

SIFIDS

Scottish Inshore Fisheries
Integrated Data System

Work Package 6 Final Report

**Development of a Pilot Relational Data
Resource for the Collation and
Interpretation of Inshore Fisheries Data**

Project code: WP00(6)SIFIDS



Published by	Marine Alliance for Science and Technology for Scotland (MASTS)
This report/document is available on	The MASTS website at http://www.masts.ac.uk/research/emff-sifids-project/ and via the British Lending Library.
Dissemination Statement	Following the successful completion of the EMFF funded project 'Scottish Inshore Fisheries Integrated Data System (SIFIDS)' the University of St Andrews and Marine Scotland have consented to the free distribution/dissemination of the projects individual work package reports, in electronic format, via MASTS and the British Library.
Disclaimer	This report reflects only the view(s) of the author(s) and the University of St Andrews, MASTS, Marine Scotland, the Scottish Government and the Commission are not responsible for any use that may be made of the information it contains.
This report was prepared for	University of St Andrews (UoSA) Contact Dr. Mark James Telephone +44 (0) 1334 467 312 Email maj8@st-andrews.ac.uk
The report was completed by	SeaScope Fisheries Research Ltd. Contact Grant Course Telephone +44 (0) 1461 700309 Email grant@seascopefisheries.co.uk
Recommended citation style	Tidd, A.N., Ayers, R.A., Course, G.P. and Pasco, G.R. (2019). Scottish Inshore Fisheries Integrated Data System (SIFIDS): Development of a Pilot Relational Data Resource for the Collation and Interpretation of Inshore Fisheries Data. Published by MASTS. 49 pp.



EXECUTIVE SUMMARY

The competition for space from competing sectors in the coastal waters of Scotland has never been greater and thus there is a growing need for interactive seascape planning tools that encompass all marine activities. Similarly, the need to gather data to inform decision makers, especially in the fishing industry, has become essential to provide advice on the economic impact on fishing fleets both in terms of alternative conservation measures (e.g. effort limitations, temporal and spatial closures) as well as the overlap with other activities, thereby allowing stakeholders to derive a preferred option.

The SIFIDS project was conceived to allow the different relevant data sources to be identified and to allow these data to be collated in one place, rather than as isolated data sets with multiple data owners. The online interactive tool developed as part of the project (Work Package 6) brought together relevant data sets and developed data storage facilities and a user interface to allow various types of user to view and interrogate the data. Some of these data sets were obtained as static layers which could sit as background data e.g. substrate type, UK fishing limits; whilst other data came directly from electronic monitoring systems developed as part of the SIFIDS project. The main non-static data source was Work Package 2, which was collecting data from a sample of volunteer inshore fishing vessels (<12m). This included data on location; time; vessel speed; count, time and position of deployment of strings of creels (or as fleets and pots as they are also known respectively); and a count of how many creels were hauled on these strings. The interactive online tool allowed all the above data to be collated in a specially designed database and displayed in near real time on the web-based application.

During development of the online tool, several challenges were faced by the project team. The most important of these was access to relevant data sets. Even though the type of data that would be useful could be specified and the data owners identified, it was often very difficult to get permission to use and display these data. Data has a value and is considered an important economic resource. There are also complicated legal concerns regarding data privacy and misrepresentation through poor interpretation. Information on where and when a fisher is active and how much fish has been caught is commercially sensitive data and fishers quite rightly want to protect their data from their competitors, especially when marine space is limited. This limited the quantities of real data sets that could be explored in this development. Instead fabricated data sets had to be used where it was felt important to demonstrate the online tool's capabilities. That said, all vessels that participated in the WP2 equipment trials and the vessels who completed surveys and observer trips under WP8, all agreed for their data to be made available to this development project, with the assurance of anonymity.

A stipulation of the development of the online tool was to use open source software wherever possible. The development team used PostgreSQL for the base relational database engine and R software for the development of the user interface. R software was chosen because it is completely open source and it allows transparent technology transfer between disciplines and researchers. It also has a large toolset of cutting-edge packages available, especially related to statistics and visualisation. The databases and the application instance were developed, tested and stored on Amazon Web Service servers.

The final design of the online application included information on substrate type, Marine Protected Area locations, activity and effort data from WP2 vessels, annual landings data (fabricated), survey response data from WP8 vessels, management areas and fleet dynamics data from official statistics.

Users were provided with different levels of access to ensure that confidential data could only be viewed by authorised users or the actual data suppliers. For example, a fisher could view their own unaggregated fishing activity data but could only view data from other fishers if it was aggregated to a minimum of 5 vessels within 5nm² cells, Regional Inshore Fisheries

Group (RIFG) chairs could only view the data supplied by their members, again aggregated to 5nm²; whereas Marine Scotland could access all raw data at an unaggregated level.

Challenges were encountered during the development project. These included sourcing and accessing data sets, getting users to decide which data would be of use to them, updating the “live” data, data volume and data storage, and data processing loads. However, all of these issues were addressed for the duration of this time limited project. It was acknowledged that changes, agreements and control procedures would need to be implemented if this data tool was to be released for general use in the future; Consideration would need to be given to managing and updating the near live data, agreements would need to be in place with data suppliers that encompass the issues of privacy, data use, publication and ownership; the host servers would need to be powerful enough to store and process the amounts of data being received; user permissions would need to be closely managed to account for changes in vessel ownership etc. and the needs of the different user groups e.g. tourists, fishers, scientists, fishery managers, policy makers, would all need to be reviewed and altered.

The feedback from users was positive. There were some user errors and technical issues that impacted on their experience but to have such useful and positive feedback at this stage of a prototype resource development demonstrates that there is a need for this type of tool. This type of resource is only as good as the data that is available to it and the tools used to extract and display the data in a user-friendly way. With refinement and the correct data resources, this tool would be a useful asset to a wide range of marine users.

As part of further work conducted by the University of St Andrews (USTAN) in using and adapting off the shelf tracking devices used in the road transport sector, the WP6 data base was transferred to USTAN servers for further development to accept data from these devices and to incorporate data derived from the SIFIDS App which is designed to capture FISH1 form data and observations from fishers. In addition, some automated analyses of vessel track and effort data has been added to the database functionality based on works conducted as part of WP8b. This additional work and the upgraded user interface is documented in a supplementary SIFIDS report.

CONTENTS

1	Background.....	7
1.1	Background to the Scottish Inshore Fisheries Integrated Data System (SIFIDS) Project 7	
1.2	Background to WP6	7
1.3	Original specification for the work package.....	8
1.3.1	Approach	8
1.3.2	Outputs and deliverables	9
2	Methodology.....	9
2.1	Modifications to the original specification.....	9
2.2	Definition and sourcing of data	9
2.3	Acquisition of the data	10
2.4	Selection of tools	11
2.5	Data locations and hosting	12
2.6	Data users, data anonymisation and aggregation levels.....	13
2.7	Development process.....	14
2.8	Performance issues and problems	15
2.8.1	Data volume and server size.....	15
2.8.2	Points versus Polyline.....	15
2.8.3	Map displays and colours	16
2.9	Project replication.....	17
3	Results	17
3.1	Database.....	17
3.2	The Application.....	17
3.2.1	Introduction page	18
3.2.2	Interactive fishing activity map (Map data tab)	18
3.2.3	Static layers	22
3.2.4	Vessel characteristics	23
3.2.5	Fishing drivers	24
3.2.6	SIFIDS catch and effort.....	25
3.2.7	RIFG reporting.....	27
3.2.8	National catch and effort	28
4	Discussion.....	28
4.1	Caveats and assumptions	28
4.2	Data sets, acquisition and hosting	28
4.3	Data security	30
4.4	Data sharing and ownership	30
4.5	The selected tools	31
4.6	The resultant application	32

4.7	The project concept.....	32
4.8	Opportunities for expansion.....	33
5	Conclusions.....	35
	References	36
	Acknowledgements and Data Sources.....	37
	Annex A – Detailed feedback responses (anonymised)	38
	Responder 1	38
	Responder 2	39
	Responder 3	40
	Responder 4	44
	Responder 5	45
	Responder 6	46
	Responder 7	51

LIST OF FIGURES

Figure 1.	The introduction page of the SIFIDS Marine Database Application.	17
Figure 2.	Interactive fishing activity map.....	19
Figure 3.	The control panel for the interactive map.	20
Figure 4.	Available vessel sessions table. Clicking on the arrows next to the VesselID column will sort by this criterion, but only one criterion can be selected for sorting at a time.....	21
Figure 5.	The track produced (green line) when a row of vessel data is selected.	22
Figure 6.	An example showing the fisher activity in the form of a day's track (green line) and the hauled pots (red circles) detected during the trip.	22
Figure 7.	Selection and labelling of layers illustrated by using "substrate" as an example. Also shown is a fishing trip track and the associated effort (pots detected).	23
Figure 8.	Vessel characteristics table for an example area and characteristic. Clicking the dropdown arrows will subset the data and display new plots.	24
Figure 9.	An example of the type of survey data that can be displayed using the WP6 application. This shows the responses provided by vessels as to why they usually stopped fishing, as collected during the WP8a surveys.	25
Figure 10.	An example of the 5nm square heatmap displays that can be requested. This figure shows the pot count for the WP2 vessel operating in the Solway Firth in July 2018. .	26
Figure 11.	Menu panel for selecting data for displaying on the heatmaps.	27
Figure 12.	The RIFG reporting page that allows data to be aggregated by time period and Scottish Marine Region (SMR).	28

LIST OF TABLES

Table 1.	Data sources that were used to build the application functionality.	11
Table 2.	R packages used to develop the SIFIDS prototype application.....	12
Table 3.	Examples of access rights for the application	14
Table 4.	Potential additional data sources and layers for the WP6 application.	34

1 BACKGROUND

1.1 Background to the Scottish Inshore Fisheries Integrated Data System (SIFIDS) Project

SIFIDS was conceived by the Marine Alliance for Science and Technology for Scotland Directorate (MASTS) to align with the three key elements highlighted in the 2015 Scottish Inshore Fisheries Strategy, which are:

- enhancement of the evidence base on which management decisions are taken
- promotion of stakeholder participation
- the incorporation of inshore fisheries activities within spatial marine planning

The overall project was commissioned to take place over a 30-month period, commencing December 2016 and running through until May 2019. The SIFIDS Project is subdivided into 12 highly integrated “Work Packages” (WPs) ten of which are focused on the development of various aspects of the system and/or the collection of data relating to the inshore sector. The remaining two WPs encompass the engagement of the inshore industry and overall project coordination and oversight. Of the twelve WPs five were undertaken by researchers and staff based at the University of St Andrews while the remaining seven were assumed by external contractors appointed through open procurement. The majority of project WPs actively sought to engage with the industry as they undertook their research.

1.2 Background to WP6

In order to efficiently utilise the data streams collected for inshore fisheries management and marine planning the build requirement was for a secure, powerful and flexible data resource solution that could be easily accessed and interrogated by fisheries managers, marine planners, scientists and potentially, individual fishers. In addition to secure and tiered access to data, there was a requirement for the system to provide a range of easily accessible outputs which should be available to defined user groups based on the access rights of that user group to the underlying raw data.

Ready access to a centralised, up to date fisheries data resource would also assist Marine Scotland in achieving an objective listed under Outcome 2 of the 2015 Inshore Fisheries Strategy¹, that of implementing an appropriate form of vessel monitoring to provide good quality information on the footprint of inshore fishing by 2020. In addition, being able to draw upon a cumulative record of fishing activity data, the regional fisheries groups would have the quality assured evidence base necessary to advocate effectively on behalf of the inshore fishing sector, within the marine planning process (as listed under Outcome 3 of the 2015 Inshore Fisheries Strategy).

Experience gained through the 2014/15 European Fisheries Fund (EFF) project ‘Evidence Gathering in Support of Sustainable Scottish Fisheries’² indicated that any data solution employed should utilise, wherever possible, open-source software. The University of St Andrews (USTAN) project team invested significant additional resources and staff time in developing, as a proof of concept, an open source standard query language relational database which provided a platform for the analysis of the AIS and contingent vessel and fisheries data. The USTAN team evaluated different open source options against several criteria, such as;

¹ 2015 Inshore Fisheries Strategy can be found here; <https://www2.gov.scot/Topics/marine/Sea-Fisheries/InshoreFisheries/InshoreFisheriesStrategy>

² Final Work Package reports and a summary video can be found here <https://www.masts.ac.uk/research/sustainable-scottish-inshore-fisheries/>

- capacity to handle large amounts of data quickly
- the capability to provide a robust spatial database management system
- the level of online support available, and
- the ease of connection with an open source viewing tool.

This work demonstrated how PostGreSQL along with PostGIS and QGIS, provided a valid computationally efficient platform for the analysis of the AIS and contingent vessel data.

Subject to other suggested modes of operation and delivery by applicants, WP6 was expected to be built upon the preparatory work outlined above to provide an integrated and systematically developed platform for the handling and processing of marine data and, where appropriate, provide a range of predefined outputs required by key user groups.

1.3 Original specification for the work package

1.3.1 Approach

1. Assess the potential relationship, as well as any desirable or required interoperability, between the Inshore Fisheries data resource and existing public and open access data resources held by Government and other relevant parties. Also undertake an assessment of the potential to generate outputs for statutory reporting processes such as, for example, the Data Collection Framework.
2. Based on user group requirements, identify key data parameters/metrics to be collected under WPs 2, 3, 4, and 5, as well as any valuable supplemental data streams which may include, but not be restricted to; meteorological, hydrodynamic and benthic habitat data, social and economic metrics, market information and profiles of fishers behaviour that could be incorporated within the data resource. Potential user group requirements will be defined under WP8B (WP8B Approach 4).
3. Identify potential data storage, sharing and ownership challenges which could limit or prevent the use of data identified under Approach 2 and, based on this information, develop a draft data sharing and management agreement designed to address these challenges.
4. Define data quality checks and data verification, security and access requirements for the various user groups proposed and the scoping of what systematic controls and actions could potentially be built into the process.
5. Develop an open source standard query language (SQL) relational database (or equivalent) which will provide a platform for the analyses of AIS (or equivalent), contingent vessel, fisheries and biological data as well as any relevant model data variables identified under WP8B. Subject to discussion with Marine Scotland, the database developed may need to be compatible, where applicable, with existing (relevant) systems and processes. The platform will need to be easy to use, based on open source software, have good online support and enable the rapid processing of relational queries on standard computers (not large clusters).
6. The data resource will be required to assist Scottish Government in making informed evidence-based policy and legislative decisions regarding inshore fisheries management and marine spatial planning out to 12nm. As such it will be necessary to develop a suite of common database queries, defined through consultation with Marine Scotland, and where applicable other stakeholders.
7. The resource is intended as a first step toward a verified evidence base for use by multiple stakeholders for a range of purposes, over varying timescales. These include; individual fishers, local/regional fisheries management; marine planning and potentially the identification of strategically important responses to social and economic pressures and potentially longer-term impacts of, and adaptation to, climate change.

1.3.2 Outputs and deliverables

1. A fully specified pilot relational, SQL based data resource of agreed specification created (subject to superior suggested options) within PostgreSQL/PostGIS that can incorporate positional data, vessel information, effort data, biological data and be flexible enough to easily accommodate other supplemental data streams identified through consultation with key stakeholders (e.g. Marine Scotland, Inshore Fisheries Groups, Scottish Natural Heritage).
2. A draft data sharing and management agreement.
3. A detailed final report, in a format to be agreed with the SIFIDS Project Management Team, which will include all necessary information describing the function and operation of the database, together with reference to all code developed for the database, quality checks and standard reporting tools.

2 METHODOLOGY

2.1 Modifications to the original specification

The project operated under an agile framework and as such, the specification and associated deliverables changed as the project evolved. The drivers behind the changes can be grouped into 3 main headings; the customer's requirements became clearer, which data sets were actually available to the project and technical issues becoming apparent during development.

2.2 Definition and sourcing of data

To determine which data sets were needed within the application a series of face-to-face meetings were held with and/or discussion documents circulated to the primary users. These included the Regional Inshore Fisheries Groups (RIFG), Marine Scotland Science (MSS), MASTS and individual fishers.

For the pure fisheries science users of the application, the general theme of requirements was as expected; summarised catch, effort, size and sex distributions for both the landed and discard elements of the catch. The aggregation levels preferred varied from trip level up to quarterly summaries and the spatial aggregation preferred tended to be at the management region level, except where specific issues were being explored and ad-hoc aggregations were needed. For users with a fisheries management interest there was added interest in data related to the dynamics of the fleet (e.g. activity drivers, vessel distribution, seasonality of effort etc).

For the RIFG type users it was a similar story but with a requirement for a more detailed spatial structure. Due to the commercial nature of some of the data (effort and catch), the RIFG type users were restricted to seeing anonymised data only. Providing detailed spatial data without aggregation or anonymity protection for the individual contributing vessels was not possible. To meet this need it was agreed to provide data aggregated to 5 nautical mile (nm) cells across the whole project region and for cells that had less than 5 contributing vessels, the data request was returned as 'data unavailable'. This ensured that no single vessel could be identified or that its effort and catch rates could be viewed. There was discussion on whether to merge adjoining 5nm cells until the 5-vessel rule was satisfied but it was felt this would lead to a more confusing visual representation of the data.

Where individuals were more inclined towards marine planning or were taking a more holistic view of usage and activities within the marine region; they required a much larger range of data sets including habitat, substrate, maritime structures (wind farms, oil and gas installations etc). There was also some interest in shore-based infrastructure and socio-economic data.

Feedback forms and dockside discussions suggest individual fishers in general were interested in overall trends but mostly considered that they knew where they fished and what they caught and seeing it confirmed on a web-page was of limited interest. However, what also came across was the potential for the underlying source data for catch and effort to provide a good 3rd party evidence base for track record, both in terms of spatial activity and catch. With this in mind, the fishers were keen to see their own data in the application, mainly to check its accuracy and reassure themselves that the evidence base being accrued was a complete record. There was also a desire to see data that could corroborate the anecdotal evidence that environmental factors (e.g. sea bottom temperature or sea state) can sometimes have a large impact on catchability of crustaceans. If this corroboration was proven and recent data made available then there is potential for fishers to use it to become more efficient, for example, by only putting to sea when forecasted catch rates warranted the operational costs of fishing.

Data sources were identified to provide the data layers required by the users;

- COMPASS (fleet-wide catch and effort data and FISH1 data for the inshore fleet)
- National statistics (fleet-wide vessel information and catch/effort)
- Work Package 8 (SIFIDS fleet vessel surveys and activity drivers)
- Work Package 4 (Socio-Economic data)
- Work Package 2A (SIFIDS fleet detailed track data)
- Work Package 2B (SIFIDS fleet detailed catch and effort data)
- British Geological Survey (Substrate/Habitat)
- Oil and Gas Authority (Sites, pipelines etc)
- Various public resources (Coastlines, MPAs, fishery limits etc)

2.3 Acquisition of the data

Although efforts were made on both sides to either release or link to COMPASS data for inclusion within the application, this proved impossible given the technical and time constraints imposed. It was therefore necessary to create dummy data sets that could be used to demonstrate to users what could be achieved given access to the right types of data.

Due to ethical restrictions and the commercially sensitive nature of the surveys undertaken as part of the socio-economic work package (WP4), the raw data was not available to the application. However, there is the potential to incorporate the published results within the application once they are available. However, at this stage this was not possible as the time-frames of both projects mean that publication of the WP4 report would not happen until after WP6 had completed.

The Oil and Gas Authority data required a license fee. Ideally these additional data sets would have been purchased to fully demonstrate the power of the WP6 application but as there was no budget for purchasing additional data sets, combined with the short project lifetime and the main WP aim being to explore the feasibility of the concept rather than build a full production system, these were not included in the development.

The British Geological Society (BGS) data was supplied for free under an academic licence via the University of St Andrews (USTAN).

The remaining data sets, e.g. Marine Protected Area (MPA) shapefiles, were all acquired free of restrictions, other than acknowledgement of source for some, and these form part of the final application. Table 1 lists the data sets used during the development of the WP6 application.

Other data sets that would have been of interest to users, e.g. sea bottom temperature, were not available at resolutions necessary to make them useful to the parties that expressed an interest and were not pursued or included.

Table 1. Data sources that were used to build the application functionality.

Description	Type	Limitations/Restrictions/Licensing
Coastline	Base Map	None
MPA	Shapefiles	None
Scottish Marine Regions	Shapefiles	None
BGS Habitats	Shapefiles	Academic licence through USTAN
Bathymetry	Shapefiles	None
3,6 and 12nm limits	Shapefiles	None
WP2A vessel tracks	1Hz latitude and longitude	Data from 13 vessels in the SIFIDS programme
WP2B Effort data	Geo/Temporally tagged String and Pot sensor data	Data from 5 vessels in the SIFIDS programme
WP2B Catch data	Geo/Temporally tagged count and size data by species and sex for retained and discarded	Data from 1 vessel in the SIFIDS programme
WP8 Observer data	Trip level activities, creel catch for selected strings, see WP documents for detail of data collected	132 successful trips and 3 cancelled trips from SIFIDS vessels over the life of the programme
Vessel File	Vessel characteristics (length, age, VCU etc.)	Static data, reliant on MMO publications, only for currently registered vessels, no historic data

2.4 Selection of tools

The project specification required use of an Open Source database platform that provided GIS capability and a large support community. As the project would need a significant collaborative input from project partners it was decided very early that a cloud-based approach would provide the most flexible platform, but without the overhead of institution-based infrastructure limitations. As USTAN had previously developed systems using PostgreSQL with good results, this was chosen as the base relational database engine. The final piece of the jigsaw is the toolset for development of the user interface. R software was chosen due to the large array of cutting-edge packages available, especially those for statistics and visualisation. R is completely open source and is widely used for promoting transparency and allowing technology transfer between disciplines and researchers. As an alternative to R, python could also have been a good workable option.

For further information on R, please see: <https://www.r-project.org>

The R version used for the application was: 3.3.1 (2016-06-21) -- "Bug in Your Hair" for platform: x86_64-apple-darwin13.4.0 (64-bit) (R Core Team, 2018)

In addition to the core R package the following additional libraries (see Table 2) are used in the production application.

Table 2. R packages used to develop the SIFIDS prototype application.

Package	Version	Description	Use
shiny	1.0.5	A package used to build the interactive web pages	application
shinythemes	1.1.1	Provides themes used to control application appearance	application
shinyTree	0.2.2	Tree control used in dropdown checkbox menus	application
shinyjs	1	A package support for commonly used JavaScript operations	application
dplyr	0.7.4	A package to assist in data manipulation	data
reshape2	1.4.2	A package to assist in cast and melting dataframes	data
magrittr	1.5	A package that offers a set of operators which make the code more readable by: structuring sequences of data operations left-to-right	data
DT	0.2	A package for rendering sophisticated datatables in HTML using JavaScript	graphics
ggplot2	2.2.1	A data visualisation package	graphics
likert	1.3.5	A package designed to assist in displaying social survey data from questionnaires (i.e. fisher decision making)	graphics
RColorBrewer	1.1-2	A library of colours and palettes	graphics
viridis	0.4.0	A package of colour palettes in R	graphics
leaflet	1.1.0	A library of opensource interactive maps	mapping
sf	0.5-5	A package for plotting data as maps, spatial selection, retrieving coordinates and sub setting.	mapping
maps	3.2.0	A package for displaying maps along with mapproj and mapdata (projection)	mapping
rgdal	1.2-16	Geospatial Data Abstraction Library to assist in spatial projections	mapping
rgeos	0.3-26	A package that implements functionality for the manipulation and querying of spatial geometries using the Geometry Engine	mapping
DBI	0.7	A package that defines the generic database for R	database
RPostgreSQL	0.5-2	A package that provides a Database Interface 'DBI' compliant driver allowing R to access PostgreSQL database systems	database

2.5 Data locations and hosting

As the application needs to have access to a wide variety of data sets, some very large and extremely dynamic, it was designed to allow its data sources to be hosted in several different locations rather than altogether in a single location. This design decision removed the need to have systems in place to replicate or refresh source data into the WP6 system, whilst ensuring the system still had access to the most up-to-date data sets. By leaving the source data with its providers we also remove the need to manage that data (in terms of backups etc.) and reduce the cost and management overheads associated with the data resource. Access control also remains wholly with the providers allowing them to retain full control of the data feed with the option of modifying or removing data streams to meet their obligations (legal

or ethical) to their data subjects. This may prove beneficial when negotiating data sharing agreements as it reduces the risk to the data subjects and data provider.

Adding additional data layers or sources is simplified as the development effort can be mostly directed towards the user interaction and visualisation of the data, rather than the analysis and design of data structures as this will have been done by the data providers, who will understand their data and its dependencies more clearly.

However, there are drawbacks to this design:

- the system is reliant on the source being available and responding within a reasonable timeframe when it requests data, otherwise the user experience is downgraded
- the security envelopes around the WP6 host and the data host need to facilitate the inter-connection
- a good working relationship between the data host and WP6 management team is essential to mitigate potential issues when changes (to structure, access routes etc) occur on the data host.

The WP6 application does have its own core database which holds the system reference tables and data layers that could not be retrieved on-the-fly from their sources.

The feasibility of this design is proven within the final system as the WP2 and WP8 data sources are hosted on a different server to WP6 and accessed on-the-fly.

2.6 Data users, data anonymisation and aggregation levels

When specifying user classes for the pilot system we identified the following:

- Individual fishers contributing the fine-scale vessel data
- People working under or for the parent organisations (SeaScope, USTAN, Marine Scotland).
- Chairs or authorised members of Regional Inshore Fisheries Groups (RIFG)
- Full public user class (for the purposes of the trial this was not considered further but would likely be included in a full-scale roll out).

For each data layer in the system we then identified what level of access was required for each user type:

- All the supplementary data layers (habitat, substrate, MPAs etc) would be visible to all users
- The publicly available fleet data would be available to all users
- The results of WP8 and WP4 would be available to all users
- Reporting at RIFG level would be available to RIFG and parent organisation users
- For the system to protect the commercially confidential fine scale vessel data it was necessary to implement access, aggregation and anonymisation rules (Table 3):
 - An individual fisher would be able to see their own data in its raw form but no other individual vessel data in any form. Vessel data ownership is granted to a given user for a period of time. This time restricted access is necessary to allow for different skippers or owners to switch backwards and forwards between vessels. This would require careful management and access to up to date vessel ownership details on an almost daily basis.
 - RIFG users would only be able to see data aggregated to cells of 5nm squares and only if more than 5 distinct vessels contributed to the value within this cell.
 - Users from the parent organisations would have unrestricted access to the raw data.

Table 3. Examples of access rights for the application.

User type	Access type	Map data page	Catch effort & page	RIFG reporting page
Fishers	restricted	Own	Own	None
Marine Scotland & St-Andrews	admin	Any	Yes	Yes
RIFG chair	IFG	None	Yes	Yes

2.7 Development process

The core database was created on a Postgres Relational Data Management System (RDBMS) hosted through Amazon Web Services. The core database is dedicated to holding static information collected from various 3rd parties and system references. As spatial information was included, a Postgres extension (PostGis) needed to be included to support geographical objects thus allowing use of SQL spatial queries. The web-page side of the application was hosted on an Amazon Elastic Compute Cloud (EC2) instance.

Some of the development of the WP6 system had knock-on effects to the Vessel Tracking (WP2) system as we needed to modify the data streams to make them more appropriate to the WP6 system or to improve response times when users were accessing data etc. This was only possible as the WP2 systems were in development within the overall SIFIDS project at the same time, this would have been more difficult or impossible to achieve for 3rd party data sources or those that were fully rolled out and in maintenance rather than full development. Although the changes were made in another work package they are detailed here as they highlight important learning that is applicable should there be follow-on development work generated from this project.

The WP2 system logs vessel positional data at a maximum of 1 ‘ping’ per second. It was apparent that users found selecting and visualising the data at a ‘trip’ level far more intuitive than selecting from a constant stream of data. This aggregation of data into real physical trip units becomes more important as users start to explore catch and effort data. Therefore, the WP2 system was modified to group the pings into ‘data sessions’, with the system starting a new session each time the vessel powered up, vessel powering up was used as a proxy for trip start and the trip ended when power was switched off. This method of trip allocation does lead to spurious fishing trips when, for example, the vessel remains in port for maintenance and is started and stopped several times or if a vessel should lose power whilst at sea. However, with appropriate geo-fencing, ping count rules and activity analysis, the spurious trips could be excluded from detailed analysis.

The application of the data ownership hierarchy and provision of data from the tracking system was handled within the tracking system database. This follows the design pattern described earlier where the data source retains control of how data is supplied. Again, this required some work on the part of the WP2 team. The data from the WP2 vessel tracking system is mainly supplied as responses to calls to stored procedures, where the procedures require the ID of the WP6 application user as well as the data selection criteria. The stored procedures ensure that data supplied back to the WP6 application is only that which is appropriate for the ID supplied and an error or omission in the WP6 code could not result in inadvertent disclosure of data.

In order to facilitate the RIFG aggregation to 5nm cells, further work was required from the WP2 system which involved the creation of the matrix of 5nm² cells and the allocation of the vessel's positional pings to their containing cell. This was implemented to improve WP6 response times because the on-the-fly allocation and subsequent aggregation of hundreds of thousands of pings to cells when queried by a user was resource hungry and resulted in poor response times to data requests. The cells account for the curvature of the earth in their 5nm size limits and therefore do not appear square when displayed in the preferred projection on the maps within the application.

2.8 Performance issues and problems

2.8.1 Data volume and server size

The WP6 development team encountered a number of issues with the rendering of GIS track data (WP2) onto the maps (map data tab). These were mainly as a result of the increase in the number of vessel trips and hence the data volume increased resulting from the high vessel sampling frequency of 1 ping per second (1Hz). The resulting increases greatly impacted the server performance (server size AWS micro, see above) when the database was queried for multiple vessel tracks and date ranges. The user, via the map area and the date range menu specifies the queries; for example, the user selects a date range together with the geographical range adjusted by the screen size map zoom function. If the user selected a large map area (i.e. zoomed out) and a large date range (over many months), the server struggled to select the data in an acceptable timeframe and the rendering of the many tracks caused the application to 'hang' and thus a re-boot of the application was needed. To mitigate this performance issue, the data selection method was altered to a 2-step procedure as follows:

1. The user selects the area and date range as before, however the query returns a list of unique sessionids and associated vesselids (anonymised) within a small table. This allows the user to view and select which vessels and on which specific dates, they would like to see track data for.
2. Once a specific vesselid and date combination are selected a query is triggered to get the individual track data from within another table and thus the user can select or deselect tracks each time from the sessionids table.

2.8.2 Points versus Polyline

Initial development used individual vessel ping data (1 per second) displayed as a series of points to make up its tracks. Utilising this method however, it was found that displaying all this information graphically caused performance issues. Initial development in R used many random normal sampling processes to sample enough data and display the tracks without too many missing data points. However, in some portions of the displayed tracks, resolution was poor, as was the rendering of the tracks. Other approaches were explored such as filtering by vessel speed within certain parameters, along with random sampling of the data. If vessels were found to be travelling at excessive speeds e.g. 30 knots, then erroneous positional data may have occurred. But still performance issues were encountered, and the process was further researched.

The WP6 development team then researched using geographical approaches within the R framework. The R framework supports many cutting-edge spatial manipulation packages such as the sf package. The sf package contains a function known as the **sf::st_line_sample** which works by sampling points on a linear geometry and 'converting' to a geometric polyline rather than displaying individual data points. The approach was considered with the user able to use slider controls on the interface to set the number of points to sample to simplify the

track by adjusting the slider density parameter. In other words, instead of seeing every ping (1/sec), the user could choose to only see the 10th or 60th etc. The modification of the individual data points was tested as geometric points versus a geometric polyline (described above) and the rendering speed assessed. It was subsequently found that the conversion of data points to a geometric polyline rendered much quicker than the geometric point data.

An advantage with respect to a "random sampling" approach was that the **st_line_sample** function allowed a more reasonable subsampling of the equally spaced point data per linear geometry, i.e. if the user specifies a sample of 1/1000, then this equates to 1 point per kilometre. A small caveat was the need to go back and forth between the 4326 lat/lon projection needed by leaflet (R package), and a metric projection needed for **st_line_sample** to work. This approach was effective for several months in the initial development but then the issue of rendering the data reoccurred when data volumes increased and subsequently it was decided that the whole process should be performed on the server. A stored procedure was created so that the incoming data was converted directly to a geometric line and stored as a column in the data table (this is described more fully in the section "stored procedures").

2.8.3 Map displays and colours

Many different map displays (from within the R leaflet package) were tested in order to determine how well they rendered live within the application when zooming in/out to a specific area for the map data page of the application. For example, Hydda.Full and Esri.OceanBasemap base maps (from leaflet) were tested in earlier application designs, but it was found that when displaying on Hydda.Full maps the tiles were occasionally very blocky and when zooming in some tiles did not appear at all. The most stable background to show the detailed tracks was found to be OpenStreetMap.Mapnik. When the user moved the geographical search area or zoomed in or out, the display performance remained stable, visible and unaffected.

In order to display the data on the pre-described basemap (map data page), several colour palettes with enough contrasting colours were tested and randomly sampled in order to show which combinations worked best when displaying several tracks at once. However, the potential number of tracks between specific date ranges had to be considered when choosing the palettes in order to ensure enough contrasting colours were available to be displayed and to differentiate between trips and vessels. The R package 'viridis' and 'RcolorBrewer' were the perfect solution to this issue due to the number of differentiating colour palettes they provided.

The database schema and R development code is available from USTAN. The code is written in a simplistic way without nested functions etc. to enable a third party with a basic understanding of R, relational databases, SQL and the shiny packages to follow, understand and adapt to their needs. A production system based on this pilot system would need to streamline both the database schema and R code to meet the project needs and the infrastructure supporting it.

Although both Postgres and R are open source and should theoretically provide platforms from which the solution can be deployed to meet project needs, this did not prove straightforward. The initial aim was to develop the system as a cloud-based collaboration and then move it across to servers within the USTAN infrastructure. However, due to security concerns, version control conflicts and library conflicts, it was not possible to implement the necessary R-studio and shiny packages at USTAN. With an isolated instance such as the Elastic Compute Cloud (EC2) development platform there is no requirement to mitigate security, version control and shared libraries that are a factor when hosting on a shared resource.

2.9 Project replication

Replication of this project can be done by setting up an instance on Elastic Compute Cloud (EC2) to host shiny and shiny server on amazon free tier.

For development and testing of the software it was essential to set up a live, globally available instance. To do this the development team used Amazon Web Services (AWS) free tier (t2 micro) and the Amazon Elastic Compute Cloud (Amazon EC2) running under Ubuntu Server 16.04 LTS

Details on running instances of this type can be found at the link below: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EC2_GetStarted.html#EC2_Launch_Instance_Linux_

Once a server is set up the developer needs to install RStudio Server, Shiny Server and the library dependencies. For details on these products see: <https://www.rstudio.com/>

3 RESULTS

3.1 Database

The WP6 core database is, as discussed previously, mostly a repository for static data. Documenting the detailed table design within this report would add very little, so has been omitted, however the structure is available as a buildable SQL schema from USTAN.

3.2 The Application

The application is available through a web portal and can be implemented across a wide range of stakeholders. It can enforce data-set specific aggregation and availability rules based on individual users or their roles within the system (as shown in Table 3). The user interface is a series of page tabs each displaying data in the most appropriate form for the data. Data sets can be downloaded for local analysis, subject to specific access rules.

When a user accesses the home page of the site they are asked to login. Once successfully logged in they are presented with the site introduction page (Figure 1):

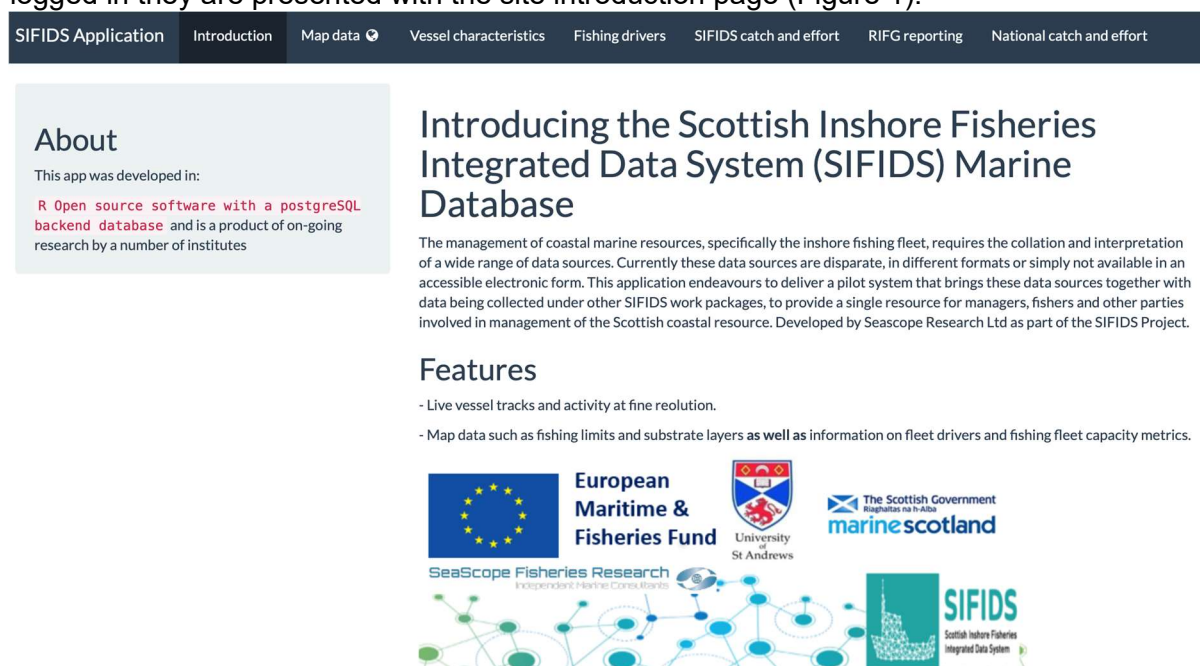


Figure 1. The introduction page of the SIFIDS Marine Database Application.

The site is broken into 7 tabs which can be accessed from the page banner:

1. **Introduction:** The introduction page for the application.
2. **Map data:** This is an interactive fishing activity map, which contains information on the study fleet in terms of their fishing tracks and activity. The user can look at vessel tracks and their activity e.g. pot deployment/string hauls etc. This is currently limited to the 13 vessels with the WP2 tracking system on board, of which 5 vessels also have the fishing effort monitoring sensors installed. There are also static layers which can be overlaid e.g. habitat, MPAs and 3,6 and 12 nautical mile limits.
3. **Vessel characteristics:** This page shows vessel information such as tonnage, engine power and length broken down by Regional Inshore Fisheries Groups (RIFG) areas.
4. **Fishing drivers:** This page contains information that was collected from fishers by observers during at sea surveys, with respect to the drivers (reasons) that influence the fishers decision making in regard to when and where they go fishing. Further details can be found here (Mendo *et al*, 2019a).
5. **Catch and effort:** The catch and effort page provides a map and tools to interrogate specific fishing areas to provide the user with catch (numbers of lobster/crab) and effort (number of pots) summaries as a heatmap. This can be summarised by month, quarter or year.
6. **RIFG reporting:** The RIFG reporting page provides tools to interrogate specific Scottish Marine Regions catch, as reported on FISH1 forms, (numbers of lobster/crab) and effort (number of pots) summaries by month, quarter or year and output to a csv file.
7. **National catch and effort:** The national catch and effort page would consist of all voluntary and potentially mandatory data submitted by fishers. This would provide a map and tools to interrogate specific fishing areas to provide catch (numbers of lobster/crab or any marine capture fish/shellfish) and effort (number of days at sea) summaries as a heatmap by month, quarter or year. Unfortunately, the real data sets were unavailable to the project team and therefore this page was designed using fabricated demonstration data only.

3.2.1 Introduction page

This page (Figure 1) provides a brief description of the project and the main features users can expect to find in the page tabs. It also lists the partner organisations and the funding bodies involved and details the open source software used to develop the application.

3.2.2 Interactive fishing activity map (Map data tab)

Overview: The interactive fishing activity map contains information on the WP2 study fleet in terms of their fishing tracks. The user can look at vessel tracks and their fishing activity e.g. pot deployment/string hauls etc. This version of the activity map only shows this level of detail based on data derived from vessels fitted with the WP2 gear sensors. The level of detail shown will be determined by the access permission granted to the user. There are also static layers that can be overlaid e.g. habitat, MPAs and 3,6 and 12 nautical mile limits. The yellow spheres shown in each port is related to the number of vessels surveyed by the observers during WP8.

The Interactive map contains 2 main parts (see Figure 2):

- Menu panel (on the left-hand-side) named as “Fishing activity control panel”
- Main panel (on the right-hand-side) which contains the map.

It should be noted that the map page is designed to always open in the same geographical area (currently Arbroath). The user can drag the map to their preferred location of interest or zoom the map out to cover a wider area. However, the user should be aware that making the geographical area too large will have a negative impact on the webpage response time due to the increase in time required to download the larger data set.

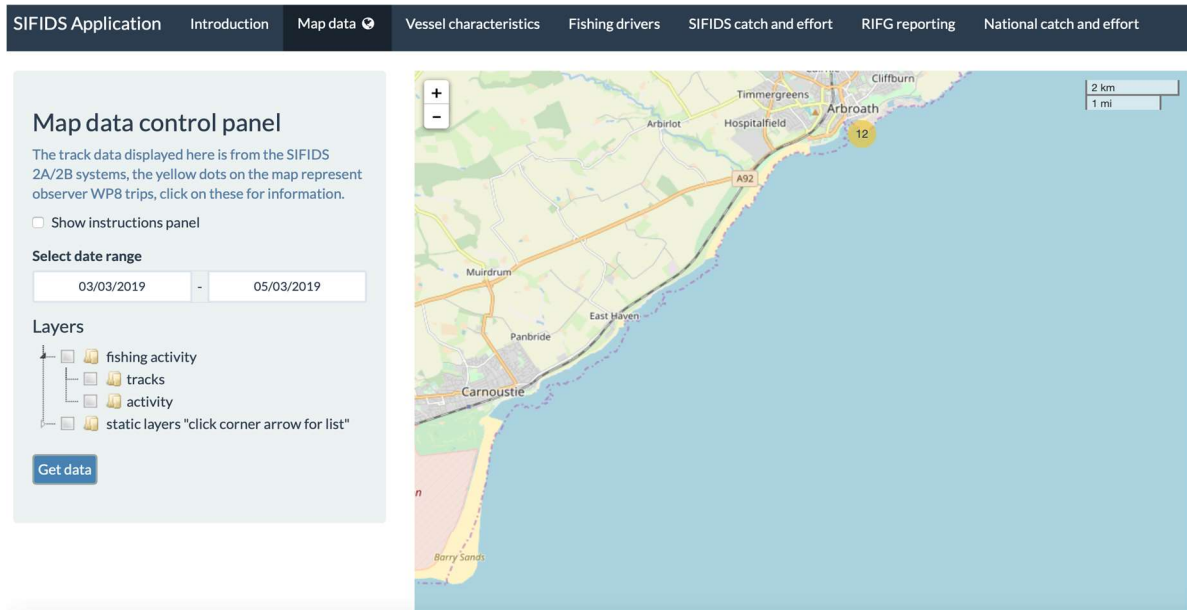


Figure 2. Interactive fishing activity map.

Fishing tracks

Figure 3 shows the control panel for the interactive map. This allows the user to select the type of data and the date range required.

Map data control panel

The track data displayed here is from the SIFIDS 2A/2B systems, the yellow dots on the map represent observer WP8 trips, click on these for information.

☐ Show instructions panel

Select date range

03/03/2019 - 05/03/2019

Layers

- ☐ fishing activity
- ☐ tracks
- ☐ activity
- ☐ static layers "click corner arrow for list"

Get data

Figure 3. The control panel for the interactive map.

The main instructions for using the map control panel are:

- Select date range for fishing activity
- Move map to desired area of interest and zoom in to the desired location by screen grabbing with the right click button of the mouse and holding this down whilst dragging. This stage is important as this procedure queries the database and returns information based on the area of the map shown.
- Select the 'tracks' checkbox from the tree folders by clicking within the clear square and a green tick will appear (Figure 4). You must select points before you can view activity i.e. pot deployments.
- Hit the 'Get track' button when ready, this captures the time and space query and returns trip data.

Map data control panel

☐ Show instructions panel

Select date range

07/07/2018 - 09/07/2018

Layers

- ☒ fishing activity
 - ☒ tracks
 - ☐ activity
- ☐ static layers "click corner arrow for list"

Show 10 entries Search:

vesselid	starttimestamp	trackpoints	track
L	2018-07-09 05:44:08	34544	true
I	2018-07-09 19:59:07	28396	true
C	2018-07-09 04:57:25	23137	true

Figure 4. Available vessel sessions table. Clicking on the arrows next to the VesselID column will sort by this criterion, but only one criterion can be selected for sorting at a time.

A vessel sessions table will appear below the button “Get data”, this shows the data available for the map area and the date range specified. The table shows the start date-time for the trip, the number of data points that make up the track and whether the current pre-built track on the server is up-to-date. If a track is out-of-date (i.e. new data has been received) it will be rebuilt before being displayed which takes a few seconds.

The trackpoints value gives an indication of the size of the track, lower numbers may be due to it being a ‘live’ track with the vessel still at sea or in the case of historic tracks may indicate vessel trials in port, maintenance etc (see page 16 for details of ‘trip’ definition from WP2).

The user can select/deselect vessels and trip data from the table by clicking on rows in the trip table (Figure 5), each selected trip will be displayed on the map.

If a user decides to move to a new map area or change date range, they need to hit the “Get data” button again to refresh the vessel sessions available.

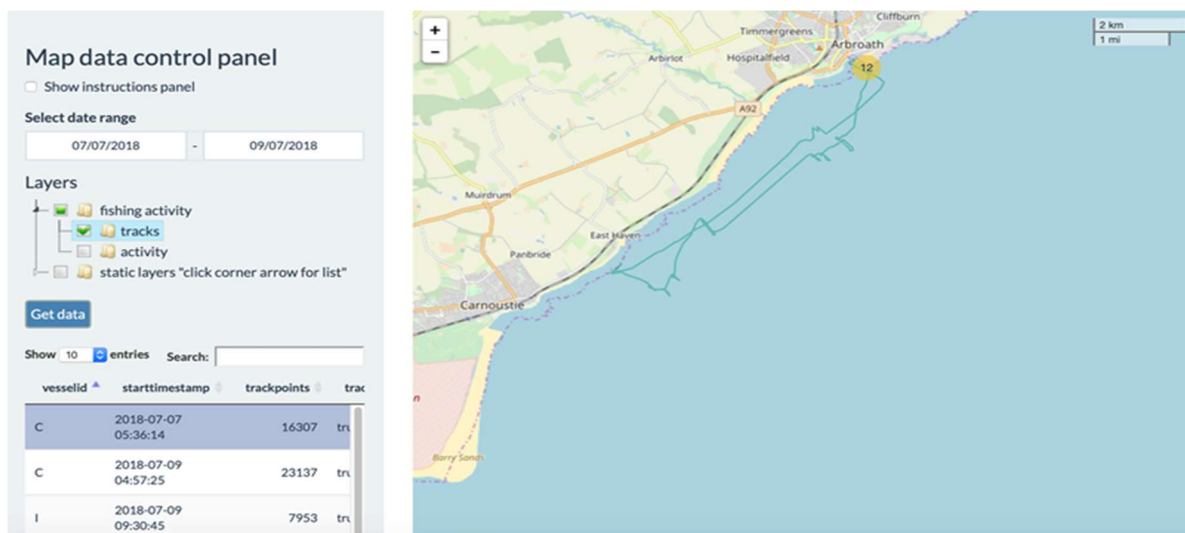


Figure 5. The track produced (green line) when a row of vessel data is selected.

Fishing activity

Overview: Fishing activity is the activity of the fisher on a trip, i.e. it shows the fishing effort deployed by the vessel on a trip. Selecting the activity checkbox, as shown below, will enable the user to select pot, haul or shoot from a dropdown menu displayed at the bottom of the Map data control panel (bottom left of figure 5). Once an activity type is selected, the data points are plotted for the tracks currently displayed in the map panel. The example in Figure 6 shows the vessel's track for the day and the pots (creels) as they were hauled and detected by the on-board sensor, as well as the substrate type from the habitat static layer.

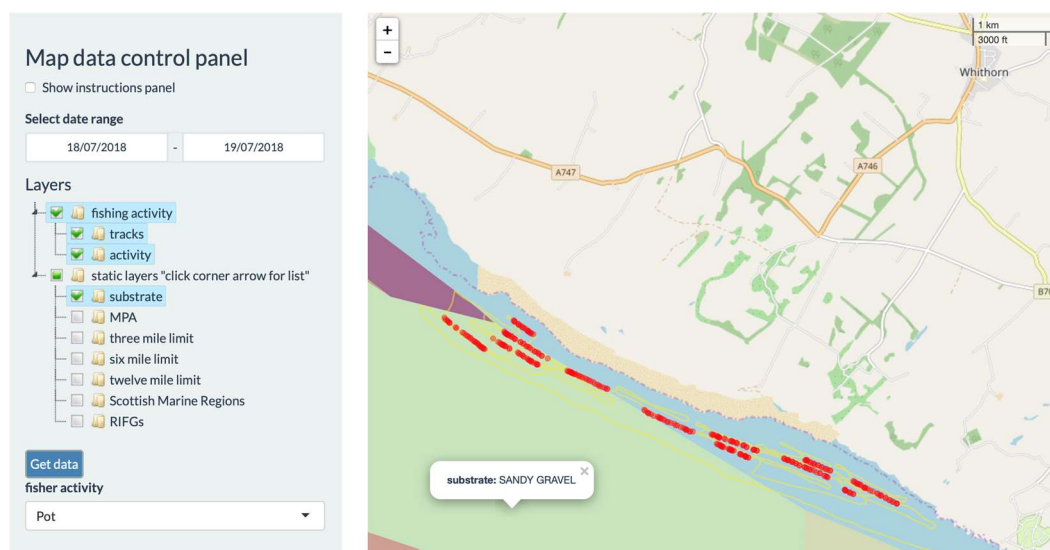


Figure 6. An example showing the fisher activity in the form of a day's track (green line) and the hauled pots (red circles) detected during the trip.

3.2.3 Static layers

Overview: The static layers options incorporate mostly unchanging data as opposed to the previously described layers above that update with fishing trip information over time. By

clicking the corner arrow, a drop-down list of layers will appear. If the main parent checkbox labelled “static layers” is selected, then all of the various layers will be shown. However, this may cause performance issues due to the volume of data being displayed and the potential of masking information through conflicting colour selections. The user can select which layers are most appropriate for their needs by switching all of the layers off or on at the parent level (top level) or choose individual layers to show using the child level checkbox menu (below parent level). Figure 7 below also shows that labelling within the layers is possible by clicking on any of the added polygon layers (RIFG, MPA, Substrate and Scottish Marine Regions) to display the layer information.

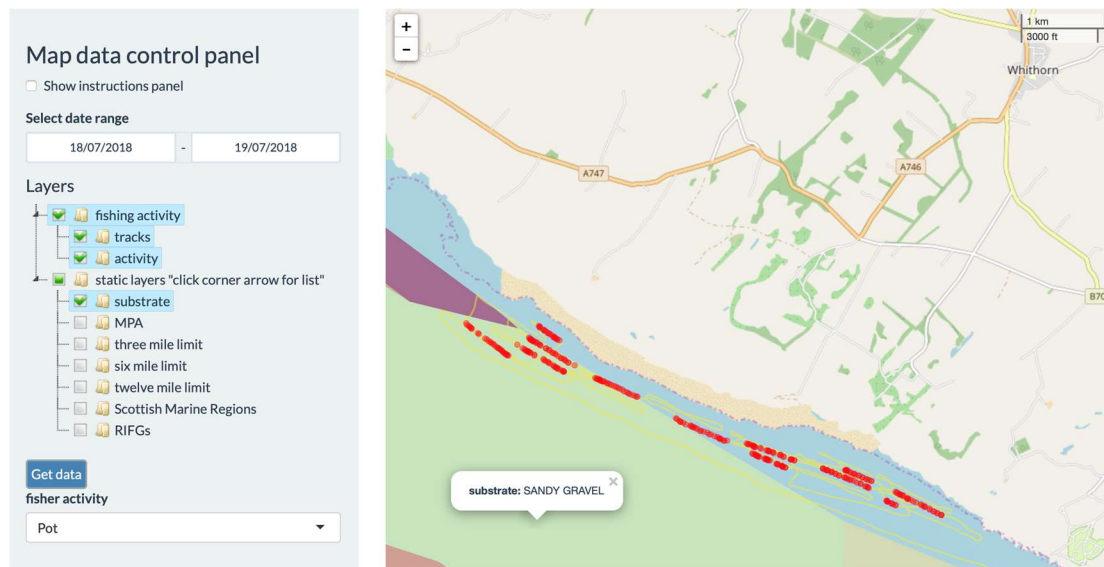


Figure 7. Selection and labelling of layers illustrated by using "substrate" as an example. Also shown is a fishing trip track and the associated effort (pots detected).

3.2.4 Vessel characteristics

Overview: Vessel data was acquired from the publicly available Marine Management Organisation (MMO) fleet data and contains information for the <12m vessels for all Scottish RIFG areas in 2017 (a total of 878³ vessels). The vessel characteristics tab contains 2 main parts (see Figure 8):

- The menu panel (on the left-hand-side) that allows RIFG area and vessel characteristics to be selected, and the
- main panel (on the right-hand-side) that produces boxplots of the vessel characteristics by homeport within an RIFG (Regional Inshore Fisheries Group) area.

This information is useful as it allows fishery managers to see how a local fleet of vessels is made up by providing data on such factors as the vessel age, overall length, horsepower, tonnage or vessel capacity units (VCUs) of the local fleet.

³ This total number of vessels appears low compared to other fleet estimates (see Scottish Sea Fisheries Statistics 2017. <https://www.gov.scot/publications/scottish-sea-fisheries-statistics-2017/pages/4/>), but as this was the only publicly available dataset at the time the project was active, it was deemed suitable to test the prototypic WP6 system

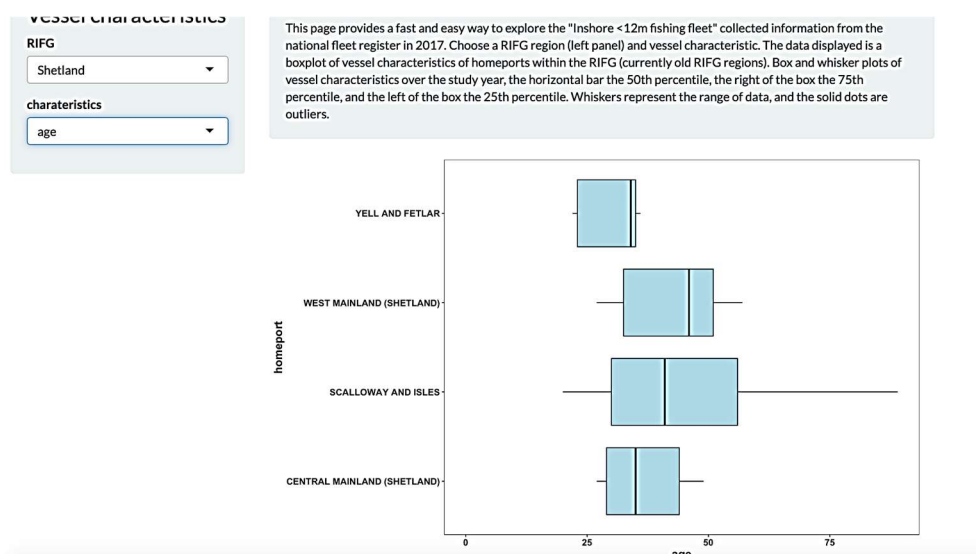


Figure 8. Vessel characteristics table for an example area and characteristic. Clicking the dropdown arrows will subset the data and display new plots.

3.2.5 Fishing drivers

Overview: The fishing drivers' data was collected by observers interviewing skippers during sea trips to try and identify their decision-making processes regarding when and where to fish. These are stored on the WP8 database and are supplied to the WP6 application as a static layer.

The fishing drivers tab contains 2 main parts (see Figure 9):

- Menu panel (on the left-hand-side). The dropdown menu has 2 options as to what factors affect fishers from 'stopping' or 'placing' their fishing gears.
- Main panel (on the right-hand-side) that produces Likert plots. Likert scales are commonly used in survey research to record user responses to a statement and in this case provide a fast and easy way to explore the "Understanding of fishing drivers" question.

Figure 9 shows the fishing drivers associated with vessels choosing to stop fishing. The scale used to evaluate each factor influencing behaviour is based on an ordinal scale of 1 to 5, where 5 is most likely and 1 least likely. Associated percentages are also displayed for each factor, which represent the % of respondents who chose a particular reason for stopping fishing. The right plot in the main panel displays the number of responses completed/missing for each factor.

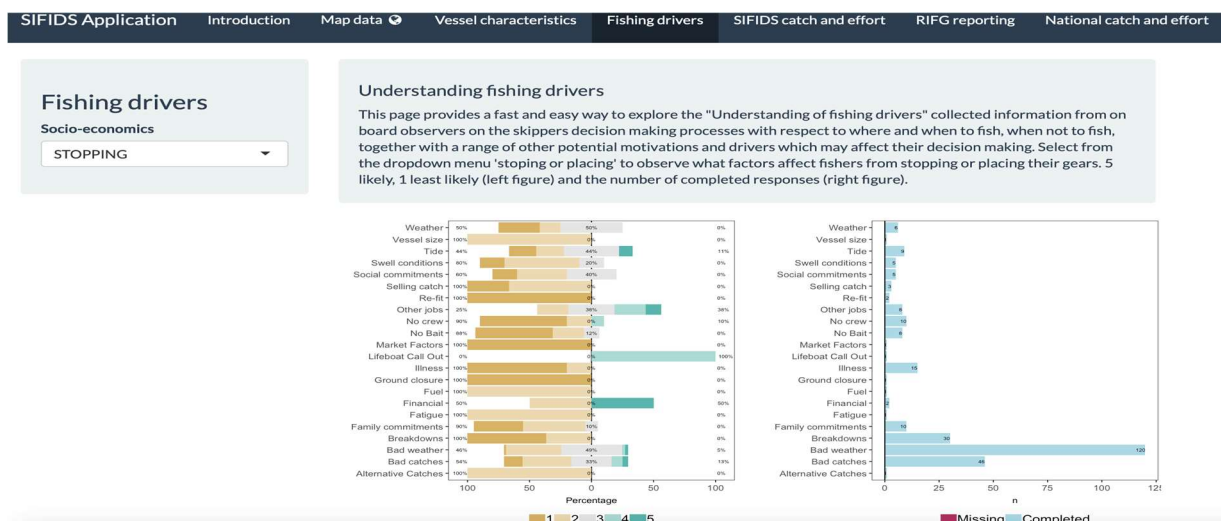


Figure 8. An example of the type of survey data that can be displayed using the WP6 application. This shows the responses provided by vessels as to why they usually stopped fishing, as collected during the WP8a surveys.

3.2.6 SIFIDS catch and effort

Overview: These data were collected during the WP2 trials and are an aggregation of the vessel track/activity data found in section 2 of this document. Effort: Pots = number of pots or creels detected during the hauling process and Catch: the aggregated catch of either lobster or edible crabs by month, quarter or year. The data are displayed depending on a user's permission level.

The SIFIDS catch and effort tab contains 2 main parts (see Figure 10):

- The Menu panel (on the left-hand-side) that allows the user to select the time period and type of data that is required, and.
- The Main panel (on the right-hand-side) that produces a heatmap of the numbers of caught lobsters or crabs as well as the number of pots. This side also allows the user to select the geographical area of interest by dragging and zooming the map area.

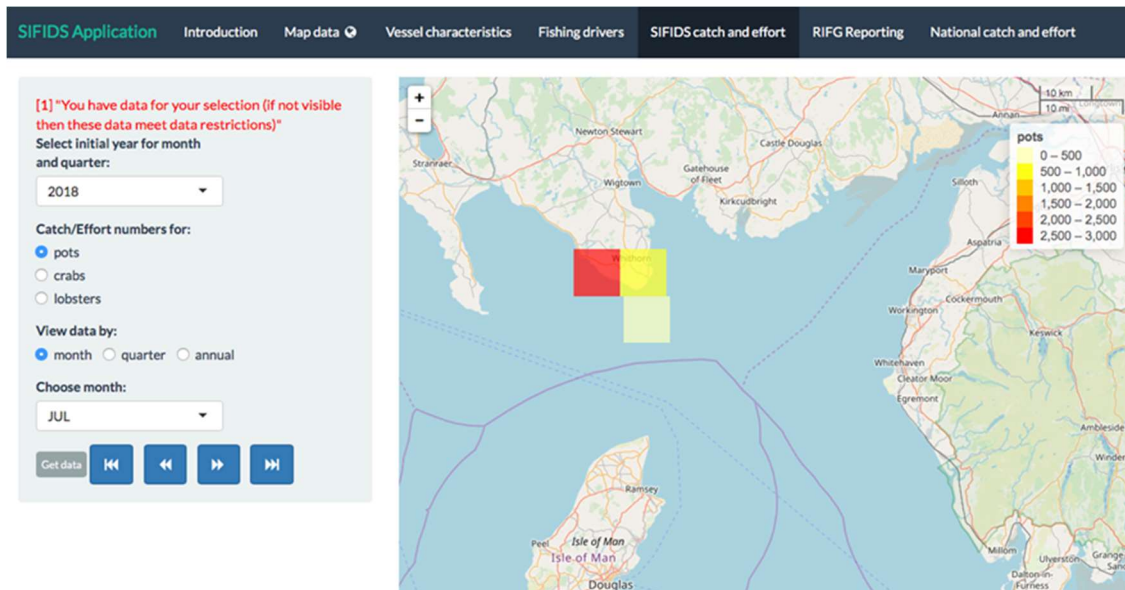


Figure 10. An example of the 5nm square heatmap displays that can be requested. This figure shows the pot count for the WP2 vessel operating in the Solway Firth in July 2018.

To select data the user should:

- Select the year in the dropdown menu (as shown in Figure 11).
- Move the map to the desired geographical area (main panel) and zoom in to the desired location by screen grabbing with right click of the mouse and holding while dragging (this step is important to limit the data set returned).
- Select either crabs/lobsters or pots from the radio button underneath the year dropdown menu
- Choose how you'd like to see the data displayed i.e. monthly, quarterly or annually from the radio buttons and select preferred period from the dropdown menu
- Press the "Get data" button and wait for data to load.

Once these steps have been undertaken a message will be displayed at the top of the menu panel informing whether or not data is being collated for display or not. The user is also able to scroll through months using the arrow keys but must always hit the "Get data" button afterwards or when moving to different map areas.

[1] "You have data for your selection (if not visible then these data meet data restrictions)"

Select initial year for month and quarter:

2018 ▲

2017

2018

☐ crabs

☐ lobsters

View data by:

☒ month ☐ quarter ☐ annual

Choose month:

JUL ▼

Get data

⏮ ⏪ ⏩ ⏭

Figure 11. Menu panel for selecting data for displaying on the heatmaps.

3.2.7 RIFG reporting

Overview: The data available for the RIFG chairs to view comes from the WP2 on board systems and is aggregated by RIFG.

The RIFG reporting tab contains 2 main parts (see Figure 12):

- Menu panel (on the left-hand-side).
- Main panel (on the right-hand-side) that produces a table of the selected aggregations.

The menu bar is exactly the same as described in the SIFIDS catch and effort overview above, and as shown in Figure 11. The only difference is months, quarters and years can be selected in multiples as displayed in Figure 12. The resulting table in the main panel then is downloaded as a csv by hitting the download button so that RIFG chairs can use the data as they wish. Note that scrolling through time periods is not an available option on this page. Also note that this page can only be viewed by RIFG chairs with the appropriate permissions.

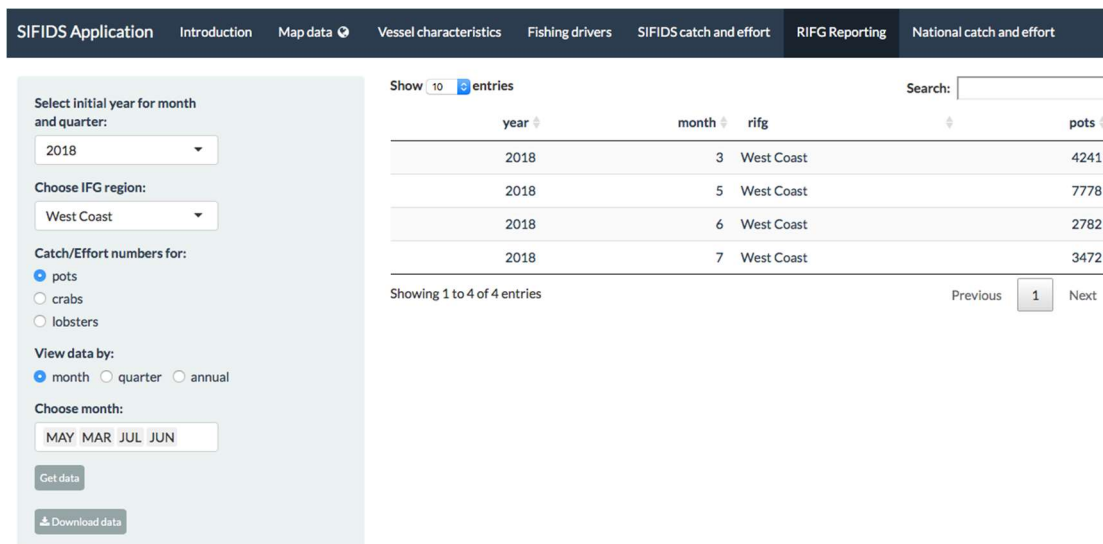


Figure 12. The RIFG reporting page that allows data to be aggregated by time period and Scottish Marine Region (SMR).

3.2.8 National catch and effort

Overview: This page is a duplicate of the SIFIDS catch and effort page described earlier except that all data used is pseudo data which was invented to allow this function to be demonstrated on the application. It has been designed to allow fleets and vessels to volunteer data that could be included in the application in any future designs of the application.

Effort: Pots = number of pots or creels detected during the hauling process and Catch: the aggregated catch of either lobster or edible crabs by month, quarter or year. The data are displayed depending on a user's permission level. Please refer to Figure 10 and the section on "SIFIDS catch and effort data" for more information.

4 DISCUSSION

4.1 Caveats and assumptions

This report does not attempt to address issues regarding best practice for managing or operating a WP6 type tool or application e.g. data security, business continuity, user and data management etc. It is assumed that the tool would be hosted within a well-founded data management environment providing all the necessary measures, tools and processes required for the application and its associated internal databases to meet the requirements of the host organisation, relevant legislation, licencing conditions and the project aims.

4.2 Data sets, acquisition and hosting

Although significant effort was expended trying to determine which tools, visualisations and data sets were required to fulfil potential user needs, this did not result in the level of cross-discipline interest that we originally expected. Giving users free rein to describe what they wanted in terms of data to incorporate into their work or decision making may not have been the most productive route. The responses tended towards answering 'what I do now' rather than 'what I could do' with the resource in the future. Perhaps a theme-led discussion would have prompted more lateral thinking, for example, how could data on bathymetry and substrate be incorporated into your work with the data on fishing activity? Unfortunately, this approach of fact finding would have required a more detailed and personal discussion of data

needs and would have impacted on the associated resourcing for this element of the project. However, this would be a recommendation for any future project trying to achieve the same aims.

The data for a tool of this type can be classed in 3 ways in terms of its dynamic nature:

Static – e.g. coastline shapefiles: These data sets are unlikely to change in any significant way over the life of the project. To minimise application to database connections and ensure ongoing availability they can (licence terms and storage permitting) be hosted within the internal tool database.

Periodically updated or expanded – e.g. quarterly national catch statistics: These are live data sets with a low update rate which can sit equally well as internally hosted or externally sourced on-demand data sets. If internally hosted, then appropriate processes must be in place to not only receive new data items but to also incorporate updates to existing records.

Dependent on the project need for audit trails or replication of queries using data that was available at an earlier time, snapshots or replication processes may be required to allow a local copy of the data to be retained and synchronised with its source. This is a question that will need to be addressed for each data layer of this nature and appropriate storage/access and synchronisation processes put in place.

Very dynamic to live data – e.g. vessel location: These data are in a constant state of flux. If the data is originated by the parent project then it may be appropriate for it to reside within the internal databases, providing its update route does not cause major conflict with another project aim (e.g. within a virtual private network (VPN)). If the data is from an external source, then local hosting with low latency replication/synchronisation is possible but would require appropriate infrastructure and processes to support it. If possible, it would be preferable to source data of this nature directly on demand, but this would be affected by the amount of processing of the raw data that is required for it to be used within the tool. A data source organisation may not be happy if the application is using large amounts of their processor cycles because it is submitting complex aggregation queries on a regular basis. In this scenario, a set of views specific to the WP6 tool on the source database may help.

Volumes of data: The other factor to consider when assessing the data is its pure volume. It may seem a straightforward decision i.e. we have a wholly static data set that we can host locally so that's what we do, but it is 7500Gb. Storage is 'cheap', but costs don't stop at the storage. The data will have to be managed and backed up and those backups managed. If the data source is reliable (low/zero downtime) and there are no bandwidth or speed issues retrieving it, then it may be more sensible to source it directly on-demand.

Since development finished with respect to WP6, research at USTAN has progressed in order to identify the optimal level of data that is necessary to collect between pings in order to characterise fishing activity. For example, small vessels of less than 9.5m in length require < 60 second intervals while vessels (med-large) >9.5m require < 120 second intervals at a resolution of 100x100metres (see Mendo *et al.*, 2019b). Bearing these important findings in mind developers can estimate what storage volume is necessary in terms of a wider roll out for the entire Scottish fishing industry especially if the average trip contains between 20,000-30,000 pings,

Implications for wider roll-out: For each data layer identified, an assessment must be made as to the most appropriate hosting and access solution for the data, bearing in mind the nature and volume of the data alongside the sharing or licence agreement that is associated with the provision of the data. This may well require high level negotiation between the project leads and the source organisation but must also include the technical teams who will have to implement the solution. Even if the executive teams agree to make something happen, a significant system incompatibility could cause delays or compromises on both sides, or in a worst-case scenario prevent the data being available.

4.3 Data security

The security envelope that encompasses the core WP6 application and its source databases is a major driver in its availability as a tool and the data it can access. For example, the WP2 data (vessel location) is provided by the vessel systems using standard 4G mobile data and as a result each data connection is likely to use a different IP address, as is the nature of mobile network connections, when communicating data. If the system were housed inside a VPN with IP address restriction the WP2 system (in its current form) would be unable to submit its data and each user of the system would need the appropriate software and access rights to enable access.

Implications for wider roll-out: The security envelope will largely be driven by the requirements of the hosting organisation and the various data providers. However, these requirements may also impact on the tool's ability to deliver the desired products to the desired user base. As such, compromises may have to be made, or a custom host platform configured, and additional costs factored into any future project budgets.

4.4 Data sharing and ownership

In general, we can classify the data within a WP6 type tool into 1 of 3 'ownership and usage' models:

3rd party licensed data (e.g. the British Geological Survey substrate layers): These data items are governed by the licensing agreement entered into when the data was provided. They may come with restrictions on onward transmission of the data, as a raw download rather than being displayed as a layer on a map or require specific citation on the pages using the data. Where a WP6 tool is using such data, it is likely that there are levels of licensing available from 'internal use only' through to 'full public access' to the data products (not necessarily the raw data). Depending on the implementation model selected, an appropriate licence would need to be obtained.

Freely available data (e.g. 3 and 6nm limits; fleet dynamics data): Although the term suggests that the data can be used for any purpose, some freely available data sets still come with licence conditions. Although this may be as simple as a citation, it may also include pointing users to the original source to provide data downloads rather than providing directly from the tool. As an example; the Open Government Licence (OGL) covering the fleet data requires the user to: -

- acknowledge the source of the Information by including any attribution statement specified by the Information Provider(s) and, where possible, provide a link to this licence;

If the Information Provider does not provide a specific attribution statement, or if you are using Information from several Information Providers and multiple attributions are not practical in your product or application, you may consider using the following: 'Contains public sector information licensed under the Open Government Licence v1.0.'

- ensure that the Information is not used in a way that suggests any official status, or that the Information Provider endorses you or your use of the Information in any way;
- ensure that others are not misled or that the Information or its source are not misrepresented;
- ensure that the use of the Information does not breach the Data Protection Act 1998 or the Privacy and Electronic Communications (EC Directive) Regulations 2003.

To see the full text of the OGL, please refer to :
<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/1/open-government-licence.htm>

Often by downloading the data and using it you are acknowledging the licence terms but without having to provide contact details or even click on an 'accept conditions' box.

Project owned data: The final class of data within a WP6 tool is the data collected specifically by its parent project. As the data originator it will be down to the project to define its own rules on data ownership and subsequent usage, taking into account data protection legislation, GDPR, FOI etc. It may be beneficial to the project, in terms of stakeholder engagement and data management, for the data ownership to reside outside of government. This option can only be explored further as part of a wider discussion with all the stakeholders and data legislation experts.

Implications for wider roll-out: As part of the data source review the various licensing options will have been explored. The project needs to select the appropriate licence for both the hosting body and the final audience for the data products. When using freely available data the project should ensure it is compliant with the implied or less than obvious conditions imposed when the data is obtained.

When developing the plans for its own data origination, careful thought needs to go into the proposal put to the stakeholders to ensure that the subject of who owns the data and what can be done with that data, is transparent. This is vitally important if the implementation model relies on vessels hosting the data harvesting systems on a voluntary basis. Experiences in this and other projects has shown that some vessels will not take part in a scheme voluntarily if party X owns and manages the data or party Y can have access to the data without even asking. There is particular nervousness associated with government departments having access to data, particularly within the fishing industry, and that the supplied information may be used for alternative purposes that reduce fishing opportunities.

The fact that data has value is now common knowledge. Any project that harvests potentially saleable data (e.g. sea temperatures, depth profiles) needs to have processes and structures in place at the outset to facilitate those sales with their own licence/usage/provision conditions. There is also a need for financial structures and processes to use the proceeds of those sales, possibly to offset running costs, redistribute to contributing vessels, or to expand a programme etc. These processes need to be transparent in the proposal put to stakeholders at the very start. Even if there are no obviously saleable data commodities at the outset of a programme, the potential needs to be acknowledged and the need for agreed methods of dealing with them in the future established.

The status of the data owner for project originated data sets should also be given careful consideration. A government department would be subject to demands such as Freedom of Information (FOI) requests for example. Obviously, these can be declined if a request contravenes the rules governing FOI but there is still the overhead of managing those requests. This could also impact on the project making saleable data products or adversely affect participation if the scheme was voluntary. Detailed discussions with legal experts in this subject area should be undertaken to determine the most appropriate hosting and ownership route.

4.5 The selected tools

R and Shiny are, in some circles, not considered appropriate tools for a production environment, with preference being for Python development. Over the last few years there

have been significant developments in the R environment and user-base, meaning it is becoming increasingly used in production systems. Should a wider scale roll-out of this prototype system become a reality, we see no reason to move away from the R, open-source, environment it was developed in, other than to update to the latest versions of the packages.

4.6 The resultant application

Following live release of the application, an informal questionnaire was circulated to a cross section of the users. The detailed responses can be found in Annex A, but a general summary by theme is given below. The results collected were not used in a statistical or quantitative analysis, rather the comments provided insights into whether the WP6 team had overlooked some aspects of the application development or could improve upon it in the future.

Ease of use: Generally, the respondents were very positive regarding the ease of use. Many found the tool to be intuitive and they did not have to refer to the user guide. However, some respondents' encountered performance issues, which possibly were due to their own Internet speeds, the way in which they selected data for display, or the internet browser they were using.

Graphics: On the whole the respondents were positive. In some cases, respondents quite rightly suggested that the colours differentiating quantities or tracks were too similar or hard to see against a background that also displayed substrate type. Others wanted to see more layers, such as bathymetry profiles or closed areas, and better labelling of catch and effort quantities.

Understanding of the data displayed: There was a positive response to this question. Some wanted to see the data displayed in different formats, perhaps with more layer options configurable to individual users. This was interesting as it may be worth considering in the next roll out whereby radio buttons are added so the user can select the different ways of viewing and displaying the data.

Mix of data: The majority of users said that the application contained the right mix of data. A good point was made that the areas for RIFG reporting were at a too large scale to be of use in a local management scenario, so maybe a future project should consider smaller aggregation levels through a consultative process.

Other useful data if rolled out: Respondents were asked what other useful data they would like to view if rolled out further. Again, this was varied ranging from static layers on other map pages e.g. bathymetric data, or national catch and effort statistics displayed as landing per unit effort profiles on the heatmaps. It will always be impossible to satisfy all users, but further consultation with different user types would allow a list of "top ten" preferred static layers to be established for display purposes. A list of the most useable way of presenting data for higher level users e.g. fishery managers, scientists, policy makers, could also be established.

4.7 The project concept

This prototype application, with further development, could potentially provide a management system needed for spatial management plans, together with information for a transparent equally participatory stakeholder driven management process. The latter is particularly relevant within the EU via the Regional Advisory Committees (RACs - Council Decision of 19 July 2004 establishing Regional Advisory Councils under the Common Fisheries Policy (2004/585/EC)), where there is a requirement to more closely involve stakeholders, such as the fishing industry and conservation groups, in management decision making processes.

For instance, when evaluating the performance of fisheries management measures, it is common to test the effectiveness of proposed management measures to alternative hypotheses about system dynamics, e.g. in terms of fishing fleet behaviour to multiple competing maritime sectors and not just population dynamics. The choice of an optimal policy is often problematic because different system dynamics models usually imply considerably different optimal policies and debates fail to resolve the plausibility of the alternative hypotheses for system dynamics (Smith, 2002). In the future, governments will need to take a more proactive approach to resolving these trade-offs, and therefore there will be a growing need for applications such as this with the inclusion of models to estimate and form the basis for evidential foundation for future seascape planning under the integrated maritime policy (EC, 2007). For instance, the output can then be summarised to define the interaction and hence the impact between fishing activity and species distribution at a variety of levels (which might include expert opinion on a species distribution over a broad geographic area) i.e. from rectangle (reported landings and days fished), to fishing effort derived from Vessel Monitoring Systems (VMS) (EC, 2003) or at the 5nm level from the outputs of the WP2 system. In addition to the SIFIDS (WP4/8) socio-economic inputs it could also be very useful to see data such as market prices, fuel price, insurance cost, employment rates etc. at as close to port level as possible. Again, there is potential to explore/infer how the existing fleet will react to changes in these costs. Unfortunately, the finished prototype application didn't have access to COMPASS (Scottish national catch and effort data) or any fine scale economic data to allow this to be explored further. This meant that although fleet interactions (for example) may have been a useful tool, we did not have real data to demonstrate a particular tool's usefulness or otherwise.

The level of aggregation is important, if the user wishes to look at the behaviour of vessels and other datasets in detail it would be beneficial to have information for all activities at a fine scale. Therefore, agreement will be needed on what the aggregation level for all commercial fishing activity and other commercially sensitive data should be. Furthermore, the core SIFIDS data is live whilst other data sources may be subject to a temporal lag (e.g. vessel file, or maritime planning) or be static data sets within the system, even though their source is a live system.

Further testing of this prototype would be needed to determine whether it could be used as a management tool but this may have cost implications in terms of server size to handle the large data streams and data processing required.

4.8 Opportunities for expansion

This expansion opportunity is mainly focussed on adding other data sets and layers. Over the course of discussions to date, various additional layers have been suggested, but due to limitations such as data access, we have not been able to implement demonstrations of these. Other data sources that could be considered useful, but the system currently doesn't have access to are found in Table 4. The potential data sources listed is by no means exhaustive and preferences will be user specific. Gaining access to additional data may require additional funding or intervention/licensing by 3rd parties, so cannot be guaranteed, but this illustrates how useful the WP6 application could be to a wide range of users.

Table 4. Potential additional data sources and layers for the WP6 application.

Description	Information	Issue
Oil and Gas infrastructure	Shapefiles	Cost: £1400+ per annum
Full catch and effort data for all fleets (including AIS/VMS)	Compass	No access to the data
Full catch value data all fleets	Compass	No access to the data
Economic data	Fuel prices, Insurance costs etc.	No access to the data

5 CONCLUSIONS

This project has shown that the development of an open-source system to achieve the initial project aims is indeed feasible given access to the necessary data sets.

To fully explore the data sets that have potential to improve the decision making or analysis processes the system intends to support, an approach that encourages cross-discipline thought processes within the user sample group, needs to be undertaken.

The largest barrier to a wider scale roll-out and adoption of the initial concept system is access to useful data. Without access to most of the data sets that all the user groups wish to visualise, the system is only a partial solution and unlikely to become a regularly used tool.

Before any significant resource is committed to building a full-scale, user-facing technological solution for the primary user base, firm commitments (beyond “in principle”) to data sharing and, we would suggest, initial technological implementation to under-pin those commitments, are essential for all the core data sets. This would help to ensure the subsequent commitment of development resources has proven access to the data that is the foundation of the system.

REFERENCES

EC (2003) Directive No. 2244/2003. Laying down detailed provisions regarding satellite-based vessel monitoring systems. Official Journal of the European Union 2003, L333: 17–27.

EC (2007). “Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions”, An Integrated Maritime Policy for the European Union, COM (2007) 575 final, Brussels, 10.10.2007.

Mendo, T., Smout, S. Ransijn, J., Durbach, I., McCann, P., Crowe, S. Carulla Fabrega, A., de Prado. I., James, M. 2019. Scottish Inshore Fisheries Integrated Data System (SIFIDS): Identifying fishing activities and their associated drivers. Report to Marine Scotland in fulfilment of EMFF: SCO1413 68pp.

Mendo, T., Smout, S., Russo, T., D’Andrea, L., James, M. (2019). Effect of temporal and spatial resolution on identification of fishing activities in small-scale fisheries using pots and traps, *ICES Journal of Marine Science*, , fsz073, <https://doi.org/10.1093/icesjms/fsz073>

R Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

RStudio Inc., (2013). shiny: Web Application Framework for R, <http://CRAN.R-project.org/package=shiny>.

Scottish Sea Fisheries Statistics 2017. <https://www.gov.scot/publications/scottish-sea-fisheries-statistics-2017/pages/4/>

Smith MD (2002). Two econometric approaches for predicting the spatial behaviour of renewable resource harvesters. *Land Economics* 78: 522–538.

ACKNOWLEDGEMENTS AND DATA SOURCES

The habitat data within the system are delivered under the terms of the [Open Government Licence](#), subject to the following acknowledgement accompanying the reproduced BGS materials: "Contains British Geological Survey materials © NERC 2018 <http://www.bgs.ac.uk/data/services/wms.html>

Scottish Marine Regions;

<https://data.gov.uk/dataset/f9ef823d-e672-4f35-8f00-41480dad7bf2/resource/b48874fc-870c-411b-865b-9e366cfe3f3d>

Scottish MPA's:

<https://gateway.snh.gov.uk/natural-spaces/dataset.jsp?dsid=MPA>

ANNEX A – DETAILED FEEDBACK RESPONSES (ANONYMISED)

The following section provides the feedback from the various users involved in the survey. These forms have been anonymised and any information that could identify a responder has been removed. Other than those redactions these forms remain as they were returned.

Responder 1

1. How easy was it to use (seeing as people seldom read user guides properly)?

Fairly easy to intuitively use. Unfortunately, the RIFG management area information did not appear to be enabled? You are right I did not at first go to the guide!

However, have I used all the features?? A checklist of what can be enabled and viewed may be useful?

2. How did you find the graphics?

They were good. Is it possible to view more as a chart with depth contours illustrated?

3. Was the way data was displayed easy for you to understand, or would other options be preferred for your level of access?

The catch / effort data was unclear in terms of the units indicated – number of lobsters or tonnes? Was it pots in the water or pot days fished etc

When more data becomes available for instance pot numbers is there a risk that the graphics will be snowed out?

Could do with a key to illustrate the background features such as 6nm line or roll over with cursor to get description?

4. Was it the right mix of data for you? *Data is currently limited, but if there is any other data you believe that would be valuable to include please indicate below*

Potentially the depth contours on the background charts? Would it be possible to identify any statistics associated with vessel track such as time from port to first fishing?

5. What data would be useful to see on a properly rolled out tool and what display options would be preferred?

Would it be possible to look at activity from a named port and indicate the spread of fishing locations the vessels go to after leaving port?

6. If you have any further comments, please put them below

Would it be possible to annotate more of the features being viewed on a chart and will it be possible to take a screen shot and identify what is being viewed. This may be most important when viewing activity within an RIFG area or a discrete part thereof?

Responder 2

1. How easy was it to use (seeing as people seldom read user guides properly)?

The more times you play with it the more you discover.

2. How did you find the graphics?

Not sure about MPAs. Creel amount colours, difficult to distinguish between dark orange and red.

3. Was the way data was displayed easy for you to understand, or would other options be preferred for your level of access?

At first had difficulty seeing creel data on map. Could only see the data for the example in the instructions. Would like to see data for the vessel in my area.

4. Was it the right mix of data for you? *Data is currently limited, but if there is any other data you believe that would be valuable to include please indicate below*

Some very interesting data in the system. Would take time to find all the information of interest.

5. What data would be useful to see on a properly rolled out tool and what display options would be preferred?

Not sure if I could see individual creels. Would be very interested in being able to count number of creels a vessels working.

6. If you have any further comments, please put them below

- Wasn't able to access using RIFG username/password
- Was using my own WIFI going in straight from the email link.
- Found calendars slightly frustrating – using ipad the top of the calendar arrows and month not visible. Entered dates manually if I could select from calendar.

- MPA information misleading – There are many parts to the Loch Sunart to Sound of Mull MPA. Seasonal closures, restricted areas for certain gear with in the MPA. This would all need to be displayed to help Fishers work out where they can and can't fish.
- Small Isles MPA does not exist. Initially proposed but never pass by law.

Summary

I'm not a great one for user instructions but following the instructions got me the results I was looking for. When I went off piste, I found it difficult to find data for my local boats. I can see many interesting bits of information coming from this system. The more you have a go with the system the easier you find your way about.

If we had some added data that the individual Fisher could use it would make it more attractive to the Fisher. I have had dealings with Marine Instruments before and they were looking at providing added benefits when you installed their system. If the system included a location alarm, if their vessel moved a certain distance from its mooring, or a bilge alarm connected to the system. If these systems become mandatory, then it would be in their interest to look after it better if it was supplying the Fisher with valuable information.

Responder 3

1. How easy was it to use (seeing as people seldom read user guides properly)?

I thought the app was well laid out and very easy to use / switch between functions.

I tried the app both on a 15 inch laptop and on my iphone, both of which worked well although on the iphone it was obviously more difficult to scroll down between the search section and the map.

I found the app was good and intuitive to use so once more data is integrated across areas will be very useful. However, on my slow wifi, it was taking a little while for the limited data that is there to download and display, also there was a significant time lag between selecting map/ track data and this appearing on the map.

The inclusion of the drop down notes section at the top of some pages works well and could avoid the need for a separate document

I would have found it easier if the area of map under scrutiny could have been selected / lassooed by a box, as in some areas the ports are closer together and some are not marked with a dot showing the number of vessels, which makes it more difficult to be sure one only has data for the chosen area.

2. How did you find the graphics?

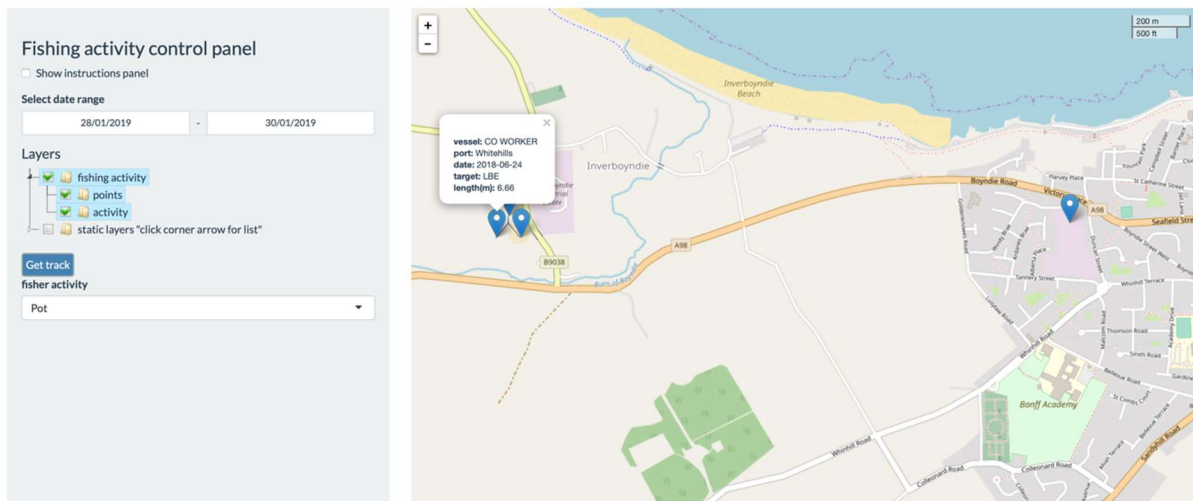
I found the colours for the 5km square boxes showing effort / catch were too similar at the top (red) end of the scale to differentiate easily between the top two grades. The colour looks different when the semi-opaque squares overlay the map compared to the legend.

Also, some of the fishing track lines under the map data section are so pale they are very difficult to make out. Are the colours of the tracks allocated at random?

The tracks are often obscured / not visible against the background bathymetry layers but this may not be an issue if one zooms in.

3. Was the way data was displayed easy for you to understand, or would other options be preferred for your level of access?

The information shown under the Map Data / Fishing Activity Control panel section seems to have allocated the vessel location as the home address of the owner/ co-worker rather than the harbour. See map screenshot attached and data here for one of the blue vessel locators.



In the Explanatory notes for the Fishing Activity Control Section, it would be useful to mention what is represented by the Activity Data, as opposed to Tracks.

Possibly also mention that where the number of track points is 0, no track will appear (obvious really, but on the small mobile phone screen, the chart column showing the number of tracks was sometimes obscured.)

Under Map Data / Fishing activity Control panel - once you have obtained the data for an area, you can do a search by vessel ID. This does not show up the trips in Date order but appears to be listed according to the number of trackpoints. (see screenshot below) Date order might be more useful.

Static layers Click corner arrow for list

Get track

fisher activity

Pot

Show 10 entries Search: mb

vesselid	starttimestamp	trackpoints	trackuptodate
MB	2018-11-20 07:46:41	188207	true
MB	2019-01-04 11:16:13	67278	true
MB	2018-11-11 05:34:04	35459	true
MB	2018-11-26 06:10:26	34671	false
MB	2018-12-01 06:04:57	33034	true
MB	2018-11-18 05:54:04	32156	true

Showing 1 to 10 of 65 entries (filtered from 815 total entries)

Previous 1 2 3 4 5 6 7 Next

(Although your notes say one can't do a search by vessel / date, the data is given for the vessels registered in an area so it would be easy enough to identify and search out a vessel's tracks.)

Under Vessel Characteristics, the Regional names are the old IFG areas, not the current regions - e.g. South West , MF&NC, South East.

The x axis of Vessel length graphs for some areas (East Coast , MF&NC, Orkney) doesn't extend to 12 metres (for 12 mu vessels). It would be helpful to be able to have a bigger scale so one can more clearly distinguish between the under 10 and under 12 vessels.

Under RIFG reporting, there are two extra regions listed called Ayers Coast NNW and Tidd Coast NNW .

The effort data was clear but where this is map based, is there an easy way to view and compare statistics between time periods? (the non map based RIFG report function does allow this, but this is at too great a scale to be locally useful.)

It might be worth adding an explanatory note section to RIFG Catch and Effort reporting page to show one can select multiple periods to get a table of comparative data. (the SIFIDS Catch and Effort / National Catch and Effort functions look similar in format but that data is heat -map based, so doesn't allow easy comparison.)

Is the top 'select year' box in the RIFG or National Catch and Effort reporting sections required? as there is a select year function box at the bottom too (In the SIFIDS effort section , this is required as the lower box only gives month/ quarter).

If the option on the lower search criteria box is set to annual reporting when you open the SIFIDS/ Regional / National effort pages, the year shown in the top box is greyed out and can't be changed, unless you change the lower selection criterion to monthly or quarterly reporting instead.

Is it the intention to only populate the National Catch and Effort Data section in future and to keep the SIFIDS data as additional historical information?

4. Was it the right mix of data for you? *Data is currently limited, but if there is any other data you believe that would be valuable to include please indicate below*

The RIFG report allows two or more years / periods to be selected for comparative data, but this is per RIFG area. Could this function be extended to allow comparison for a slightly smaller area although obviously not at too small a scale which would allow estimation of the number of creels used by individual vessels?

5. What data would be useful to see on a properly rolled out tool and what display options would be preferred?

Will fishers be able to interrogate their own data in more detail?

Is this database to be accessible to MS Compliance for enforcement purposes, or will they use other means to check vessel tracks / activity patterns, e.g. where these may encroach on a protected area?

Presumably this database would only remain current if it is used to record the national statistics / effort / catch data from FISH1 forms ?

If different groups/ the public have access, will this be restricted

Other useful features might be:-

An option to save searches / heat map data for later use (although one can take screen shot)

Comparative effort data by local area / fishery over a range of time periods.

Landed quantities and port of landing data would be interesting but probably too complex / variable to be feasible.

The port characteristics / infrastructure investigated under the SIFIDS project would also be interesting and useful as a reference source, but the data is not complete.

(Note - For info only - during early 2019 the North and East Coast RIFG is planning a survey of local harbour facilities, management etc., to be completed by local fishers / fishing associations. There is a possibility this will be carried out by other RIFG areas.)

6. If you have any further comments, please put them below

The Section 4 data on fishing drivers is interesting, but a partial and historic snapshot . For non-statisticians I wonder if this could be referred to in a simpler way / list . I presume this data will not be added to in future?

The button on the tool bar marked 'Map Data' brings up the 'Fishing Activity Control Panel', so should this tab be renamed since other tabs also have mapped data?

Responder 4

1. How easy was it to use (seeing as people seldom read user guides properly)?

Good, I didn't look at the guide and think it all made sense (except see notes below)

2. How did you find the graphics?

Good – scaling of the left-hand panel sometimes a little out. Need to zoom out further on map display.

3. Was the way data was displayed easy for you to understand, or would other options be preferred for your level of access?

Format is good. I would always want to extract data wherever possible, which currently only available on

4. Was it the right mix of data for you? *Data is currently limited, but if there is any other data you believe that would be valuable to include please indicate below*

No current application for data

5. What data would be useful to see on a properly rolled out tool and what display options would be preferred?

Intro page suggests live vessel tracker, similar to MarineTraffic.com, which would be interesting

6. If you have any further comments, please put them below

Couldn't get SIFIDS Catch and Effort page to function correctly – everything came back with "You selected a year, month, area combination contains 0 data" . Maybe no data in there yet?

Wasn't sure of difference between SIFIDs and National Catch and Effort datasets. Why separate?

Options to view: pots, crabs or lobsters. Unsure what this means – does pots contain all crab and lobster? Or nephrops?

Unsure about map data scale – selection boxes say Catch (t)/Effort(days at sea) – where (t) is tonnes? But e.g. on National catch and effort, annual catch of crabs in pots indicates upto 2500 tonnes per day. Scale needs to be in the legend.

What does ‘stopping’ mean in relation to drivers? – Stop fishing for the day? That page needs better explanatory text. Is data self-reported, or calculated somehow from multivariate data?

Drivers page could have a dropdown box (or checkboxes) to select drivers, so you can compare your choice of factors, rather than view all at once.

Many of the factors in here also cry out for more explanation – if there is more related data perhaps it could be selectable. E.g. depth as a factor for placing... could click on ‘depth’ to see more data on depth distribution.

Would be interesting to include catch and effort, per vessel, in the vessel characteristics tab. (as B&W plots to show distribution of catch, effort, and catch per unit effort.)

Big datasets e.g. habitat layer in Map Data take a few seconds to load each time you reposition the map, which looks quite glitch.

Also needs a legend/key for habitat type layer

What are Tidd and Ayers Coasts, in RIFG page?

Responder 5

2. How easy was it to use (seeing as people seldom read user guides properly)?

The core software menus were easy to use and navigate without reading the instructions, however the search functions and data fields used labels that weren’t all clear in their meaning, and some were quite unintuitive, even if they worked well once you knew how they functioned. (Thinking particularly of the way multiple time periods are selected for the RIFG reporting from the drop-down box. It works well, but I haven’t seen that system used often in other software, so might be useful to make this clearer, and it lacks a ‘Select All’ option which involves much clicking when doing an annual breakdown by month.

3. How did you find the graphics?

For the most part it was clear and nicely presented. I was viewing on a 1440p PC monitor, and this resulted in text/numbers on the ‘Vessel characteristics’ and ‘Fishing drivers’ charts that was small and difficult to read. I suspect this was also the reason why the map only took up a small portion of my monitor even when I was viewing in full-screen mode.

4. Was the way data was displayed easy for you to understand, or would other options be preferred for your level of access?

When I could navigate the search features and get data displayed, it seemed to be presented in a clear and straight-forward fashion that was easy to understand.

5. Was it the right mix of data for you? *Data is currently limited, but if there is any other data you believe that would be valuable to include please indicate below*

The data available would be interesting in terms of being able to identify changes in local fishing patterns, and potentially useful in terms of finding vessels that might be causing problems in or around MPAs.

6. What data would be useful to see on a properly rolled out tool and what display options would be preferred?

Being able to overlay the data from the different map functions on each other would seem to be useful.

7. If you have any further comments, please put them below:

Conducting a search of a lot of position data (for the whole of 2018 over a wide area) and trying to display a lot of positions for one vessel caused the system to hang, and then fail. Did not have the time to test this more thoroughly, but it seemed to struggle with large quantities of data in general which caused a few hiccups. I presume this will be addressed in the release version.

From a user perspective, it is not entirely clear why there are multiple map sections, rather than a single map function with multiple search options.

Presumably due to the quantities of data involved, but the map-search function having a limited zoom-out obviously limits the active search area, where otherwise it might be useful to do a wider search and then narrow down onto points of interest afterwards.

I had difficulty identifying tracks on the zoomed out map, but unclear whether this was due to the data not loading in properly, or by design. Unable to do better searches on this due to the data crash issue mentioned above.

Responder 6

1. How easy was it to use (seeing as people seldom read user guides properly)?

It was fairly easy to use after I spent some time playing around. Below are some points that may improve ease of understanding:

On Intro page;

- Provide hyperlink to SIFIDS Project webpage (<https://www.masts.ac.uk/research/emff-sifids-project/>)

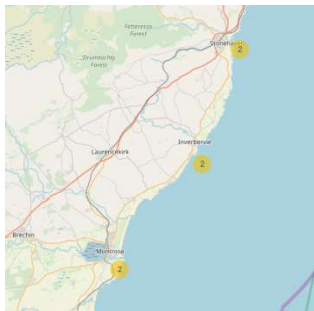
On the **Tab for Map Data**:

- When I select 'points' it shows me a continual fishing vessel track (not points), so can you rename it to 'vessel track' please?
- next to 'activity' it would be helpful to insert some adjacent text in brackets telling user to 'select from options below'
- rename 'static layers' to 'marine boundaries and habitats'
- Under static layers rename 'three, six and twelve' to '3 nm limit', '6nm limit' and '12 nm limit'.

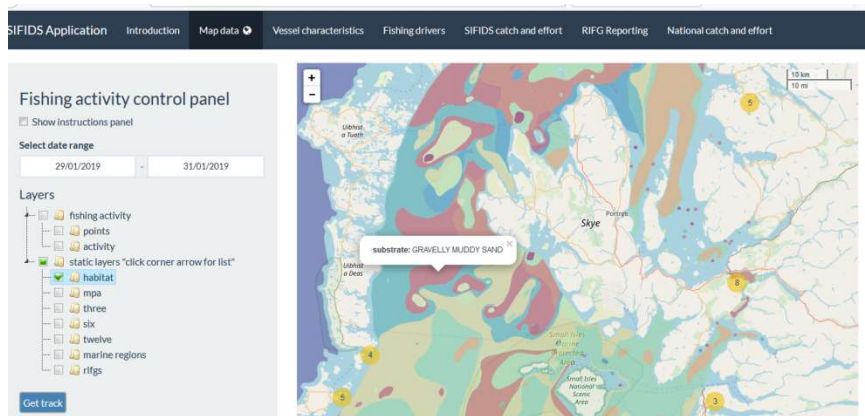
2. How did you find the graphics?

Clear and well presented, however some of the layers took a while to load (mainly static layers). In addition, the server also 'crashed' for a while and I was unable to access the database for several days. Below are some points to consider;

- I couldn't see an explanation anywhere for what the yellow circles with numbers are (below), so can you provide one please? (P.s. I know what they are because I'm a SIFIDS person, but other non-sifids people won't know)



- Map currently doesn't zoom out wide enough to show the whole of Scotland out to the 12nm limit, can you change this please? I.e. allow map to zoom out much wider.
 - Can you add a key for the habitat types, it might make the maps easier to interpret. Clicking on the colour to get habitat name is a good feature, but if someone wanted to print off a map showing how fishing patterns overlap with habitats then a key would be needed.
 - There is very little coverage for habitat types in some areas (e.g. Inner Sound Skye), why is this?
-

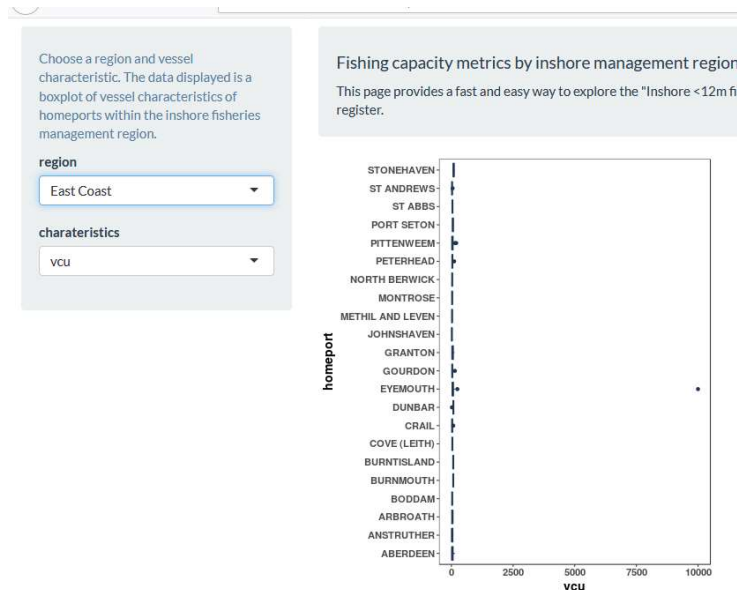


3. Was the way data was displayed easy for you to understand, or would other options be preferred for your level of access?

Generally fairly easy to understand. However I suggest the following:

On the tab for vessel characteristics;

- What exactly are 'inshore management regions'? I think you have listed what appear to be the old Inshore Fisheries Groups, which have subsequently been changed to the Regional Inshore Fisheries Groups (e.g. The SW and NW should now be amalgamated to the 'West Coast RIFG', and the MF& N coast and East Coast should now be amalgamated into 'NE Coast RIFG').
 - Can you please provide a bit more detail about what the box plots are showing? e.g. inter-quartile range, average and outliers (I think that is what I am seeing)
 - can you also add an option where people can view a spreadsheet summary of the box-plot data, or allow them to download the actual summary statistics?
 - for VCU on the East Coast & South West, there seems to be an outlier that totally skews the plot and makes it hard to read. Can you remove this outlier or think of another way to view the data that is more meaningful (see below)
-



Tab for Fishing drivers

- This is the least easy to understand data set.
- I think the introductory text needs to provide a better explanation of the study and how the data is presented. E.g. start by saying "In 2017/18 x-number observer trips were carried out on static gear vessels around Scotland, and fishers were asked how they decide when and where to fish and what factors might stop them going fishing...." "The data presented below shows ..??? (explain)".
- Under plot 1 you need to be more clear about what the scale means. I know you explain in the left tab, but eyes don't naturally look there when trying to figure out a graph. So I suggest you put the explanation directly under plot 1 next to the key showing shades of brown and blue, & state that the "scale represents the extent to which factors influence fishers decision making, 5 = likely, 1 least".
- In the 2nd plot what does 'missing' and 'completed' mean?

Tab for SIFIDS catch and Effort

- The meaning of the term "catch/effort" is not immediately obvious. I assume you are showing effort data and not catch data here. In which case you should rephrase to "Effort according to target catch". Then state what effort is measured as, e.g. "effort is recorded as number of pots in water per fishing day".

What does 'pots' refer to (below)? Is it Nephrops pots/creels? Or some kind of mixed catch in a pot? If Nephrops then change to 'prawn creels'.

Catch/Effort numbers for:

- ☒ pots
- ☐ crabs
- ☐ lobsters

- It's a shame the grid squares couldn't be smaller (I know you have to stick to Marine Scotland specifications). A grid of 2nm x 2nm would be ideal for local management I think.

RIFG reporting tab

- no data to view at the moment.
-

4. Was it the right mix of data for you? *Data is currently limited, but if there is any other data you believe that would be valuable to include please indicate below*

Tab for national data

- Under options for gear type & species it would be useful to have fishing gears listed that are more relevant to Scotland's inshore fisheries. E.g. would be good to include:
 - **Gear types to add;** Dredge; Demersal trawl; Hand Gathered; Handline
 - **Species to add;** king scallops, queen scallops, prawns; whelks; mackerel
-
-

5. What data would be useful to see on a properly rolled out tool and what display options would be preferred?

-
- Total landings by port (monthly and yearly averages in tonnes and value)
 - Numbers of vessels registered per port in following size categories (u10m, 10-12m, 12-15m and >15m).
 - VMS data that can be filtered by gear and target catch (e.g. scallop dredge, prawn trawl, etc)
 - Scotmap data
 - Landings per unit effort data (LPUE) per target species at 'an appropriate local scale' (2nm x 2nm)
 - It would be useful for managers to be able to download maps of different data layers, e.g. maps showing habitats with fishing effort data superimposed.
 - Would be useful to have a 'stock status' tab with links to most recent Marine Scotland Stock Assessment data for inshore species (mainly crab, lobster, Nephrops, scallops)
 - Would be useful to have a tab listing current research projects relevant to inshore fisheries in Scotland, plus contact details (or links) to relevant institutions.
-

6. If you have any further comments, please put them below

Responder 7

1. How easy was it to use (seeing as people seldom read user guides properly)?

Fairly easy, the only section that wasn't straightforward to me was the Map data, I had to read the instructions. I didn't know you had to choose an area in the map first and that the search would be constraint to just that area. I thought you chose a period of time and then it would zoom to the areas where you had fishing tracks.

2. How did you find the graphics?

Good.

3. Was the way data was displayed easy for you to understand, or would other options be preferred for your level of access?

The fishing drivers section is not easy to follow.

4. Was it the right mix of data for you? *Data is currently limited, but if there is any other data you believe that would be valuable to include please indicate below*

Yes

5. What data would be useful to see on a properly rolled out tool and what display options would be preferred?

Number of vessels fishing per grid cell, number of fishing trips per vessel

6. If you have any further comments, please put them below

I had issues with the Map data layer – the tracks would only appear briefly, then disappear, then it would disconnect me from the server.

Is there no catch data for crabs and lobsters from SIFIDS? I couldn't see any representation of this.

In the national catch and effort I would get the message "Congratulations you have data for your selection" but there was no data displayed. Gear type has to be linked with the target species. So far I had haddock caught with pots.

RIFG reporting – no data displayed – unable to download data.



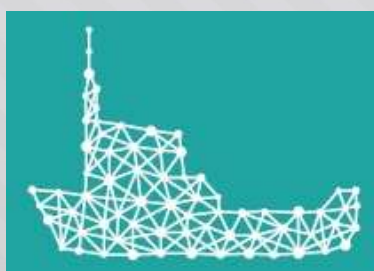
University
of
St Andrews

marinescotland



The Scottish
Government
Riaghaltas na h-Alba

SeaScope Fisheries Research
Independent Marine Consultants



Marine Alliance for Science and Technology for Scotland
C/O Scottish Oceans Institute
East Sands
University of St Andrews
St Andrews
Fife
KY16 8LB
Scotland/UK

T: +44 (0) 1334 467 200

E: masts@st-andrews.ac.uk

W: www.masts.ac.uk

MASTS Company Number: SC485726; Charity Number: SC045259

ISBN

EMFF: SCO1434

This project was supported by a
grant from the European
Maritime and Fisheries Fund

