositioned such that the Latitude-Longitude are seaward of the intertidal area. In addition, survey-events that correctly occur in intertidal space should be assessed against the classification of intertidal biotopes.

2.5 Remit

The remit of this report is three-fold.

1) to establish the extent of the dry records for Wales and then to reposition these records as correctly and accurately as possible using available information.

2) To scope the extent of records that:

   i) are incorrectly placed in the intertidal zone (all depths below CD)

   ii) are correctly placed and are entirely in the intertidal zone (all depths above CD), but which have not been assessed against intertidal biotopes

iii) may be correctly placed because they started or ended in the intertidal zone (minimal depth above CD).

3) to establish an objective process by which erroneous data identifed in 2) can be corrected or improved.

3 METHODS

3.1 Identifying dry records

'Dry' records were identified using the QGIS package (QGIS Long-term release 3.4.15). A polygon representing the area in Wales above the height of Mean High Water Springs (MHWS) was created by closing a high water polyline along the England-Wales border using a polyline of European Regions. These shapefiles are feely available in the Ordnance Survey Boundary Line product (https://www.ordnancesurvey.co.uk/business-government/products/boundaryline; October 2019 release). The Boundary Line product relies on capture and maintenance to a generalised scale of 1:10,000.

Locations (as Latitude & Longitude using coordinate reference system EPSG:4326, WGS84) of Seasearch records were plotted against this MHWS polygon. Survey records were considered 'dry' when they occurred inside (i.e. above) MHWS (Fig. 1).

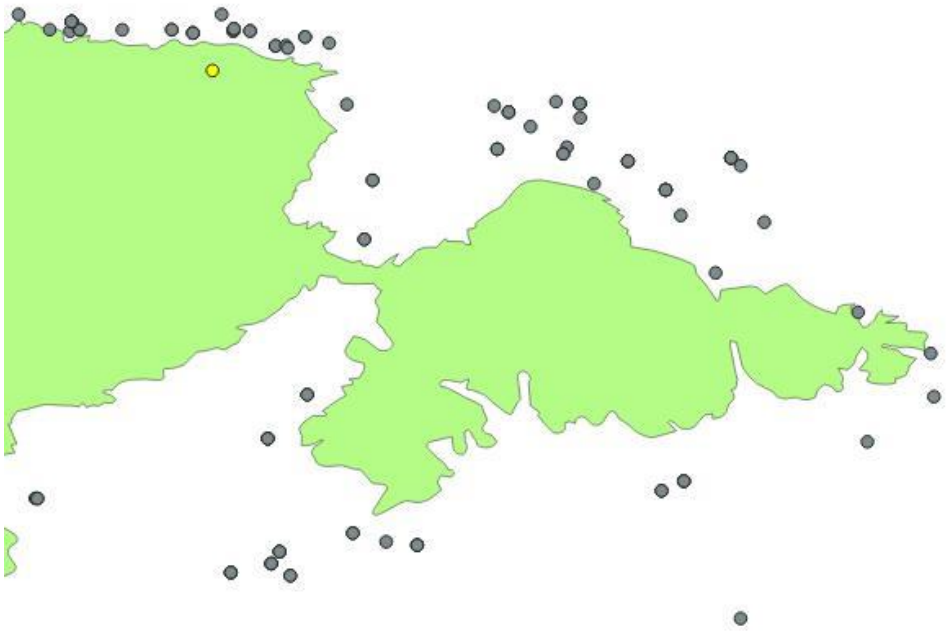

Figure 1. Locations of Seasearch dives around the east of Skomer, Pembrokeshire (green shading). Grey dots are dives that are marked as being below Mean High Water Springs (MHWS), the yellow dot is above MHWS and therefore, incorrectly, on land.

3.2 Correcting dry records

The list of 'dry' dives was processed in sequence, beginning with the oldest. For each dive, the original position in relation to MHWS was reconsidered given a range of information. The information available included the details on the original recording form (specifically the fields SurveyName, EventName, LocationName, Location description, SurveyEvent Comments, Sample Description, and Sample Comments), GIS layers for MHWS and bathymetry, aerial imagery from GoogleEarthPro and GriReferenceFinder online tools, plus expert knowledge from experienced record-verifiers for the area. This information was synthesised to allow the most likely position of the dive to be inferred. This new position was then placed as a marker in Google Earth Pro, which provides the Latitude and Longitude in Decimal Degrees (Fig. 2), WGS84. Decimal Degrees were then converted to Degrees Decimal Minutes (DDM, the required format for input into Marine Recorder). A spreadsheet was prepared listing the new locations for each dry survey-event, along with an explanation of why and how the postion was altered (Appendix 2).

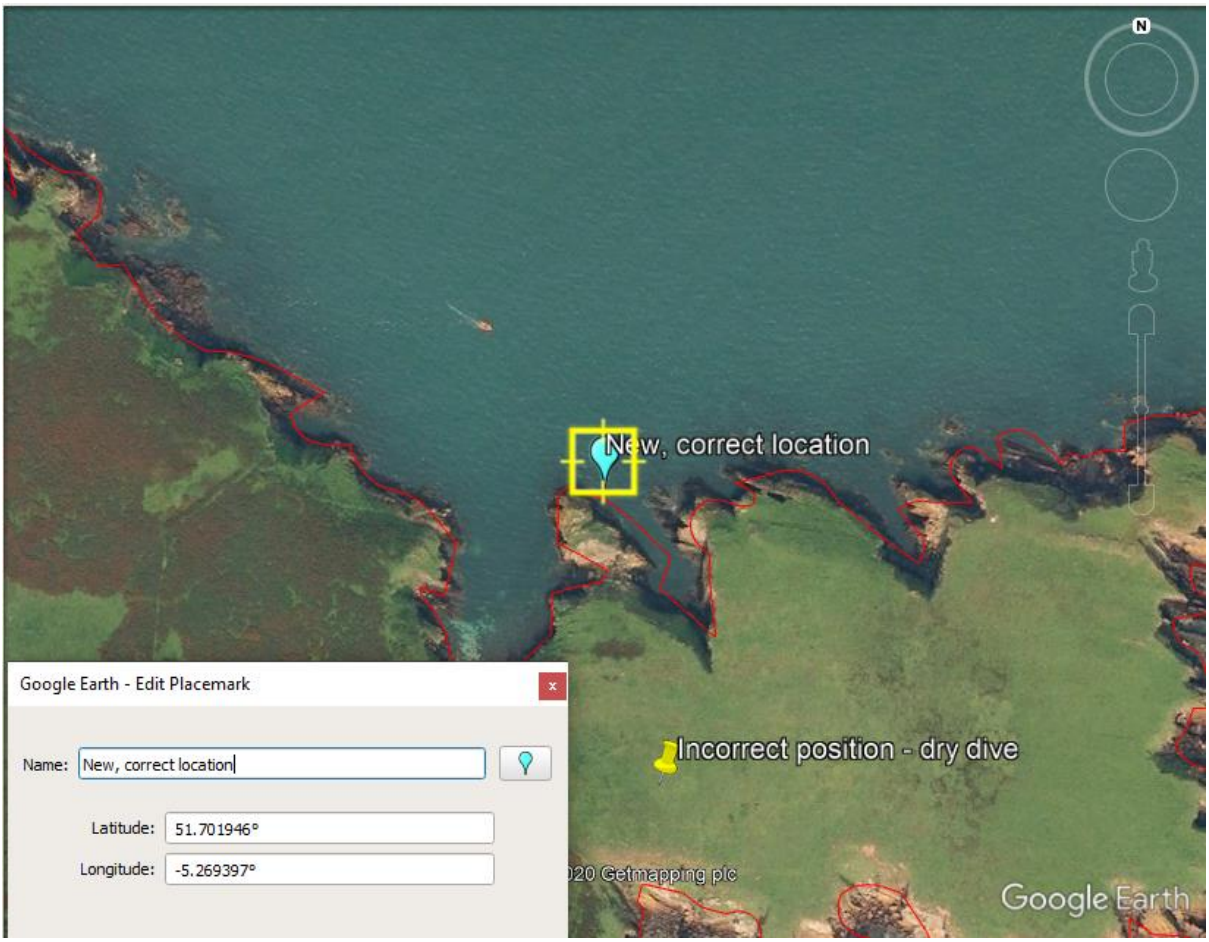

Figure 2. An image grab from Google Earth Pro showing the incorrect, 'dry' location for a dive on Skokholm Island, Pembrokeshire (yellow pin) and the new, correct location (blue marker) derived from information about the dive.

For line-type survey-events (on the **Event** page in MR), the **start** and/or **end** fields were updated with new positions in DDM format. If necessary, the **Co-ordinate system** field was set to 'Lat Long (WGS 84)'. The **Derived from** field was set to 'Derived from Google Earth aerials'. The free-text field on the tab labelled **2. Description** was updated with text describing why and by how much the location was altered and whether the co-ordinate system was changed.

For each survey-event, positions of **all** samples also had to be updated. For samples in line-type survey-events (on the **Sample** page in MR), **start** and/or **end** positions were updated with new locations in DDM format. If necessary, the **Co-ordinate system** field was set to 'Lat Long (WGS 84)'. The **Derived from** field was set to 'Derived from Google Earth aerials'. For line-type survey-events with a single line-type sample, the start and end of the sample were set as for the survey-event. For line-type survey-events with multiple line-type samples, the position of the samples depended on available information, but the start and end of each samples was typically the same as the start and end of the survey event. No instances were found for line-type survey events with point-type samples, although this combination is realistic possibility.

For point-type survey-events and their point-type samples: **start**, **Co-ordinate system**, and **Derived from fields** were updated as above. Line-type samples within a point-type survey event should not exist.

3.3 Identifying intertidal records

The same mechanisms that cause dives to have 'dry' positions (Section 2.3) can also cause dives to have intertidal positions, but intertidal positions are not necessarily incorrect. It is possible and reasonable for Seasearch surveys (particularly by snorkelling) to be done partially or entirely in intertidal habitat. A subsequent assessment of survey positions was done in QGIS using a shapefile representing intertidal areas derived from Phase 1 habitat (NRW 2007).

The upper boundary of the intertidal biotope layer did not coincide perfectly with the the layer for MHWS. The eliminate the risk of double counting records that fall above MHWS and also within the intertidal area, survey events above MHWS were first eliminated using the 'difference' geoprocessing tool in QGIS (these have already been dealt with above). Remaining records were then checked to see if they fell within the polygons of the layer with intertidal habitats ('clip' geoprocessing tool).

Survey events that were entirely intertidal, i.e. all recorded depths were above CD (whether or not they were positioned in the intertidal area) were quantified. The number of these which were positioned inside (correctly) or outside (incorrectly) of the intertidal polygons was identified. Illogical depth entries (upper depth is deeper than lower depth) were also noted.

4 RESULTS

4.1 Dry dives

Of the 3007 suvey-events (dives) collected in Wales since 1983, 125 (including 181 samples) plotted incorrectly above MHWS (Fig. 3).

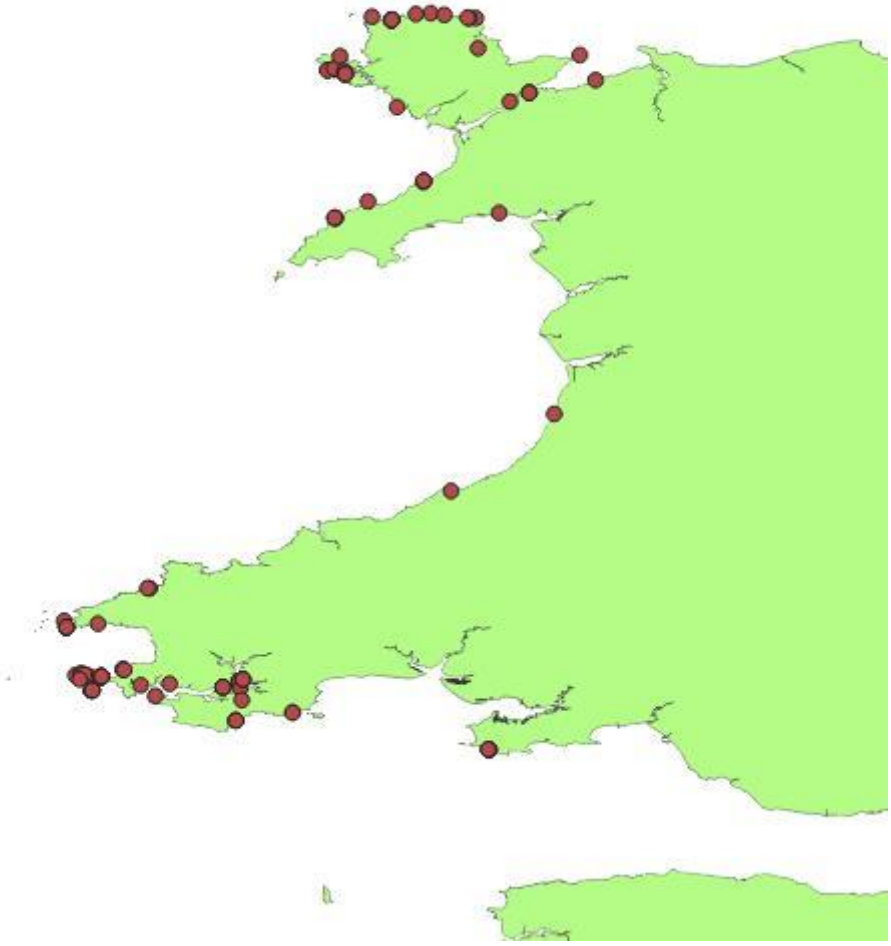

Figure 3. Positions of Seasearch 'dry' dives around Wales (1983-2018).

There was no obvious increasing or declining trend in numbers of errors, although there have been very few errors since 2014 (Fig. 4). The mean (± s.e.) number of errors per year was 5.32 (1.07). The years 2008,2009 and 2012 had relatively large numbers of errors.

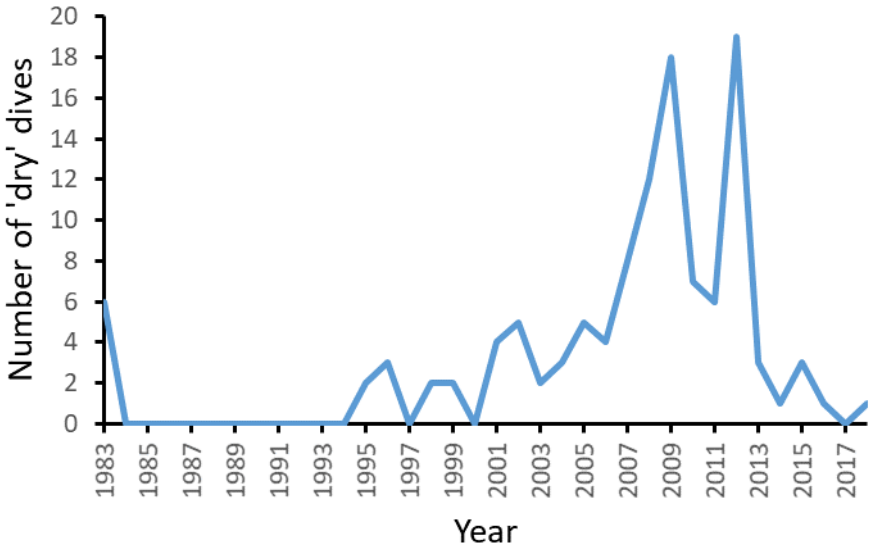

Figure 4. Numbers of incorrect 'dry' records from Wales, each year. The zero values between 1984 and 1994 were because no records were made.

Of these 125 'dry' dives, three were line surveys. One followed a curve in the coastline, meaning that the mid-point was on land. The original start and end points were acceptable. This was difficult to correct, because calculation of the midpoint is an automated process in MR and cannot be edited manually. In this case, the line-type event was converted to a point-type event, with a new, 'correct' mid-point added to MR. The original positions of 'entry' and 'exit' were included in the description of the event. Two were straight lines where the start and/or end fell on land. Positions were corrected in MR.

Four dives were classed as areas with an unreasonaby large size. These survey events and associated samples were converted to point records with a sensible position given information provided in the description. The remaining 118 dives and associated samples were point records. Four dives were marked as being on Trefor Pier (Gwynedd). The pier is not a solid structure that extends to the seabed, yet it is marked as a feature on the MHWS polygon. These dives actually happened below the pier. In the interests of avoiding future confusion, the positions of these dives have been adjusted to be close to the side of the pier.

All other dives were 'dry' due to some combination of reasons 3 – 5 (Section 2.3).

Corrected positions were made available to NRW initially as a snapshot that includes the new records from 2019. Corrected positions will be broadly and freely available via the JNCC following the next release of their public snapshot and via the NBN Atlas following the update in June 2020.

4.2 Intertidal dives

After 'dry' dives had been excluded, there were 267 dives with positions that overlapped intertidal polygons. Of these, eight occurred entirely in depths above CD (i.e. they were actually intertidal dives). Biotopes for these dives should not be subtidal. If biotopes were originally selected from subtidal options, then these records should be re-assessed against intertidal biotopes.

A further 14 dives occurred entirely in depths above CD, but these were positioned outside of the intertidal polygons. For these records, the positions should be corrected so that they overlap with the polygons in the biotope layer and if neccesary, appropriate intertidal biotopes should be selected. One hundred and fifty-five dives had all their depths below CD, but were incorrectly positioned in the intertidal area. These positions need to be moved so that they are seaward of the intertidal polygons. This process is analogous to the correction of 'dry' dives. One hundred and three dives had their shallowest depths above CD and their deepest depths below CD. An intertidal position for these could be quite reasonable and it would be challenging or impossible to determine whether the recorded position was not representative of the actual position of the dive. One dive had a deeper depth shallower than the shallowest depth, this typographical error needs to be corrected.

5 CONCLUSIONS

With the advent of the Seasearch data officer position (December 2018), the importance of positional accuracy has been re-emphasised to Seasearch data enterers and verifiers. In addition, all Seasearch records now undergo a rigorous and comprehensive check of positional information after data have been entered to MR. This is in recognition of the 1:10:100 principle (Labovitz et al., 1993) which suggests that remediation costs more than prevention and that costs of remediation are far smaller than the costs of persisting with bad data. Any errors or questions are returned to the data enterer or verifier for checking and correction prior to the records being made publically available. The chance of dry records appearing in the dataset is now extremely small. On-going elimination of such errors during verification and prior to release of data should prevent the need for similar reports and corrective work in the future. Other forms of positional error may still persist.

Four percent (125 from 3007) of all Seasearch records from Wales were incorrectly placed in relation to MHWS. These positional errors for survey events and their associated samples have now been corrected (Appendix 2). Just under 9 % (267 from 3007) of records are currently positioned in the intertidal area have a questionable position in relation to the intertidal area. For about two-thirds of these, it is possible and desirable for information to be corrected or improved (i.e. re-assess biotopes against intertidal categories, correct erroneous entries for depth, adjust locations to be inside or outside of the intertidal polygons according to the actual depths of the dive). For those with an intertidal start and a subtidal end (or vice versa, according to depth) it may be quite reasonable for the dive position to fall in the intertidal area. In the absence of detailed description on the survey form to indicate otherwise, my recommendation is for these positions to be left as they are.

The precision of divers' positions is seldom likely to be such that different parts of the survey event are linked to their own latitude and longitude. Where dives are fully intertidal, biotope allocation should be made using the littoral component of the JNCC habitat classification.This is normally done using the combination of species recorded and their relative abundances (SACFOR) alongside the description of the habitat. In the case of intertidal dives for Wales, there will also be additional supporting information available from the phase 1 intertidal biotope map (NRW 2007). I estimate that this work could be done within a 2 month period. Records for other countries will be checked and rectified when funding and time permit.

6 SUMMARY

Positional inaccuracies can enter the Seasearch database through human error. One serious positional error is where marine biological records appear as being on land. For Seasearch records in Wales, 118 of these errors were identified and corrected as far as possible, using a range of information. Other spatial inaccuracies include position in relation to the intertidal area. To scope out the extent of these inaccuracies, the different possibilities for errors in this context have been described and their occurrences quantified. They have not yet been corrected, but the information and processes required have been described. New data validation processes have been implemented that drastically reduce the possibility of 'dry' or incorrect intertidal records entering the dataset in future. Such improvements will boost the correctness, credibility and usability of Seasearch data.

7 REFERENCES

Labovitz, G., Chang, Y.S. & Rosansky, V. 1993. *Making Quality Work: A Leadership Guide for the Results-Driven Manager*. University of Texas Press: Austin, Texas.

NRW. 2007. *Intertidal Phase 1 Habitat Survey*, Natural Resources Wales. Available at: http://lle.gov.wales/catalogue/item/MarineIntertidalPhase1HabitatSurvey/?lang=en. Accessed on 28th February 2020.

8 APPENDIX 1. Seasearch quality control process.

Seasearch Quality Assurance Procedures

Seasearch diving and recording is carried out by volunteers. Many of them have a professional background in marine biology and conservation but many do not and are self-taught naturalists. The document sets out the processes which are used to assure the quality of Seasearch data so that they can be used by professionals with confidence.

Seasearch Training Programme

Training is available at three levels to all participants.

Observer Level – this is aimed at volunteers without previous experience of marine recording in British and Irish waters. It comprises a one-day course followed by two survey dives where the individual records are reviewed and discussed with a tutor. The Observer qualification is awarded after completion of a further 3 survey forms.

Surveyor Level – this is aimed at experienced Observers and others with previous relevant experience. The training comprises a two-day course which involves the completion of two Survey Forms (one from video and one from an actual dive). The Surveyor qualification comprises completion of a further 5 Survey forms, two of which are supervised by a Seasearch tutor, and the completion of an ID test.

Specialist level – this is aimed at experienced surveyors to either increase their skills in survey methodologies or individual groups of plants and animals. Courses are workshop style and are led by experts in their field. They are often attended by professional biologists as well as Seasearch surveyors.

In addition to the training process Seasearch produces a series of **ID Guides** aimed at improving in-water ID skills. These comprise:

Seasearch Guide to Marine Life – introductory level containing a selection of widely observed species of plants and animals. (Much expanded and updated second edition published December 2018)

Seasearch Guide to Sea Anemones and Corals of Britain and Ireland – comprehensive guide to all of the anemones and corals found in shallow waters, the only guide of its type. (Two editions)

Seasearch Guide to Seaweeds of Britain and Ireland – again the only guide to be illustrated with in-situ photographs to complement recording by collecting specimens. Equally popular with littoral recorders and divers. (Two editions)

Seasearch Guide to Bryozoans and Hydroids of Britain and Ireland – these are difficult groups to identify but important in biotope terms as they often form significant animal 'turfs'. This is the only guide to contain *in situ* images as opposed to line drawings alone.

Seasearch Guide to Sea Squirts and Sponges of Britain and Ireland - as with bryozoans and hydroids, these groups can form the dominant animal cover in the right conditions but are often confused. As with the other Seasearch guides, this book concentrates on *in situ* features to allow recording without specimen collection. Most of the sea squirts found the shallow waters around Britain and Ireland, together with the more easily recognised sponges, are included in the guide.

These guides help to ensure high quality records as many of our volunteers use cameras and are able to check their images with those in the guide.

Quality Assurance Process for Recording Forms
Validation and verification of the data follows a three-stage process:

Initial validation can be carried out locally or by the National Coordinator depending on who first receives the forms. It comprises allocating a Seasearch number, checking the completeness of the form, checking the position given and checking the species lists for any unlikely species. If there are queries then these are raised with the recorder and photographs requested to check identifications, especially of unexpected species. Either the recorder or the validator can assign a '?' to a taxon record which is then included in the database as an uncertain record. Supporting verification of an identification, in the form of confirmation by a recognised expert, can be appended to the taxon record within Marine Recorder (*e.g.* "identification confirmed by Bernard Picton" for a rare/unusual nudibranch).

Data Entry into the Marine Recorder database is carried out by a small group of experienced personnel, the majority of whom are professional biologists or extremely experienced recorders. There is a manual and supporting guidance for data entry to ensure consistent standards. The person entering the data can add significant value in the way they describe habitats and they also allocate MNCR Biotopes to the habitats identified in the Survey forms. This is a specialised skill which we do not expect volunteers to have. We have produced two manuals to aid the process and again maintain consistency of approach. At this stage the person entering the data can again refer back to the original recorder to clarify any points.

Merging and final checks are carried out by the National Coordinator, supported by the Seasearch Data Officer. This stage consists of merging all of the separate local datasets into a single UK/Ireland file prior to checking and distribution of the data. Once merged, a 'snapshot' of the data is created which enables checks to be carried out of species (looking for unusual or questionable records), completeness of data and consistency over the dataset as a whole. A map is also created which plots all of the records received and this is also checked for significant positional errors. Any changes required are agreed with the person responsible for entering the data and must be carried out by them to avoid the creation of duplicate datasets. The National Coordinator is responsible for distributing the data to the NBN, JNCC and other users.

Ongoing Data Management
Queries arising from users of the data normally come to the National Coordinator (some through the NBN) but may also arise at a local level. They are discussed and amendments made as appropriate by the holder of the dataset at the local level. Any amendments are incorporated in an, at least, annual update of the whole dataset.
This process we believe makes the Seasearch data reliable and of a professional standard. Whilst many of our volunteer recorders are experts in their own right, that is not always the case and the process ensures that records made by less experienced volunteers are thoroughly checked by experienced people prior to appearing in the dataset.

Charlotte Bolton
National Seasearch Coordinator
Marine Conservation Society
Overross House, Ross Park
Ross on Wye HR9 7US.

Seasearch QA procedures (v2 – updated by CEB November 2017; v3 – ID guide update (CEB Dec.2018))

9 APPENDIX 2. Changes made to ‘dry’ positions.

Click on the icon to open the spreadsheet.

Wales_dry_records_
1983-2019.xlsx