June 2025

FMRI DIGITAL INFRASTRUCTURE SCOPING STUDY

USER AND STAKEHOLDER RESEARCH REPORT

FMRI FUTURE MARINE RESEARCH INFRASTRUCTURE

British Antarctic Survey NATURAL ENVIRONMENT RESEARCH COUNCIL

National Oceanography Centre

SAMS

University of Southampton

UNIVERSITY OF PLYMOUTH

UEA

Met Office

MEDIN marine environmental data & information network

UKRI Natural Environment Research Council

Department for Environment, Food & Rural Affairs

KEY FINDINGS

TOP 5 AREAS WHICH REQUIRE SUPPORT - BASED ON IDENTIFIED NEEDS

01

CREATING VALUE FROM DATA - THE BIGGER PICTURE

Users face friction and inefficiencies because of inconsistent foundations across the digital infrastructure, tech. architecture and data standards.

Digital infrastructure and data are not well interconnected; there is no overarching data aggregation and distribution strategy. Metadata capture is perceived as an issue; frequently it is considered last or “tacked on” (even though part of funding grants), this creates critical issues for future AI readiness. Without these foundations in place, it’s time consuming and more difficult to get the ‘bigger picture’ value from the data and improve optimisation, reporting etc.

How to address gaps:
Improve find-ability, signposting and user experience across digital ecosystem, service design programmes. Eliminate duplication, increase re-usability. FMRI could lead in promoting / endorsing preferred data standards. Incentivise or regulate metadata capture for Scientists at grant level (not just in the Data Centre form ‘addendum’), invest in digital tools to support in automated metadata capture. invest in historical metadata capture.

Supporting Key Insights: 1,2

02

NEED FOR A UK-WIDE OCEANOGRAPHY FRAMEWORK

Gaps in communication between infrastructures and users exists due to the scale of the ocean science ecosystem. There is a lack of awareness around ongoing initiatives; data management, tools and infrastructure. Digital interoperability between organisations is inconsistent (data does not flow well, and digital integrations are done ad-hoc). There is no underlying digital backbone or framework across NERC, or supporting Organisations / Institutions.

Across Organisations, users end up working in silos - which makes it harder and more time-consuming to achieve ‘big-picture’ understanding of data, and results in rework and inefficient use of existing digital tools.

How to address gaps:
FMRI are ideally placed to lead a UK-wide framework (Please refer to the MFRI DISS Report document for more information on the type of organisational structure this could take). We can aim to reduce ‘data silos’ through infrastructure, technical architecture, service-level design and digital tools (signpost, access, notify). This will require considerable investment in under-represented skillsets within Ocean Science (see page 30).

Supporting Key Insights: 2, 5, 6, 7

03

SUPPORT TO KEEP THE PACE WITH ONGOING DIGITAL EVOLUTION

Scientists feel unprepared and unsupported amid rapid changes (AI, cloud computing, and autonomy) despite recognizing their potential.

Projects are outpacing the capabilities of existing digital infrastructure (see Page 37), revealing inefficiencies in delivering scalable digital tools. Key roles—Technical Architects, Digital Product Owners, UX/UI Designers, Front-End Developers—are often missing or siloed, resulting in ad-hoc tools and poor interoperability. **Users struggle to adopt or innovate new digital technologies because the burden of cost (eg: subscriptions, NVIDIA hardware) and funding gaps hinder agile experimentation and restrict access to necessary technologies.**

Little automation exists before data reaches data centres (e.g., QC/QA), causing delays and hindering 24/7 monitoring, as outdated field equipment still relies on manual upkeep.

How to address gaps:
Well designed) digital tools can support where time & resource are lacking to increase efficiency and effectiveness. People want tools to allow them to be independent, but they want them to be easy to use and not having to manage the infrastructure – improved user experience and applying system design thinking need prioritisation. Thinking styles and existing constraints need to be clearly understood and respected, to manage change during digitisation. But not everything can be solved digitally. Investment in people (increased training & collaborative opportunities) need to be factored in. In terms of automation there are good examples where it works (ARGO) from which we can learn and optimise.

Supporting Key Insights: 4, 6, 7

04

CONNECT FIELD & OFFICE - END TO END SERVICE DESIGN

Users experience a disconnect in working hours between field and office teams and this creates issues with real-time coordination and responsiveness.

Autonomy requires 24/7 support. What happens if digital systems break out of hours, how are scientists supported?

Not all the parts of the autonomous fleet delivery chain are resourced 365/24/7. The current infrastructure has not been scoped to support that type of uptime or significant increases in fleet size.

How to address gaps:
How tTo create the future’ infrastructure - service design across the full data flow needs to be prioritised from a user/operator point of view.

Introduce improved tooling or digital infrastructure systems to reduce human dependency on solving issues which arise from (for example) out of hours data transfer, or piloting vessels / fleets at volume. Note: there are already projects funded by NERC underway demonstrating large fleet management with AI support – these pilot studies are examples of what can be achieved (see Page 37).

Supporting Key Insight: 3

05

FOSTER ENGAGEMENT WITH EXTERNAL ORGANISATIONS AND COMMUNITIES BEYOND OCEAN SCIENCE- IN BALANCE WITH INCREASED SECURITY NEEDS

Scientific and research organizations have traditionally focused on research outcomes, not always aligning with the decision-making needs of policymakers, private companies, or the public. This limits effective engagement and global impact, highlighting the need for a shift in communication. **Users outside the ocean science community struggle to engage meaningfully due to barriers in data accessibility, a lack of supportive digital tools, and added cybersecurity constraints.**

With FMRI’s goal of "opening access to under-represented groups" and the cross-government ambition of "open by default", we must enhance engagement to deliver public good, while balancing cyber-security requirements and ensuring the right information reaches the right audience.

How to address gaps:
Support efforts that enhance access to and understanding of ocean science data for non-experts, improving ocean literacy. Promote digital tools that ensure publicly funded data reaches end-users securely, with audit trails to show research value.”

Supporting Key Insight: 8

KEY USER NEEDS

KEY INSIGHT	USER NEED (KEY FOCUS AREA)	IN DETAIL
01 Inconsistent foundations across the digital infrastructure.	Reduce manual transfer of data from sensors to digital data platforms.	Page 22
02 Disconnected systems = data silos. Getting ‘bigger picture’ understanding from data.	Fewer ”data silos” (better signposting, ease of access, eliminate duplication, increase re-usability). Standardised data formats across sensors and data software for wider interoperability.	Page 24
03 Disconnect and time considerations between field and office.	Improved tooling or digital infrastructure systems to reduce human dependency on solving issues which arise from (for example) out of hours data transfer.	Page 27
04 Digital infrastructure investment needs to include investment in collaboration, people and skills.	More effective and efficient collaboration with others (digital infrastructure, software and platforms). Up-skilling in new technologies (AI).	Page 28
05 Some types of oceanography research are more suited to digitisation/systematization, AI and autonomy.	Prioritise the creation of software, systems and tools which streamline repetitive data collection and administrative processes.	Page 29
06 To enable innovation, room for short term experimentation, collaborative project opportunities & training need to be fostered.	If capability is increased (eg: autonomy, AI) a mechanism needs to be introduced which allow trials and testing with new digital infrastructure and equipment.	Page 31
07 Thinking styles and existing constraints need to be clearly understood and respected, to manage change during digitisation.	Respect and work with, not against existing ways of working. Co-creation, communication needs to be factored into any digital infrastructure upgrade.	Page 33
08 Increasing need to connect outside of Organisations with “Non-Ocean Science” communities – but balancing this with security needs.	Access, use and levels of comprehension and understanding of ocean science data without having specialist knowledge in Ocean Science / Oceanography	Page 34

KEY INSIGHTS

01

Inconsistent foundations across the digital infrastructure.

Users face friction and inefficiencies because of inconsistent foundations across the digital infrastructure, tech. architecture and data standards.

Some data is available, but signposting is often lacking, and some is also kept within institutions (not publicly available unless specifically requested) - because it takes work and effort and money to make them available. The digital infrastructure which currently supports the flow of data from collection and measurement devices to the hands of the Scientists currently relies on manual transfers and lengthy processes which currently often rely on human intervention (frequently at inconvenient times, out of hours).
There is currently no uniformity of data upload from instruments across Departments, Organisations or even Countries (Local / Global).

“**System interoperability is a key enabler.**”

Note: To digitise end-to-end processes , enable UK-wide data connectivity and access technical audits would need to be carried out. A ‘cross-Organisational’ level approach is required - ideally over a sustained period (e.g: 10+ years).

02

Disconnected systems = data silos. Getting ‘bigger picture’ understanding.

Across Organisations, users end up working in silos because of gaps in the ecosystem, which makes it harder and more time-consuming to achieve ‘big-picture’ understanding, and results in rework and inefficient use of existing digital tools.

For users to gain value from datasets, a lot of resource currently go into processing (from raw to analysed). This effort is then duplicated across different Scientists within, and across Organisations (e.g: repeatedly downloading the same datasets for modelling takes up space, increasing storage requirements). For a variety of reasons, there is a lack of consistency in data formatting and determining the value of data across the complex user ecosystem (See diagram on **Page 17 and Research Findings in Detail, Page 19).** We can aim to reduce these ‘data silos’ through infrastructure, technical architecture, service-level design and digital tools (signpost, access, notify).

Read more about the how we might do this in Section 3 – Recommendations, Page 47.

03

Disconnect between field and office hours

Users experience a disconnect in working hours between field and office teams and this creates issues with real-time coordination and responsiveness.

More specifically - a service-design disconnect exits between the Autonomy Piloting Team, Vessel Research Scientists and Data Centres. The ability to collect and measure certain types of data in the ocean is weather & season dependant. Gliders and other autonomous vehicles need to be deployed under certain weather conditions. Long-term data series need constant monitoring.

For example: if a glider is piloted out of hours, and an issue arises out-of-hours at a data centre during the transfer, often the issue can’t be resolved for up to 48+ hours, or over longer holiday periods. This can result in workflow data-transfer issues.

KEY INSIGHTS

04

Digital infrastructure investment needs to include investment in collaboration, people and skills (eg: training in AI).

Investment in optimised collaboration, people and skills isn't treated as a strategic priority so opportunities for innovation are missed. Closer collaboration and diverse expertise cross-Org is needed to design cohesive, user-centred, and future-looking solutions. Platforms are not enough. Any increases in data collection (more instrumentation, increase in coverage areas) requires a respective increase in digital tooling and investment in skill development and/or resourcing. New technology introduces new levels of complexity and requires a mindset shift (**see Insight 7**)**.**

There is a learning curve required for the AI expert to effectively understand and work with oceanography data. Conversely, the field of Oceanography currently has a "handful" of ML/AI specialists. The shortage of skills is a limitation. Therefore, we either need to look at more training and new role creation and/or support individuals to collaborate more effectively with the use of improved digital infrastructure and tools.

05

Some types of oceanography research are more suited to digitisation/systematization, AI and autonomy.

Users benefit unevenly from digital tools because some areas of science are inherently less automation-ready and require more resources (greater time, cost or complexity) to digitise. Most manual process are "research side". Science and processes which focus on systematic, repetitive data collection are "ripe for digitisation".

For example: Hydrographic surveying has 'narrower' groups, doing the same things repetitively on cruise, and from project to project - in comparison to Biogeochemistry. "Basics" like contextual events logging, issue tracking and administrative processes were also highlighted.

In each specialism there are many potential areas to do things differently - for example, using autonomy. Some Scientists also require a deeper level of customisation on existing platforms (eg: API's, own tools, plots etc.)

Research Findings In Detail, Page 29.

06

To enable innovation, room for short term experimentation needs to be fostered.

Users struggle to adopt or innovate new digital technologies because the burden of cost and funding gaps hinder agile experimentation and restrict access to necessary technologies.

Software subscriptions (eg: Cloud, SAAS) and hardware costs (eg: NVIDIA) are stifling to science discovery and innovation. Funding/costing models differ to private Companies (links to Insight 04, collaboration reduces resource costs, but software licenses are required to facilitate asynchronous collaboration).

The Science funding landscape is changing (and has changed) drastically, with a reduction of short-term study funding. If capability is increased (eg: autonomy, AI) a mechanism needs to be introduced which allow trials and testing with new equipment (eg: short term / proof of concept studies), along with the technology and digital infrastructure to support these studies.

KEY INSIGHTS

07

Thinking styles need to be clearly understood and respected, to manage change during digitisation

Introducing new processes and technology involves managing change. Digital maturity differs across Research Organisations. The ‘culture of collaboration’ starts at undergraduate level, and needs to be factored in, when designing solutions. **Ingrained thinking styles and differences in mental models create addition friction during digital transformation and collaboration.**

During interviews, Scientists across different departments and Organisations used words like ‘reluctant’ and ‘mind-sets’. Data values differ across different teams, and Organisations.

An example use case: a Scientist may want to collect data, protect its IP, and only expose it to ‘public’ view once it’s been peer-reviewed. There is a reluctance to share work until published. On the end, a Developer may want to use the research imagery as soon as it’s collected to develop pattern recognition models. There is a culture of using GITHub, and open source is the standard.

08

Increasing need to connect outside of Organisations with “Non-Ocean Science” communities – but balancing this with security needs.

Users outside the ocean science community struggle to engage meaningfully due to barriers in data accessibility, a lack of supportive digital tools, and added cybersecurity constraints.

With the FMRI objective to “pursue an approach that is outward looking and opening access to under-represented groups”, as well as the cross-Government ambition of “open by default” – we need to actively engage with the public (eg: Business owners, students), the Government and other Stakeholders to deliver public good - while keeping the balance on cyber security and ensuring the right information lands with the right audience at the level it’s required.

Improvements in digital infrastructure could allow audit trails to better demonstrate the value of research while increasing security. In terms of ocean access - although it’s important to note that not everything can be solved digitally – opportunities for teaching and training (eg: virtual reality AUV piloting) are currently largely unexplored.

Research Findings in Detail Pages 18-45

CONTENTS OVERVIEW

01 Study Overview
What we did and why

This section includes:

- Key findings summary – **Pg. 3**
- Research goal for FMRI DISS - **Pg. 10**
- Research objectives and methods – **Pg. 11**

02 Research findings in detail
What we found out

Go to page 19

This section includes:

- Ecosystem map – **Pg. 19**
- “Current state” User Journeys
  - User Journey 1 (Vessels) – **Pg. 20**
  - User Journey 2 (Autonomous Platforms) – **Pg. 21**
- Research findings supporting each Insight (in detail) – **Pg. 22**
- Survey findings – **Pg. 39**

03 Recommendation
What’s next…

Go to page 47

This section includes:

- Recommendations
- Potential opportunities service map (“How might we” statements)

Section 1:

STUDY OVERVIEW

Research goal

*"**What is the right spectrum of capabilities** that are affordable, and that support the Scientists going to places they want to go to, to measure and observe the ocean in the right way…*

…the question we're trying to get to grips with is – do we want to just change that slightly with this injection of cash that we're aiming to get – that is more useful in the future. Absolutely not an either or – ships will always be there.
Do we need more ships, more autonomy, a different mix that will be used in different ways?

…what would be the future operating costs of this kit and how we would do it?"

- How do we design for the future?
Leigh Storey (FMRI, SRO)

NERC has initiated the Future Marine Research Infrastructure programme (FMRI) with a mission to advance marine science by equipping researchers with cutting-edge technologies and expertise for a sustainable future.

FMRI provides the framework for securing and delivering a strategic investment in new marine research infrastructure, including autonomous (robotic) and Artificial Intelligence & Machine Learning technologies, that are integrated and interoperable with other existing and planned NERC and national infrastructure.

The FMRI Digital Infrastructure Scoping Study aims to:

- present the current state of the UK marine digital infrastructure ("Landscape Analysis")
- present the requirements that are recommended for upgrades for the future state of the UK marine digital infrastructure

This **User and Stakeholder Research Report** outlines the methodologies, processes, and insights that informed the recommendations presented in the final report.

Participant overview: mixed methods

22

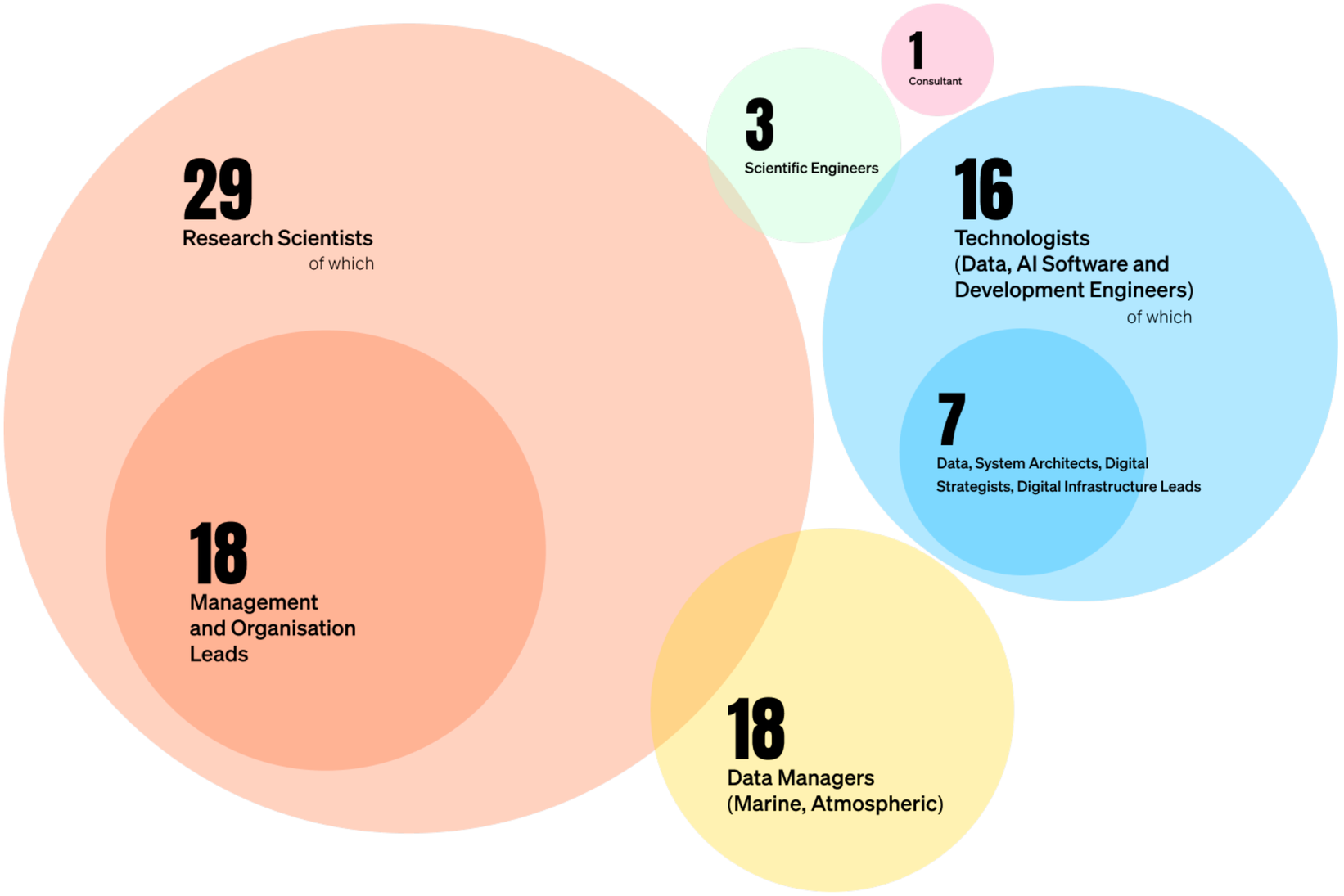
User & Stakeholder Interviews

45

Survey Respondents

Research Scientists include (but are not limited to) these topic area specialists:

- Observational Oceanography
- Physical Oceanography
- Underwater / Ocean Robotics
- Remote Imaging
- Marine Physics
- Geochemistry
- Marine Geoscience & Geology
- Climate Modelling
- Climate Change
- Zooplankton
- Artificial Intelligence



Note: The survey was sent out to a wider group of respondents, however the numbers are based on the responses which were submitted by 14 May 2025.

Research approach: what we did

The study

FMRI

Digital Infrastructure Scoping Study

The report submitted to the FMRI.
A summary of evidence and information which will support the FMRI in decision making.

This report

Moderated contextual interviews

A

User & Stakeholder interviews provide in-depth insight into who the users are, their current **experiences, and challenges**.
We carried them out to collect attitudinal metrics, and to help us understand and identify opportunities for the ‘future state’.

Max-Diff (Best/Worst) Survey

B

Many areas of opportunity were identified.
The survey focused on understanding the *perceived* priority of discovered needs across our sample user-base.

User Journeys, How Might We statements and Service Map

We plotted the current state challenges and pain points across **two User Journeys**.

C

Then we used the opportunities from the research findings to create **How Might We statements.** These highlight and give an idea of potential costs & viability, and map directly to Fig 4. in the report *(“High level architecture for Option E”)*.

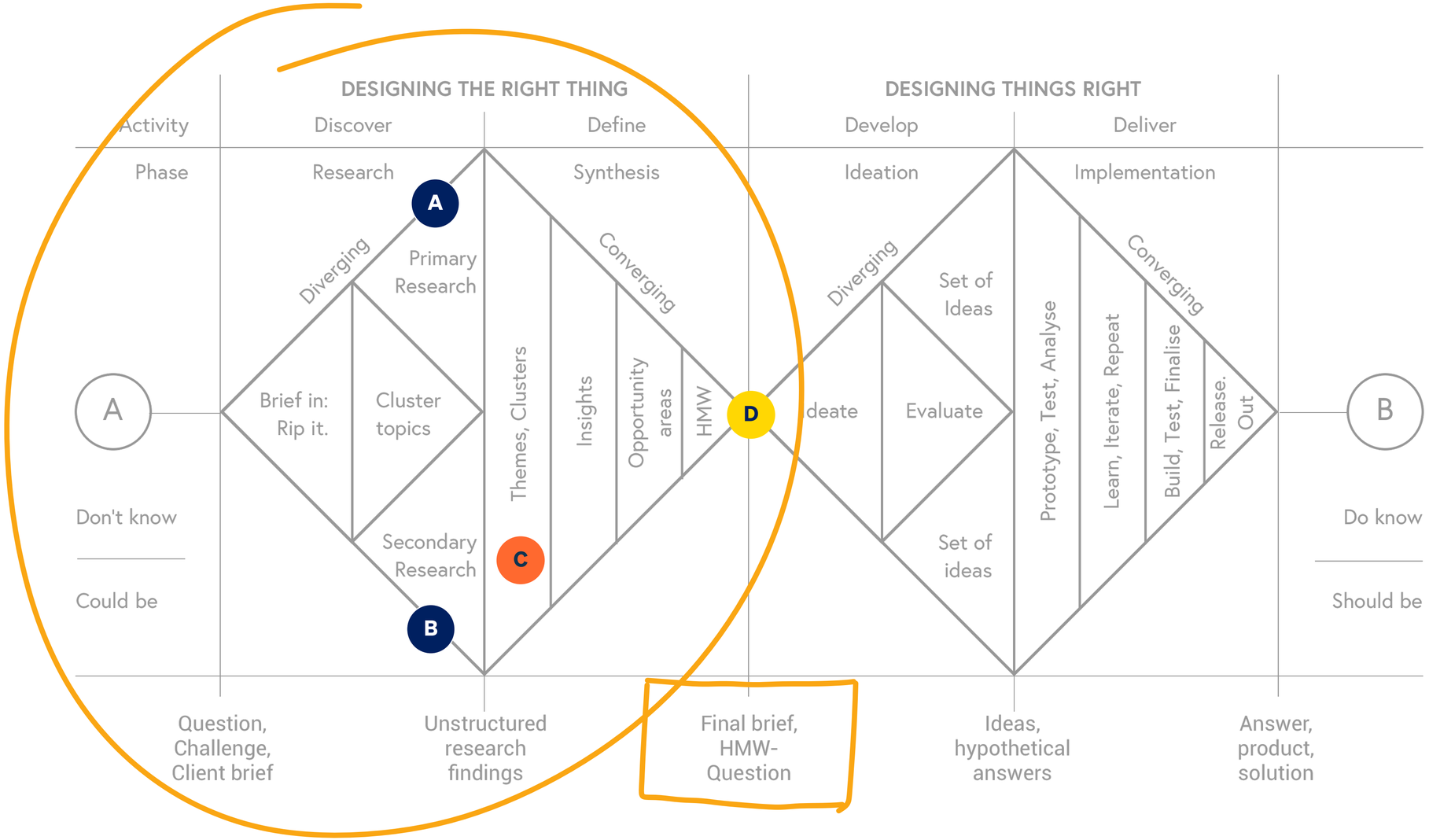
D

These will be plotted across the **potential Future state service map for the final report.**

E

Limitation: This research was conducted remotely. We were unable to observe participants in-situ on vessels or using AUVs due to time & availability (scientific research currently in progress).

Research approach: how these parts fit into the design process



Double diamond diagram - source: British Design Council

Qualitative research

22 User & Stakeholder interviews

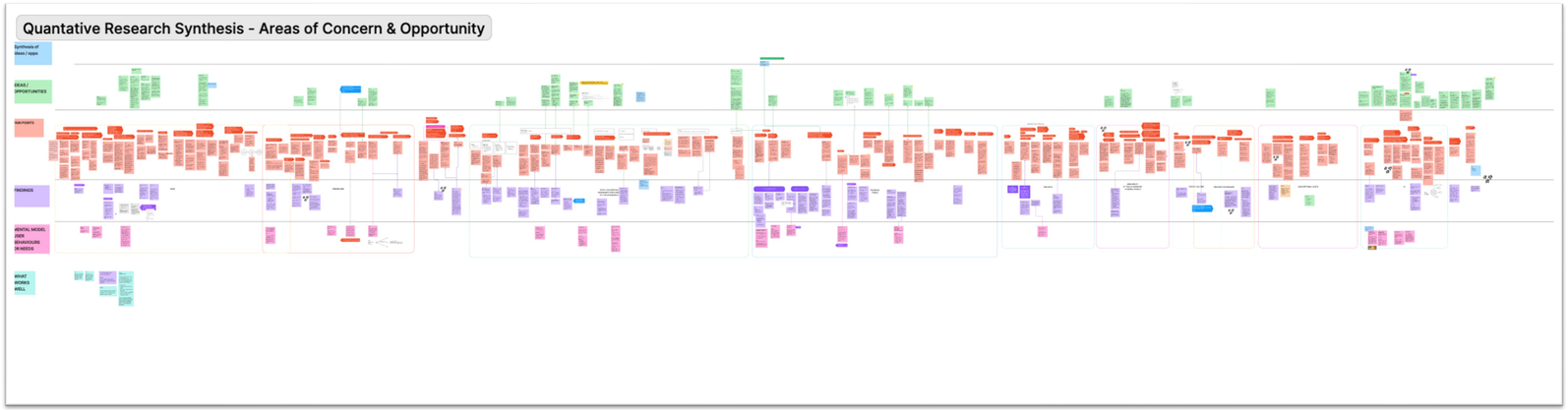
Twenty two moderated online contextual interviews were carried out with users & stakeholders of the current digital infrastructure to map pain points, needs and dependencies in the current infrastructure setup. Our goal was to understand connections and requirements at a higher level – and to uncover potential opportunities for the design of a future digital infrastructure.

Questions focused around key areas: daily user task flows and interactions with current digital infrastructure, technical setup (storage, cloud, AI etc.), skills and gaps, costs & complexities.

Users & Stakeholders work at:

- National Oceanography Centre
- University of East Anglia
- Plymouth University
- British Geological Survey
- University of Southampton
- Scottish Association for Marine Science
- Met Office
- DEFRA - Marine and Fisheries Directorate
- NERC UKRI
- MEDIN
- The Crown Estate
- UK Centre for Ecology & Hydrology (UKCEH)

Each interview was transcribed in Figma after each session (from recordings).



Figma board showing the collected research from all 20 user and stakeholder interviews.

Note: The large number of interviews is to ensure we covered participants across a wide variety of ocean science subject areas, with a focus on NOC vessel and autonomy operators .

Qualitative research

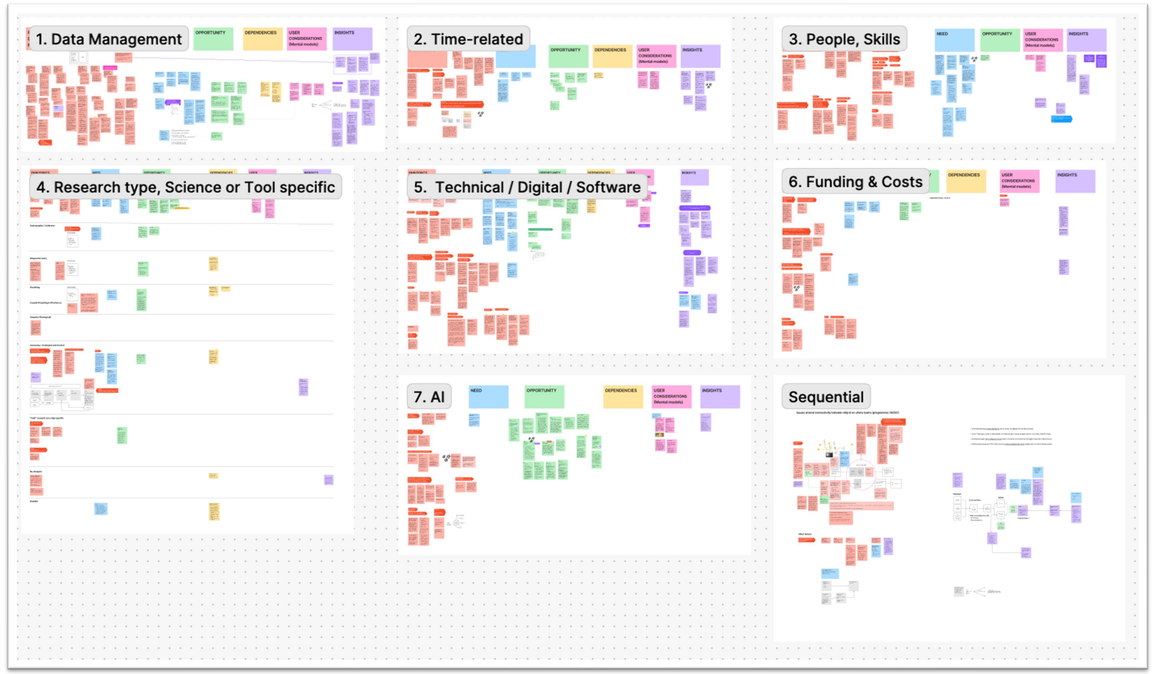
Affinity mapping to uncover areas of focus

Different affinity mapping techniques were used to group the feedback from the user interviews, and discover patterns across the data:

- Sequential (chronology of steps / choices)
- Descriptive (pain points, needs, opportunities, dependencies, user considerations)
- Key words (time, data, people/skills, training, AI, research or science type, technical, digital/software, funding/costs)

These were collectively analysed and synthesised into Key Insights (Page 3) Identified opportunities were used for How Might We statements (Page 17).

‘User need’ focus areas were defined, and the **sequential feedback** was mapped into User Journeys (Page 16 and 17).



Figma board showing the affinity mapped user interview insights in their six thematic groups: Data Management, Time Related, People/Skills, Research or Tool specific, Technical (Digital Hardware, Software etc), Funding/Costs, AI.

Note: No AI tools were used in mapping the themes – this was carried out collaboratively with the NOC Team.

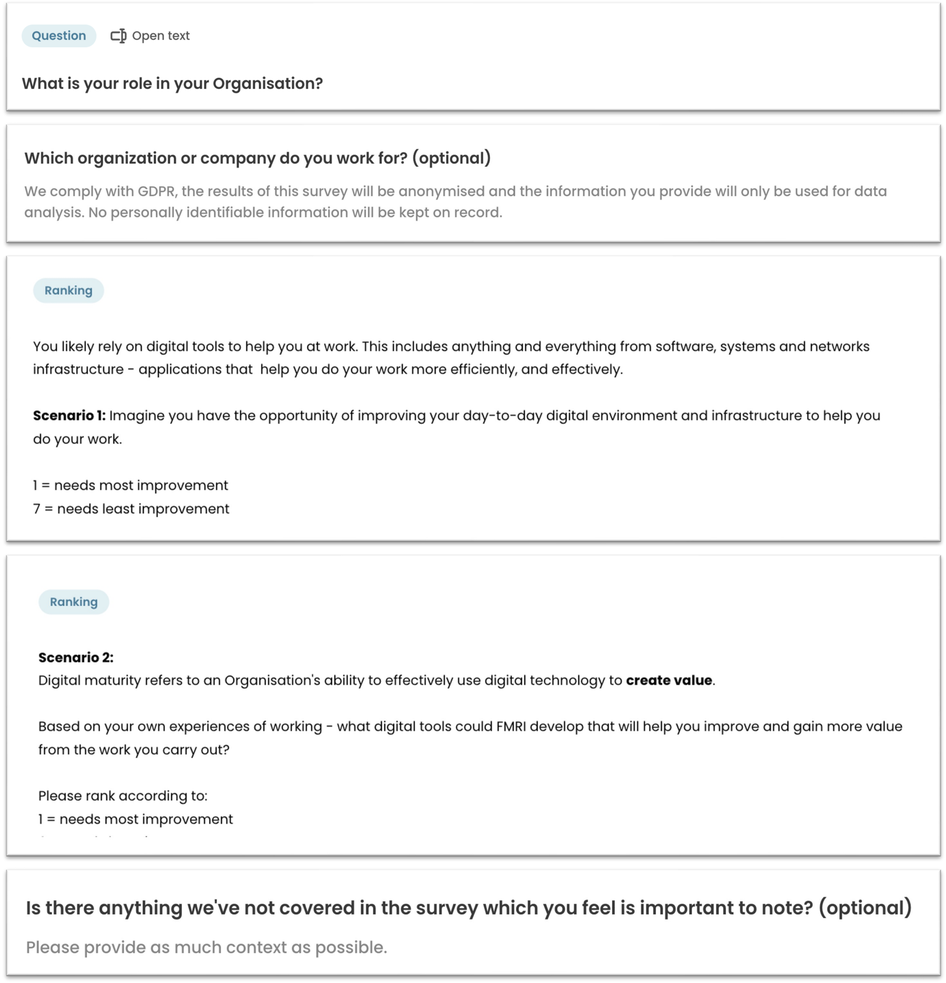
Quantitative research

MaxDiff Survey

Survey is now closed.

The survey is based on the 6 broad themes which came out of the synthesised qualitative findings from interviews with over 14 subject matter experts & users across NOC, BAS, SAMS, The Met Office, DEFRA Marine and Fisheries Directorate, The University of Southampton, Plymouth University and MEDIN. The options are broad and not exclusive – this is intentional. The task is to rank both questions in order of preference.

Aim: to quantify which of these areas is '**most attractive**' or '**most desired' to users** of the UK marine & oceanography digital infrastructure. **This is an attitudinal exercise – and we are using it to gauge user-perceptions and to help us prioritise the How Might We statements.**

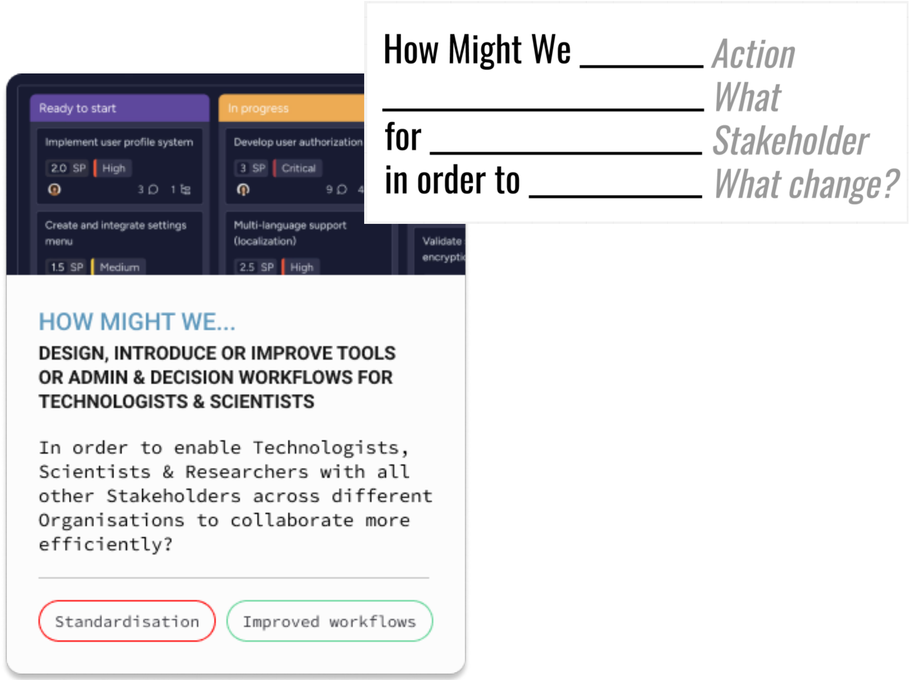


Research synthesis and recommendations

“How Might We” questions

We then wrote How Might We questions to stay focused on the problems and opportunities uncovered during research and ensure we provide flexibility for delivery teams to ideate solutions.

We matched each HMW statement with the relevant area(s) in the high level architecture diagram in Fig 4 in the FMRI DISS Report to indicate where this requirement stemmed from, or which area it would impact (some span multiple areas). The NOC team assigned level of impact, level of certainty and estimated costs for each HMW (criteria: low, medium, high). **This has been shared in the report, for feedback and adjustments from the entire FMRI DISS Team.**



Some background on HMW:
How Might We statements purposefully prevent the inclusion of solutions in HMW questions and – because doing so restricts the pool of possibilities, and fewer ideas can be generated by the teams who will be delivering on the solutions.

These statements all stem from user insights and identified business challenges - ensuring that the potential solutions developed will be grounded in reality.

- A Guide to How Might We statements

Section 2:

RESEARCH FINDINGS IN DETAIL

Note: These are sample selections of quotes from 43 stakeholder user interviews as well as qualitative feedback from the survey (Q5). Full transcripts available on Figma.

The wider ecosystem
Interconnections and overlaps

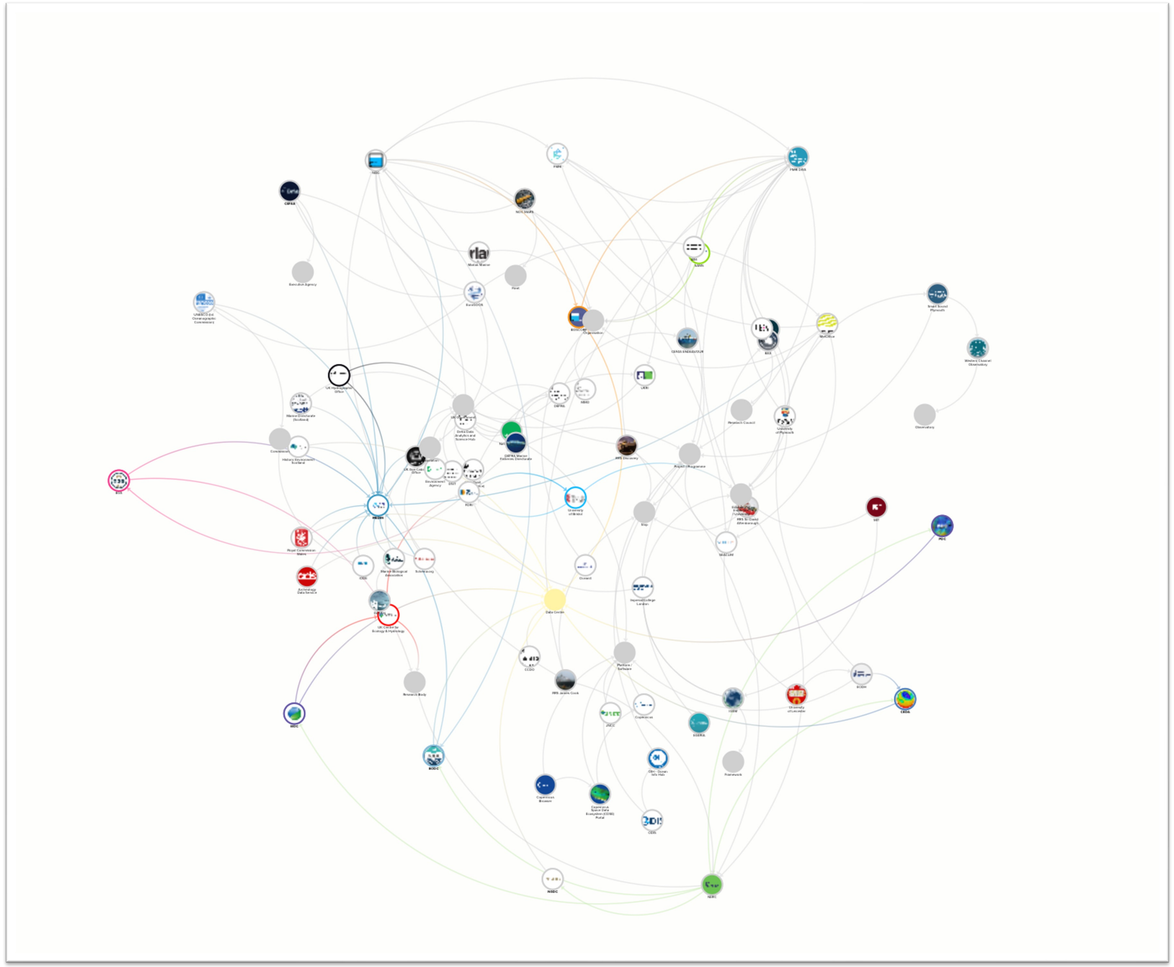
The marine infrastructure digital ecosystem is vast and interconnected, much like the ocean itself.

As with any large ecosystem – for value delivery or service innovation to happen, we need to enable knowledge sharing and partnerships.

This ecosystem map has been created from publicly available data (and is open-source) to help visualise these connections and to create a better understanding of the scale of the network the FMRI will have influence over (see **Page 31**). This ecosystem map includes the interconnected elements mentioned during interviews.

“We’re not a large enough section to function independently – so we’re quite reliant on partnerships…”

Research Interview Participant



USER JOURNEY 1

Cruise/ Vessel expedition focus

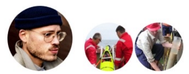
Please see PDF for detailed view

User Journey 1 - Expedition / Vessels

Created by
Ola Podgorska
Lead Digital Design Researcher
National Oceanography Centre

Contact
olapod@noc.ac.uk

Profile



Ruan Ship Science Systems Team Lead

On expedition along with:
Charles, & Georgios Research Scientists - utilising glider (autonomy) and core-sampling tools during the expedition to measure & collect data and samples.

"Our responsibility ends pretty much after the glider data is sent to the Data Centre. Everything up to that point falls with us, and it's our responsibility to get it working..."

Jobs to be done

- Ensure all systems are operating smoothly during expedition. Team is responsible for operating, maintaining, and developing the onboard scientific and IT systems.
- Work with Programme team, to identify cruise requirements for each programme.
- Review post-cruise assessments, to identify and optimise (what we do well, what we could improve?).
- Annual management of maintenance for our systems. Goes into annual refit tasks. Maintenance upgrades.
- Ensure Research which has been planned in can be undertaken. Take direction from Scientific community to spend time / money to introduce new scientific capabilities.

Step	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
Journey point	→ Pre-season cruise planning	→ Prepare ship for expedition	→ Contextualising metadata to gliders	→ Cruise requirement form submission(s)	→ Expedition begins	→ Research underway	→ Dip clear received	→ Ship to shore data collection and transfer	→ Onshore: Data Centre	→ 2 years later - data required for further use
Actions	**Research trip is planned (years to months in advance)** **Using the** Marine Facilities Planning software (MFP), Ruan plans the research trip. "You can make as good an attempt as you possible can, ahead of time, to be able to plan and sort out ***some of the stuff...*** "Unfortunately there are a few excel spreadsheets.... leave them separate, ...its' easier to leave them separate"	**The ship is physically loaded with relevant tools and measurement devices, as well as supplies etc. Manual processes carried out on ship to ensure digital tools function.** "We need to physically get in there to run new cables through. We have to do this every week... everywhere there is a sensor - we need to do this. Not only Science sensors, but the bridge and ship system. Eg: installing new SatCom system last year, this year GPS system needs to be replaced..." *"At the moment all of our risks are recorded in Excel spreadsheets - I feel like there must be a better way of doing this..."*	**Inconsistent metadata & issue logging to sensors.** "In the moment- you are trying to fix the problem. Y***ou need a system that can take complex outcomes of the problem and put it into something that 10 years down the line, can be looking at the data, and see something that doesn't seem right, and be able to match that anomaly, with what actually occurred during those few hours the problem was arising.***" "No equivalent to MERMAN for Sensor Data" "They may not necessarily be accurate. We're moving to a paradigm, where we will have a continuous logging database. Which then increases the importance of the protocol around metadata management. You'd need to make sure everyone knows - looking afterwards - that metadata has changed at some point. " *"Ships have (issue tracking tool) one but it's very poor - not good enough for what is required... useful tool to have - both on Marine IT side, and Science side."*	**All forms filled out on systems to ensure relevant research can be carried out at sea, including Diplomatic Clearance sent to FCDO as a Word Document via email. This is signed off in their Exclusive Economic Zone.** "Lots of different people involved in these things - putting the same information into multiple forms. Doing some sort of environmental impact in, and I'm putting the same information into the Diplomatic clearance forms..." "Dip Clear is out of our control - applied for in time, sometimes granted last minute (eg: day before)"	**Different types of Scientific research will be undertaken on each cruise.** **Simultaneous use of different systems can introduce acoustic interference. This can reduce the opportunity for other types of Science to be carried out.** **Note:** On RRS DA - synch hardware Kongsberg K Sync reduces conflict, but priority also affects collection... "bathymetry ping rates were often <5% of what could have been achieved if operating stand alone..." "There are always trade offs... if you want to do a certain type of science, it would mean you can't do a different type of science - the way the ship is set up..."	Underway and other ship data collected. During the cruise, the Scientist will capture all the research manually (in whichever way works best for them). Post-cruise, a BODC document is used to capture all this once it's been standardised. **Sometimes, metadata is not captured because of all the other tasks having to happen.** **When sensors are changes** When the glider sensor is changed, this will affect the metadata (if metadata changes, continuous logging will be affected).	"We're not sure from our side how the Diplomatic Clearance are being requested - because they don't align with the equipment we've got on the ship - and sometimes the wrong Dip Clear get submitted that don't cover equipment that's actually requested on the Diplomatic Request form - it's not under my remit. Where we get stuck with - my team operate the equipment and we have to check if we have diplomatic clearance to be able to use it - and we come to the point where the Scientist wants something on, but the Captain says no (as no Diplomatic Clearance is available)." **NOTE:** **When speaking to the RRS James Cook PM's, they have mentioned this is a known issue (currently being addressed), and is something that can be accommodated by the MFP software platform (a 3rd party - "they are very flexible to our needs").** **This involves digital development resource & allocation of funds, time & digital development resource. In terms of the MFP, each funding application will keep the MFP running (a portion of each expedition/cruise funding is dedicated to this).**	Currently, until the Scientists come back to shore, the larger the dataset - the more difficulty there is to transfer it. Ship to shore file transfer and management solutions varies depending on Organisation. Private companies like Fugro have their own proprietary solution for this use case but it is internal only (patents pending so we were not able to get more information at this time). At NOC, this relies on the data being transferred (either through satellite / internet stream, or physically being delivered on a drive to the Data Centre - depending on size and time this will take to transfer). This is not 'future-proof'. **NOTE:** There are other 'off the shelf' solutions available for external file transfers - but these solutions are primarily for file transfer via HTTPS or FTP. They may not have metadata level automation. *"The only time this isn't the case has been where data volume has been so large that it was faster and more robust to simply post a HDD to a partner (one recent example was 67b of data that they could only access at ~ 1.5 Mb/s over ftp). "* *"It's what Researchers do, but again on their own computers - there's no easy way to see the data..."*	Data centres then have the task to standardise the data received from the Cruise. However there are also issues with this - in terms of the volume of historic data which has not yet been standardised, and the data which is not currently linked to cruise data (eg: samples). **Currently, there is no unified software platform facilitating seamless communication, data transfer, or interoperability between the three key entities involved in marine scientific expeditions:** (1) Scientists conducting research and collecting data on expeditions using autonomous vehicles and onboard tools (2) the Programme Team acting as the operational interface between the scientists and the National Marine Facilities (NMF) or chartering organisations (3) Data Management Centres & Standard Bodies (e.g., BODC, Copernicus, MEDIN), which are responsible for reviewing, processing, validating, and archiving collected data. ***SAMPLES*** "There's no link with the data we hold back to the Cruise data...it doesn't link to the other data that's being created as part of the expedition..." Currently data is not collected/organised and linked by location or expedition." ***3rd PARTY DATA - eg: Wind farms*** "The PI contacted the Developers, they sent this data to the PI... then to me..."	**The cruise happened approx. 2 years ago. The Researcher who carried out the work has since left the Organisation. All the visualisation tools which were used on the cruise are separate and there is no overall context of an issue which occurred during data collection which is now presenting the current AI & Data Modelling team to spend considerable time trying to figure it out.** "Historical data is NOT AI ready, not machine readable... it's the going forwards..." *"In the moment - you are trying to fix the problem. We need a system that can take complex outcomes of the problem and put it into something that 10 years down the line, can see something that doesn't seem right, and be able to match that anomaly, with what actually occurred during those few hours the problem was arising."* "We start a new database every cruise - so that makes metadata management relatively easier. However we're moving to a paradigm, where we will have a continuous logging database. Which then increases the importance of the protocol around metadata management. You'd need to make sure everyone knows - looking afterwards - that metadata has changed at some point. We don't have the tool to do that at this point." **[See Step 6]** **Existing data visualisation tools on vessels currently:** • Electronic chart display and information systems (ECDIS) • Variable and flexible display configurations (configurable software) • Real time event log info • Interactive maps with spatial information • Different user permissions • Security (bridge data vs. other data) • Separate spreadsheets - especially with Admin processes
Touchpoints	Marine Facilities Planning software Various Spreadsheets (Excel etc) 	Cable transits inside ship (Physical) Sensors (Physical) Various Spreadsheets (Excel etc) 	Data Centre websites (eg: BODC) Various Spreadsheets (Excel etc) Sensors (Physical) 	Environmental impact forms Dip. Clear Forms / System Data Management Plan Documents MFP Data / Documentation (Marine Facilities Planning & Inventory) 				Email Hard Drives / USB Sticks (manual) Satellite transfer (ship to shore low width orbit satellite links - utilised for large data vol streaming - WIP)		Communication (phone / conversation) Various Spreadsheets (Excel etc) Email
Pain points	**Pain point** What's in the way of me achieving my goal? Disconnected systems and info sources Inability to predict weather, and factors which may influence the expedition which are not yet known.	**Pain point** What's in the way of me achieving my goal? **Design of vessel not always optimised for efficient digital infrastructure utility.** Ship ergonomics do not consider increasing needs of digital connectivity infrastructure (eg: getting to cable transits between decks and bulkheads, via fireproofing and waterproofing. behind cable trays to update sensor tech). Risk & compliance processes are all manual (excel)	**Pain points** What's in the way of me achieving my goal? No 'contextual' or 'issue' tracking metadata is captured automatically (from sensors linking to metadata collected). Has to be collated manually across multiple spreadsheets, systems and notes. Lack of digital tooling / systems to support this: *"We don't have the tool to do that at this point."*	**Pain points** What's in the way of me achieving my goal? Forms in different systems - relies on team to maintain order and progress applications Disconnect between various systems (not automatically connected - especially across Continents / Globally - eg: Diplomatic Clearance processes differ depending on EEZ).	**Pain points** What's in the way of me achieving my goal? **Negotiations, with Scientists, Captain and the Cruise Team.** Compromises made (based on decisions depending on negotiation / constraints of what is possible during cruise). No support tools for this. All manual / discussion based / email etc. Usability of MFP is not optimised for self serve - eg: equipment use: greying out one piece of equipment if another is selected. Relies heavily on expert system user knowledge.	**Pain points** What's in the way of me achieving my goal? Still awaiting Diplomatic Clearance! Continuous logging of metadata - protocol around metadata isn't as well ironed out / considered on vessel when sensors are switched / changed out. Those we spoke to do not have an automated procedure for this (eg: NOC, BAS, Fugro) Manual capture of data by Scientists means that often standards are not maintained, consistency is lacking - until it's passed onto the Data Centre which applies these standards (MEDIN). Often Metadata is left out completely (left to last or de-prioritised).	**Pain points** What's in the way of me achieving my goal? 3rd Party system (MFP) - feature updates & interconnectivity need to be scheduled in and costs accounted for. Interconnectivity only applies to those Vessels which use the MFP system (not extensive to wider ecosystem of vessels).	**Pain points** What's in the way of me achieving my goal? Bottleneck in collaboration and analysis: other scientists or technicians may not have access to the data if it is isolated on one person's drive, slowing collaborative workflows, allowing real-time analysis or quality assurance (or building tools to enable real-time data management to occur). Loss of data (risk) MFP (or other Vessel planning software if outside of NOC/BAS) and Data Centres servers / databases are not linked (data transfer happens through manual intervention - except for Autonomy - See UJ 2).	**Pain points** What's in the way of me achieving my goal? Lack of historical digitisation of sample data, and no link between samples collected and relevant cruise. Fragmented data flow: Data collected in the field is often transferred through manual or ad hoc methods, increasing the risk of data loss, duplication, or errors. Delayed Insights: The time lag between data collection and processing/analysis slows down scientific output and operational decision-making. Disconnect between data coming in from any other external sources.	**Pain points** What's in the way of me achieving my goal? Currently no issue tracking system cross-org blocks data-driven decision making. Ability to provide data onwards for visualisation / direct data access is not as efficient as it could be (signposting, access etc). Loss of contextual understanding over time. Data silos - no clear visibility of what is available 'cross-Org' without investing time into search
HMW examples	1 2	68 14	48 16 4	1 25 27 28	3 18 54 55 59	15 16 48	4	20 29 57	45 46 52	47 48 66

USER JOURNEY 2

Autonomy / Piloting focus

Please see PDF for detailed view

User Journey 2 - Autonomy & Piloting

Created by
Ola Podgorska
Lead Digital Design Researcher
National Oceanography Centre

Contact
olapod@noc.ac.uk

Profile

Catherine - Autonomous and Robotic Systems Engineer & Glider Pilot.

Working with:
Ruan - Ship Science Systems Team Lead
Chris & Georgios - Research Scientists

The potential of AI tools to support Piloting:
"If you add 30 gliders, and want to deploy them all in the same place, to do the same thing. Can we do that with our current capacity? No - but we probably don't need a massive increase... piloted in roughly the same way - an easy fix.

But if somebody came along and said here are 30 more gliders, we're going to have two gliders on 15 different projects, and those projects are going to change every year, and they're all going to have different sensor configurations, we (currently) have no way to cope with that sort of scale..." **[How Might Wes - 4, 8 & 9]**

Jobs to be done

Autonomous underwater and surface vessel operations including piloting, maintenance, troubleshooting, and fieldwork.

Vehicle preparation, maintenance, integrate new sensors, calibrate and enable new glider behaviours, mobilisation.

Pilot the long range fleet (primarily gliders and USVs, but also ALRs when necessary) as part of the team's rota. **Time away on field work (both shore-based and at sea:** Required (on occasion) to work away from home during field operations and out of hours whilst piloting. Occasional seagoing time on Natural Environment Research Council (NERC) and other ships.

Work with scientists and researchers to help design glider operations to best support their research projects. Focused on solving challenging long-term marine science problems, underpinning international and UK public policy, business and societal outcomes.

Step	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10
Journey point	→ Planning and programme design	→ Preparation of sensor (Slocum Glider)	→ Slocum Glider in mission (data transmission)	→ Mission issue / incident handling	→ Satellite transfer / auxiliary data	→ Data interruption or failure (error state)	→ Recovery (return to shore)	→ Post-mission data extraction	→ Data transfer to data centre	→ Data delivery to scientists
Actions	**A request will be made through to Marine Autonomy as part of any upcoming grant application (parameters of which will have initially been discussed with the Scientist/Researcher depending on what the science requires). NMF Head of MARS / Engineering Manager Approves Technical Viability of Request. This will also need to be approved by NERC. The Principal Investigator will then submit the grant application, and once confirmation of grant funding is received - the NERC Programme portfolio is set up (portfolio requests are based on prioritisation criteria).**	**Gliders fitted with requested sensor payloads, ballasted and undergo pre-deployment testing.** **The gliders are shipped to deployment location (on ship or shore) and once configured, are then deployed at sea or at land (depending if on ship or shore deployment).** **Each glider has a digital logbook for pre-mission status and ongoing documentation.**	**Glider pilots are available 24/7 to ensure operations. Once the glider has been depoyed (from ship or shore), Pilots currently pre-set the route and monitor it during operations to maintain optimal parameters during mission.** **Experienced pilots are always in demand, and Trainee pilots require quite a lot of support (training happens on real missions).** **The Glider begins collecting and transmitting sensor data:** • Near Real Time (data sent through satellite comms when glider is in the water, depending on the glider model, but the more widely used (Slocum) sends decimated data. • Real-time monitoring tools are active where available. • Metadata is collected but not automatically linked to the vessel metadata for sensor issue tracking. This is currently done manually.	**Example of 'typical' issue during mission:** • Glider loosing a wing • One of the payload sensors breaking / being broken **Field crew (e.g. ship-based) will intervene to retrieve, repair and re-deploy glider—even outside regular hours. The mission continues, and data transmission resumes.**	**Data is streamed through satellites to the Data Centre. Satellites also record Level 0 data (terabytes per cruise - for eg: plankton bloom data at sea-level surface, mapping sea grass, carbon offsetting etc.). This is currently backed up to Liverpool servers at NOC.** **Volume: Level 1 data is sent to NOC - 3 files per track, 40GB for each day, 3x days per campaign.** **Time to process: 1:1 - Times it takes to record, time it takes to process.**	**If an issue occurs - for example, the glider data download fails on a weekend, the Data Centre is not staffed at the time - so there is no immediate response.**	**The next step is for the vessel and gliders to return to shore. The instruments are retrieved and offline data downloads initiated.**	**After each mission, the raw data is manually downloaded from each instrument.** **If the instrument was not part of the MARS / NOC Marine Equipment Pool, the relevant organisation will retrieve their own instrument, and data.**	**The raw data is then passed to responsible Data Centre (BODC, or other). The manual cleanup and Quality Control (QC) processes begin.** **NOTE - IN PROGRESS: Data model for sensor data descriptions, deployments and datasets [NOC / BGS / CEH etc in discussion, NOC iADOPT standards, Vocabularies]**	**Final datasets sent to Research Teams and Research Scientists.** **Remote Sensing Data: Level 2 product - server in Southampton 40 mins per file, 14 files per day**
Backing quote from research	**Current deployments per year:** • 20-25 glider deployments • 1-3 ROV Campaigns per year • 2-10 AUV Campaigns **Battery life is always a consideration in route planning (if battery life is used in an optimal way, more data can be collected)**									
Goal	What is the aim / goal or purpose? Successfully scope and plan out the annual deployments based on the scientific research which has been granted funding. Ensure battery life is optimal for science / data collection required on glider route.	What is the aim / goal or purpose? Successful preparation of sensors, testing and deployment of glider in determined location.	What is the aim / goal or purpose? Enable the collection and streaming of data from glider and sensors on board (delayed mode or real time data).	What is the aim / goal or purpose? Successfully fix any issues in data transmission when they occur, and ensure this gap can be 'accounted' for / contextualised in future (as needed).	What is the aim / goal or purpose? The goal is to ensure that satellite data is delivered to researchers as quickly and as completely as possible—whether in near-real-time or delayed mode—so that it can be effectively used for scientific research.	What is the aim / goal or purpose? Correct any data transfer issues as swiftly as possible to re-connect the data flow and prevent / reduce gaps in collected data.	What is the aim / goal or purpose? Retrieve instruments and prepare / start data downloads (MARS team @ NOC)	What is the aim / goal or purpose? Successfully extract data from sensors and autonomous devices / vehicles.	What is the aim / goal or purpose? Successfully retrieve datasets collected from sensors and send to Data Centre. Pusher App (push data to Data Centre eg: BODC) Slocum recovery data - .ebd (Engineering), .sbd (Science) + Cache files CTD's - Digital package emailed (ASCII file)	What is the aim / goal or purpose? Scientists (need: processed data) • Get insights from data collected Customers (need: raw/original data) • Understanding processing steps • Understand what the data was used for initially Govt (need: link to funding) • Demonstrate value of data
Touchpoints	C2 Platform (users require UI relevant to autonomous vehicle in use) Various Spreadsheets (Excel etc) Communication (phone / conversation) Email	Glider Digital Logbook - "GlideLive" Fitting with payload (Physical) Slocum/Glider/AUV (Physical) Various Spreadsheets (Excel etc)	Enable the collection and streaming of data from glider and sensors on board (delayed mode or real time data). Data Centre Monitoring Dashboards (File formats depend on platform - BODC ERDDAP, MetOffice .covjson generated) C2 Platform	Data Centre Monitoring Dashboards C2 Platform Microsoft Automate (Vessels) Risk - Spreadsheets (Vessels)	Sent via FTP (login / password only) ***NB** Potential security risk*	Data Centre Monitoring Dashboards C2 Platform IMS / Sharepoint / GIT (Vessels) Email - Cross-Org Comms *"Communications are only done via emails, no teams messages - it does hamper efficient comms..."*	Slocum/Glider/AUV (Physical) SD card physically removed from the glider (eg: by MARS operator)	C2 platform (MARS) Inventory Management System / MFP (Vessels) Various Spreadsheets (Excel etc)	Data Centre systems/software (eg: BODC) used to archive data (.RXF format / NetCDF)	
Pain points	Most non-MARS users come to C2 with an existing familiarity with a specific platform type. (Original goal of C2 and some of the comms-backbone work that underpinned SoAR and MAS-DT was a aspiration to create a platform agnostic system. Now work is underway to ensure platform-dependant UI's are available in CT as the UX of the platform was not considered in the first MVP iteration: *"The last one I did in 2022 that was a sea-glider, and the whole plan was to test on the C2 interface... and I had to arrange data transfer so I could get it on my own website to pilot the glider... because I found it impossible to use the C2 tools at the time...It didn't have enough plots... I know it was on the casts to be developed..."*	None reported, but full service design study has not been carried out. **Future potential** *"We've always relied on bottle data. Taking it to the lab... we haven't got real-time biogeochemical data traditionally. Gliders are starting to help us with this. I think that is the most exciting thing in terms of observations... that can transform what we do - alongside the modelling and AI - the new observations are really powerful for us..."* • Also see ROV Core samples - HMW 61	Pilots can adjust waypoints or sampling strategies only when the platform is at the surface and has connitivity. Standard gliders currently lack AI/ML capabilities to adapt routes mid-mission (e.g., optimising data collection or battery use). **Note:** NOC is developing this autonomy for both onboard and server-side systems. Currently sensors such as CTD's and Gliders are not connected within a network. This requires pairing & matching datasets manually, which is time consuming for Scientists & Ship Crew. If fleet size increases, need for trained pilots increases - supporting piloting software (eg: AI) Demand on Experienced Pilots (planning & training, monitoring etc)	No automated flagging or context and sensor-specific metadata linkage (vessel & autonomy metadata) when failures occur. *"I know across NOC there are different things - MARS use JIRA, NOC IT have started using a new system...CPS is just starting to use a new service... ideally if we had one [joined up] that would be great."* **Current projects & future potential** Sensor metadata management open source systems (federated - linked data stores / services)	Transient nature of satellite data. Data products are dated so MARS team have had to find workarounds to handle spatial data sources (when addresses of data products change over time, when it's important to the user to identify how old the data is). This introduces complexity for any analysis automation.	**Large volumes of data transfers is not sustainable / standardised / scalable currently** *"It's not a commercial system - it can take a long time - standard output - 10-15 minutes per data file, a data file may be 150 MB, 15 minutes of radar data - but it's not needed to (historically), when it's just 2 of you..."* *"It will run through the processing even if it has gaps - in the science, basically."* No formal process for archiving or standard data handling and reporting across sensors. This results in one-off systems with little reusability/scalability.	**Gliders do not work within 'office' hours** Disconnect between the Data Centres and the Autonomy Operations. **Note:** Gliders are partially observable systems. NOC has been working towards AI fault detection for remote systems. Currently reactive, need for optimisation, to prevent data loss occurrences. **Future potential** *"Link positional inputs from the gliders on CT, into the Ship's bridge AIS display.* *Whenever the glider surfaces, the system sends a position to the server in the lab, and that will be communicated to the AIS tracking system on the bridge - so they can see the location of vessels and surfaced gliders. As you have more and more platforms in the water in a small operational space, that situation is a lot more challenging."*	Disjointed view of autonomous vehicles nationally - unless within MEP, and logged / maintained by NOC / MARS. **Future potential** An open source, National equipment registry / dashboard to provide real-time equipment status (and location, based on security level clearance / checks) or open source 'extension/API' of the MEP.	Disconnected systems / data silos. No shared visibility across the UK of all deployed assets unless they're part ofinterred into the Marine Equipment Pool. C2/MARS, IMS / MFP (Vessels), Other systems across UK used by independent fleets / not part of MEP. Usability / UX of the C2 platform does not yet meet all user requirements (see Step 1 pain point) - backlog of features, lack of resource (UX/Development etc). **Current projects & future potential** Sensor data management platform. Cloud based / MQTT, atuomated processing, QC, infilling, UI, API etc). Source: ***CEH work - Matt Fry*** **Current projects & future potential** Stores and patterns for large gridded time series datasets. Eg: Pangeo community approaches (JASMIN, NOC + OTHER ORGS)	Cybersecurity risks (depending on dataset sensitivity & value) Size of data (transfer processes are currently dictated by what's possible based on network / satellite connectivity, location, etc. - real time data transfer from vessels not currently feasible for larger datasets, hampers potential of real-time data transfers in future. Manual QC processes (higher risk of user error, system disconnects introduce 'red herrings' etc.) **Future potential** Leading systems for storage and management of image, video and UAV data. Standards based on RDA group, involving BAS + Others. **Current projects & future potential** Automation of QC processes, open source tools for data QC, machine learning / AI automation Source: ***CEH work - Matt Fry*** **Current projects & future potential** Mechanisms for automated publication and update of published sensor datasets into NERC data centres.	Time it takes to get insights from data *"(Speaking about MARS) We know where all the platforms are, but not consistent information from them. Even that interface (C2), it gets difficult to assimilate that information [and get insights from it]. It's time consuming to do..."* Lack of contextual ship to sensor metadata: if any gaps in data are identified in future, it's a manual, laborious process to try and understand what happened contextually – involves back tracking/ manual investigation and connecting different documents, logs etc. *"When we did the commission for the data management systems - we went out to the community and got feedback from them - and broke it down into aspects of marine data system. We had a workshop and asked them to prioritise... the underway data model came out on top, and the contextual logging events was the second - and we've spent less time on that. So if I were to spend money on, I'd spend it on that..."* Occasional failures during transfer or processing - identifiable after full transfer occurs (reactive) *"It will run through the processing even if it has gaps - in the science, basically. So that's happened on a couple of occasions, when I've had data uploaded through BODC..."*
HMW examples	10 3	31 56 60 61	4 5 6 8 9 21	45 46 14 9	37 55	4 19 30	20	7 19 23 40	43 37 39	18 11

INSIGHT 1

Inconsistent foundations across the digital & data infrastructure

“

But maybe we should think of stuff in a digital sustainable way, **maybe there should be systems that communicate or connect like puzzle pieces - and maybe that should be discussed across multiple organisations**. . .

Because, for example the non-commercialisation aspect, I think that will change across UKRI, and for then to be able to be ready and say “that was a software, that was used internally, it was open source, now how do we use it externally...like the software I showed you - could that be run off a tourist vessel?... maybe, but we don’t know how we would facilitate that at this time.”

Interview – British Antarctic Survey

INSIGHT 1

Inconsistent foundations across the digital & data infrastructure

"Tools, data APIs, libraries etc **that interface well with standard user software** such as Matlab, Python, R, GIS (ESRI/QGIS) to name a few." [Survey Q5]

"**A lot of what we've got is legacy**... the data centres are working to update, move with the new way that data is being collected - making that available in new ways" [06:00]

"There is almost a leap that needs to happen... **how can we make sure it's collected (as FAIR data)**?... there's so much data - **it's all collected in different ways spreadsheets sometimes...** some of it is moving into real time..." [08:20]

"The way it gets solved [currently] is people ping emails back and forwards, where is piece of Data X from Cruise Y, and eventually you'll find someone who knows - **which is a terribly inefficient way of doing it.**"

In terms of "importance", data quality is always top of the list. It is no use having AI scrapers doing clever things with the data if it's junk underneath. And there is a *lot* of bad tide gauge data in the last few years. The gaps are there for a reason. [Survey, Q5]

"**Evolution happening - not very quickly.** Partly to do with funding - there is the will to do it… it's the cost of that... vast amount of data - shift into the new – some **things it may be easier [to do] when you are starting from scratch!**" [17:00]

*"Is there an API standard for marine environmental data? **Our pilot projects have found the answer** to this question is "yes, the OGC Environmental Data Retrieval API standard is probably suitable for most marine data applications… however, we will need to develop specific use cases to establish where investment should be made with regards to increasing accessibility to marine data. **We are really keen to know if anyone else has been working on the same API standard that we have been exploring or indeed if you have trialled alternatives**.' [Presentation shared]

[MEDIN have trialled an API standard (DASH - 3 data centres implemented the Open Geospatial Consortium (OGC) Environmental Data Retrieval (EDR) Application Programming Interface (API) Standard (OGC-EDR-API) as a pilot].

"**Don't have the technology yet - in terms of doing that on a fixed basis...** people can go do it on their phones..." [Not recorded]

"**Standard understanding and best practice around large-scale data storage infrastructure would be quite amazing...** we're dealing with - because as you're aware - different levels of data coming from new data collection. How do you store it as cheaply and efficiently as possible etc." [12:46]

"Obviously, there's varying technology stacks depending if you're in Azure, if you're Microsoft, Amazon... **understanding what's best for my purpose?**" [13:09]

"**One of the biggest challenges we face is very much around standardisation.** We're getting a wide variety of physical environment, biological environment, datasets coming in... while MEDIN plays this role as a standard unifying standard body, **nonetheless it is a challenge to get lots of private contractors to utilise these, and deliver a standardised package**..." [03:07]

***"There's no link with the data we hold back to the Cruise data**...it doesn't link to the other data that's being created as part of the expedition..."

Currently data is not collected/organised and linked by location, or expedition – no links across the NERC landscape…" [41:39]

"Groups doing their own thing...how they manage, how they standardise data... **we used to be 9 different organisations - they've all got their own history and do their own long-term monitoring. We've been a single entity for 25 years now, but we've never had a push to 'this is how you should do it Org wide'**…each team does their own thing, there's efficiencies to be gained - standardised, improved practices... we want to get to a point where we have a more integrated network..." [Interview]

***Note**: Several projects over the years attempting to fund. Not considered 'successful' eg: NERC DigiRep: BAS/BDF/BOSCORF and MEDIN Pilot API standard study "MEDIN doesn't have much development budget at all, at the moment it's come to a halt [9:55]."

"Things always crash, go wrong - none of these are operational services....it's a dice roll" [Interview]

"Making it FAIR, **sharing it, is tacked on - when everyone is running out of time and money** on the projects." [Interview]

"Supporting a promoting a federated approach to data systems and governance is for me the way forward. **The key to integration and standardisation is to promote mapping and interoperable / flexible standards at the interfaces between these entities** to ensure more automated and streamlined flow of data and information without sacrificing specialist needs." [Survey Q5]

"Having something behind the scenes - **this is the formats we use, this is the standards we should apply to...**" [NEED - 42:10]

INSIGHT 2

Disconnected systems = data silos. Value in the ‘bigger picture.’

“

At the moment, from my perspective at least - I see things as very disjointed. I see NOC, the centre of excellence - generating all the control systems for Autonomous vehicles. But it’s then very difficult to know - okay that group owns that asset, NOC owns a lot more assets - and then how it’s all shared and planned...

And then the planning side, through MFP, it’s almost impossible to get itineraries out working between BAS, NOC and NERC. Having a digital infrastructure like that that’s maintained, internally within NERC is key. If there is a new feature - like what we’re generating - we could slot into the MFP, but currently we cannot, because it’s too close source software wise…

Interview – British Antarctic Survey

INSIGHT 2

Disconnected systems = data silos. Value in the ‘bigger picture.’

“It’s great to be thinking about all the fancy stuff, but **these are only going to work if we get the basics right**… the model is only as good as the data that goes into it.” [27:52]

“**There’s no equivalent** to the Marine Environment Monitoring and Assessment National database (MERMAN) **for sensor data**…” [21:45]

“it can take a long time - standard output - 10-15 minutes per data file, a data file may be 150 MB, 15 minutes of radar data… it’s one to one – times it takes to record to time it takes to process… (***systemising it***) **is not needed to, when it’s just two of you….**”

“**I think making the data accessible - especially so that people can see what data looks like** - other Scientists… for example on the Robotics side, some people may not quite realise the amount of data you can get.” [05:29]

“These technological developments are really exciting. **Are these going to be available to other users once ready via the marine equipment pool?** (AtlantiS project)

“That huge availability of the physical samples that we have - the amount of digital data associated with them is very limited. 20-25% analytical data, basic images 40% …**a very small proportion have been digitised**” [17:55]

Although possibly already covered by some of the themes highlighted in the survey, it is worth explicitly highlighting the requirement for international and cross-discipline interoperability and standardization of our digital infrastructure.
The requirement (scientific and sometimes legal) **to feed UK data into international repositories needs to be taken into account** when developing new tools/systems.” [Survey, Q5]

“At the moment… we can deliver the data to you but its not in a particularly high-tech or advanced way. You can download it. **Our issues with standardisation, mean that we can’t provide an API of all the bathymetry data to whomever wants to query it…**” [15:48]

Investment and endorsement of up to date and widely accepted standards in fields with commercial data collection occurring (ornithology, geophysics etc) will make a significant difference, allowing the academic and private sectors to align.

Approaches to what we store/how we store vast amounts of data would also be of value.” [Survey Q5]

“One of the things I end up doing a lot of - is reporting (project management). I do always think it’s the most inefficient way of doing things**…. I feel like I’m just giving the same information in multiple different formats, and portals**… [11:35]

”Getting information from X number of people - getting that information from them is pretty hard…” [12:56]

“I would say **MEDIN is the closest we have to it, but as an Organisation it’s probably quite heavily under-resourced,** which makes it really hard. It’s the sort of thing - **to get it through to policy at a Government perspective, you’d do a lot of good for that, but there’s no other unifying force**.” [05:28]

“If you’ve got silos…we can’t then coordinate it with everything else…” [04:18]

“**Make sure we can interact with other programmes going on internationally…** how do we then make sure whatever infrastructure we’re putting in place isn’t siloing us from working with others…” [35:54]

When it comes into a digital infrastructure context, the framing of that - that means our big challenge - is taking, oh god, we’ve got lots of stuff to standardised, **we want to put it out as one big standardised API… it’s not something we can physically do. We don’t have the resource, or the tooling, or the expertise in some cases…**” [03:40]

“In a magic future world, where we can do what we want - **something where you have a UK wide, endorsed standard, that’s kept up to date… but the ability to take that onto a platform, and then share it** with all of these people looking for it - via an API, spatial mapping service - that is the dream end goal, from my perspective.” [04:21]

“Pretty much everything - if you get a spreadsheet or a MATLAB file from a Colleague, you pretty much have to figure out what they’ve done, in what order, and what each column means… if it comes from a Data Centre it’s all standardised…” [10:06]

“**Each participating data centre experienced different issues associated with applying the standard,** such as time outs associated with large datasets; differences in the interpretation of the standard which made writing standardised call the to each API more difficult, the potential loss of detail/nuance in making marine data available in this way, **a potential disconnect between data and metadata that would need to be addressed**.” [Presentation shared]

Similarly, **initiatives to align data centres and archives through interoperable systems and shared standards will greatly improve usability.** Enhancing data access through machine-readable metadata is another important step in this direction.” [Survey. Q5]

”Therefore, it’s critical we can show progress and know we are addressing areas of knowledge gaps or inefficient ways of working, or ineffectiveness.

To be able to do that you need a really comprehensive overview as to what that evidence portfolio looks like - and if you don’t have a firm grasp.. then it is hard to move towards continual progress.” [05:28]

“Different types of data collected… **we keep things that provide long term value,** and at what level – what is useful for people now, and what will be useful to people in the future”

Note: Value is currently determined by individual Data centres, many use the NERC checklist. This checklist has a clause (5.) about re-use and storage formats but **these are in the Appendix, and not part of the checklist itself.**

“**It was noted that storing copies of datasets with differing levels of automated processing would have an impact on disk storage**” [RRS DA Data Systems User Workshop Report, BAS]

SUPPORTING EVIDENCE: INSIGHT 1 & 2

How do we quantify the value of data & research from conception through to production?

Currently, the Ocean Science ecosystem produces a mass volume of high value data with weak connections to its 'quantifiable' value. Projects get funded, Research is carried out – often resulting in digital components – but these results are not linked, nor organized in a way which would enable someone to demonstrate and provide transparency and traceability across an end-to-end data flow at the push of a button (and even less so across academic and commercial sectors*).

Data Centres assign data value at intake level – but digital considerations are not specifically featured on the Intake sheet (a stand-alone document). The 'burden of proof' lies on the Data Centre employee or initial contributors – who may not have full visibility or even knowledge of the potential of it's reuse over time, filing the sheet on a server somewhere… yet this data has exponential reuse value – especially in newer, AI training use-cases (and importantly, if reuse is to be undertaken ethically).

A wider framework allowing a wide variety of users to easily interconnect and demonstrate the value of the research will simultaneously *"improve the quality and provenance"* of data, while also creating funnels for potential future opportunities.

Note: Solutions exist (eg: Zenodo, the CERN open-source project, assigns a cite-able DOI to raw data) to support in solving this issue, however even if open source tools like this are adopted, the Governance aspect (**See Insight 4**) remains: encouraging and supporting researchers in using tools like this consistently and reliably, and ensuring data is cited properly in their own papers – across the UK, and Global ocean science ecosystem.

…by tracking of data from conception through to production (full data life-cycle)

"It's very difficult to do - with this kind of thing - the need to not store everything - it will be dependent on THEIR needs and the project, rather than someone else may come along and what they need… who's to say what is valuable?"

"We make sure we don't remove RAW values…so people can go back to the RAW sensor data."

[CEH interview]

*Investment and endorsement of up to date and widely accepted standards **in fields with commercial data collection occurring** (ornithology, geophysics etc) will make a significant difference, allowing the academic and private sectors to align.

Approaches to what we store/how we store vast amounts of data would also be of value. [Survey, Q5]

"This would be applied at all NERC Data centres – CEDA, EIDC, NGDC, PDC and BODC."

"Process is part of the outline data management plans – we have a "**Value checklist**" **(see image below)** – done at Data Centres themselves… partly for us to work out within our data policy BUT **some of this could be different - because some of it relates to long term monitoring…**"

[12:00]

Important criteria: These are primary criteria and answering 'Yes' to at least one of the question below should probably result in selection for retention.

Policy
Are the data a result of full or partial NERC funded activities?
Do the data fall within the selected Data Centre's Collection Policy? If no – refer to NERC Data Coordinator or pass to the correct data centre.
Scientific or historic value
Are the data a unique unrepeatable measurement of the environment?
Do the data have a broad geographical or temporal extent that makes them useful to others?
Do the data have historic value i.e. do they represent a landmark in scientific discovery?
Do the data include changes in processing methods, new standards or set any precedents?
Do the data support current projects or trends in science?
Are the data likely to meet the future needs/direction of the scientific community?
Do the data contribute to a pre-existing collection?
Is there potential for re-use of the data?
Are the data likely to be cited or referenced in a publication?

"Protocols and processes to **protect the longevity of data products both in terms of storage and access, but also in terms of documentation and updating** (are needed)." [Survey, Q5]

"We are very much **intending to always deliver value for money** - obviously we are public funded and therefore **we need to demonstrate to treasury that we are doing the best we can with those public funds.**" [05:16]

Some data end-reuse cases identified:
AI model training
Analysing priorities
Analysing risks
Conducting research & development
Conducting statutory reporting
Informing coastal planning decisions
Informing marine planning decisions
Informing operations
Managing marine resources
Raising awareness & education
Validating data from other sources

Also see Page 35

MEDIN study – 120px
Re-usability of marine data across the following sectors:
• Academia / Research
• Commercial industry & Consultancy
• Government and Maritime Authority

INSIGHT 3

Disconnects and time considerations: field and office

Field work is expensive, and currently some tasks which are expected of Researchers & Autonomy Pilots (eg: training, administration) could be carried out in more cost-efficient ways. The disconnect between ‘Glider or Research hours’ and ‘Data Centre / Office hours’ has not been factored into the design and delivery of the services (ie: value delivery).

“For most of those steps - because the team in MARS are piloting the gliders out of hours - if that process falls over out of hours - if they can’t directly fix it, they can find someone to fix it on the weekend.

But if it happens at BODC, it’s out of hours (which is fine, nobody is paid to do it - we understand why that happens).

There is no funding to pay BODC staff to be on call just in case a server falls or fix whatever pathway processing problem - when the Customer sees it - of course they’re going to see it - **we’ve got two days without data - why is that? Because it fell over at 6pm on a Friday afternoon**... [18:00]

“It’s expensive to send people to sea - they should not be doing university teaching or admin at sea..”

“Sometimes the notice will come through on a Wednesday, to say we’re going to do the maintenance this weekend - and that’s fine, we appreciate the fact that they’re trying to do the maintenance on the weekend, and it’s spoiling someone's weekend to do the job, so it doesn’t impact everyone's work - but the gliders and the vehicles that are out in the water, are still out at the weekend…

…the days when they are being deployed or recovered, are whenever the weather is best – it will be a Saturday or a Sunday, just as likely as any other day.

So sometimes we only get a couple of days notice that the servers are going to be down at a critical point when we need communications to the gliders calling in via satellite. Sometimes we don’t get enough notice to be able to work around that – that can be a problem.” [12:20]

“We’ve been doing the project for three years, and I don’t think we’ve had a Christmas break where something hasn’t gone wrong. **And as soon as it happens over Christmas break - there’s not really an option to get it fixed.**”

“I don’t think we can fix without funding”. It is a frustration to the Customer – not to us, we understand it. [19:02]

“Takes a lot of my time to train people - when I’m training young pilots, junior pilots... **getting texts and emails at 10pm at night...**

“We have to do this every week - getting to cable transits between decks and bulkheads, via fireproofing and waterproofing. behind cable trays to update (any new tech/sensors)…”

“Which we did manage, but it wasn’t trouble free… **we had data glitches, and things that didn’t happen – and of course they happen on a Friday evening**… we could make it work, but it wasn’t trouble free…” [11:11]

“**Critical elements like data governance and standards are eating into our staff’s limited time**. This reduces time available to our limited resourcing to engage with the real value proposition…”

“I can tell you straight away - I am currently organising this research cruise and **I’m completely bogged down by paperwork... I’m not having any time at the moment to think about the science we’re doing**.” [12:34]

“At the moment assimilation corrects forecast data… model tends to drift away from forecast.

Particularly in spring, our forecasts are too high than what they are in the real world.” [16:27]

”Difficulty getting access. It’s just a time consuming thing...land owner agreements. **We’ve got a 5 year delivery - you can spend a year or two just getting permissions.**”

It’s horrendously expensive to put stuff out in the field... although it’s cheaper than putting stuff in the water, but it’s a lot of time.

And we became expensive... the economies of scale of working with a provider really works out well for us.”

“our problem is so simple.... there’s no automatic backup of data from when it’s collected... it’s all done manually... ties up people’s times and their machines while that’s transferring...” [27:10]

“we have NONE of that at the moment... efficiencies would be gained for the entire staff.”

”that would save a huge amount of our resource to do other things - produce more data, whatever...” [27:55]

“lots of small wins we could have... gone from a manual to automated approach - they’re massive blockers that we have”

“Difficulty getting access. It’s just a time-consuming thing...landowner agreements. **We’ve got a 5 year delivery - you can spend a year or two just getting permissions.**” [Not recorded]

INSIGHT 4

Digital infrastructure investment needs to include investment in collaboration, people and skills

Adoption of new tech does not only rely of funding of the hardware and software but also the communication, governance, training and education aspects. Saleability relies on investment in people & teams who are required to operate at an increased scale, as well as the tools to empower them.

"One of our short comings - identified a few years ago. **The number of skills required is too broad to be able to focus properly on all the things.**" [08:21]

"Even just within our own field - **it's very hard to know who is doing what, what that evidence base looks like** - and therefore how we can continue to improve it..." [04:48]

"A broader issue of **how do you keep people in research & innovation - especially if the skills they have are quite good for industry - and they pay a lot more...**" [28:44]

"**Training and communications**, getting sector buy-in."

[Survey, Q5]

"**IT are not currently prepared for a massive roll out based on what we'd like / future scope**. There can't be a reliance of on-prem servers…" [Workshop - RISC Project: Cyber Physical Infrastructure]

"...ocean is a 'small bit' of what the MetOffice does - most of it is atmosphere, **we're not a large enough section to function independently - so we're quite reliant on partnerships,** especially with NOC. We use the same ocean model (NEMO, configurations). [25:43]

"Annoyingly, the delta costs are also difficult in terms of "transient people" - so you **have to allocate "time to learn" and still expect to lose a lot of those "trained people" because you can't keep the delta people.**" [Workshop - RISC Project: Cyber Physical Infrastructure]

"One of the challenges of technology… gliders is a really good example – to build gliders you need, someone who really understands machine learning, you need someone who understand how fronts form in the ocean, and you need an RSE, and a data architect. **All those groups of people need to work really closely together – to be able to develop a system. these technologies are too complicated for one person to do**…

…**how you exchange information between each other and how you're efficient, and how you move forward with a bit of an understanding of each other's role - being able to communicate with each other is a bit of an open challenge as well.**" [43:12]

"**We need to have better digital strategy; to maintain people in place** - for example - our software engineering teams here are based on research grants, not national capability. [42:10]

"**I would target the Governance**... **if you know who to interact with - you know how to in the best way - you know when, to to ensure you achieve each others' shared ambitions**, as well as individual aims or goals..." [41:37]

"It allows the discrete areas of focus - be it academic, government - **if we all have finite resources, this is what each other are doing - and this is how we can address these overlapping interests...**" [42:52]

"**you can't have too many experts in the same area. Especially with things that are running 24/7, for months at a time... you need people who don't necessarily have ALL the expertise, but the tools to do that...** we've got to develop them and just make them available..." [10:07]

"I guess the thing I always pay attention to is the complexity, demand, availability conundrum. **Having more than say about 10 people, makes the team more difficult to manage...**" [07:07]

"**Reducing barriers to adopting AI techniques**—through better data formats, interoperability, and clearer documentation—is also key**. Efforts to improve onboarding materials, tutorials, and community guidance can make a big difference.**
[Survey, Q5]

"Ensure business-as-usual tasks aren't under rated… **we can't afford to upgrade every work flow to new technology**…"
[Survey, Q5]

"**Up-skilling staff.** Despite Government calling for staff to be upskilled in AI, it **is still a struggle to secure resources to train staff.** If you could provide training or access to training courses…"
[Survey, Q5]

"There's no point in throwing money [just] at buying more autonomous vehicles. **You also need to make more money available for Scientists to make small grants, to try and increase that pool.**" [21:52]

"Not just time on cruise, **staff-time as well to make sure analysis is factored in** - the understanding is **where the value comes** from." [Workshop - RISC Project: Cyber Physical Infrastructure]

INSIGHT 5

Some types of oceanography research are more suited to digitisation/systematization, AI and autonomy.

Contextual Logging Events

"When we did the commission for the data management systems - we went out to the community and got feedback from them - and broke it down into aspects of marine data system. We had a workshop and asked them to prioritise… **the underway data model came out on top, and the contextual logging events was the second - and we've spent less time on them. So if I were to spend money on, I'd spend it on that…**" [46:22]

Good suitability High impact

API Standards

"**Other API standards may be a better natural fit for some types of data such as Sensor Data.** "

[when speaking about the Open Geospatial Consortium (OGC) Environmental Data Retrieval (EDR) Application Programming Interface (API) Standard (OGC-EDR-API)]

• Different API standards for different tools / platforms

High impact

Vessels

• Simultaneous use of different systems (acoustic interference)
• Competing expedition priorities
• Logistical processes
• Ergonomic access / digital ship infrastructure design

NOTE: "Off the shelf" systems are not integration-optimised / may not allow it

"**One thing which would be really nice Link positional inputs from the which we have talked about in the past. gliders on CT, into the Ship's bridge AIS display. I have seen this implemented on a NATO vessel run by the Italian Navy. A system - wherever the glider surfaces, it will send a position to the server in the lab, and that will be communicated to the AIS tracking system on the bridge - so they can see the location of other vessels around in the water around them, but also the gliders that have surfaced - which means those locations don't have to be inputted. It takes about 5 minutes. (Now) it just takes forever - a very manual process!**" [33:00]

Good suitability

Ocean & Coastal modelling

• Large size / volume of image, video and acoustic data
• Need to compare observational results against model results
• Improved usability could potentially reduce data load on servers (reduction of duplication of data)

"**they all have the same data… depth, latitude, longitude, and time… that's all I'm interested in, in working out where those CTDs were nearby the Gliders. I also know that the data I've download and am storing, someone has already downloaded before… but I don't want to ask the entire Company "oh did anyone download Wind Speed for North Sea in 2023… but I know people have done… some data I know is available in the modelling group - but there's probably a lot of replication of data…**"

Good suitability High impact

Autonomous platforms - training

• Distributed fleets require more advanced piloting expertise (training & tools)

• Opportunities for real-life training are also vital, and some things cannot be replaced by digital processes (although training could potentially be augmented with digital tools).

"**For most of the platforms I use, you put them in the water. You can do theoretical exercise and bench testing, but the real problems happen at sea - and not on the bench - you need to do it for real!**" [30:07]

More complexity High impact

Sedimentology

• Currently no link with the data back to the Cruise data
• Sample data doesn't link to the other data that's being created as part of the expedition within MFP.

"**it's quite a controlled - standardised enough of the types of samples, the way it's processed to get it ready for an analysis, the data that comes off - it would be quite an easy problem - for somebody with the right time and skills…**" [27:10]

"Geological data seabed and subsurface data should also be considered/included., and not only ocean column data. There is a marine data part of the National Geoscience Data Centre (NGDC), however funding is proportionally less compared to a solely marine data centre such as BODC. **There may be more opportunities for collaboration on AI and spatial modelling techniques and infrastructure, storage, and also on data collection.** [Survey, Q5]

"The area of digital surrogacy of physical samples. Marine geological cores are invaluable types of marine samples that allow scientists to study the seafloor and subseafloor environment. Access to the physical samples themselves is often restricted. However, with **the ever-expanding range of analytical techniques than are employed to measure various properties of these materials and the range of imaging techniques that can capture image data, it is possible to envision a future infrastructure where it will be possible to produce digital 3D (and 4D) surrogates of the samples. These samples could in turn be stored in digital geological core repositories which stand in for the physical sample repositories themselves**… [Survey, Q5]

Good suitability High impact

Autonomous platforms

• Occasional issues with data transfers & lack of contextual issue /error handling (between ship & autonomy metadata)
• AUV maintenance, calibration of sensors & admin optimisation.
• User need to ensure familiarity of platform type (Slocum, AUV etc) and Operational software (eg C2) match (UX/functionality)

"**I think there are some subtleties around how you match platform capabilities with autonomy capabilities. Rather than have one feature-rich autonomy solution there could be many specialised ones.**"

Good suitability

Hydrographic surveying

• Narrower groups
• Doing same things repetitively on cruise
• Repeated from project to project

"**We spend too much time, manually transferring data around…an area to focus on if we are to systemise things as this has a lot of 'ease' of automation, systemisation due to the kind of data which is produced (similar every time).**" [33:51]

Good suitability

Biogeochemistry

• Multiple groups, lots of different tasks
• More complexity to productise processes as they are varied
• Potential for doing things in different ways (eg: using gliders to collect Nitrate & PH data)
• **Heavily reliant on other teams: require real time biogeochemical data (not historically available - gliders make it possible)**

Reliance on others

Benthic oceanography

• Environmental factors (smaller time windows to collect data)
• Competing demands and priorities on a single vessel during expedition
• Lack of data around Antarctic.
• "Data domination" - the total data created during a cruise, will be completely dominated by video.

"**Currently this an issue with the Benthic data which is sitting there - no one is looking at it regularly. If we want rapidly developing science at NOC - need to provide access to be available relatively quickly…I'll go for lunch, and it will be available for me…" not IT will put it there in a week or so**…""

Good suitability More complexity

Re-analysis

• Relies very heavily on datasets from other Organisations (eg: Copernicus / BODC etc.)
• Cannot do the work without this data.

Reliance on others

Metadata

"We've started doing some stuff, particularly around metadata creation. Nothing groundbreaking, **Basically using a large language model to try and help key metadata fields on MEDIN metadata.** That was step 1 for us, when as soon as we have a functioning form of a chat model, based on a large language model. [08:38]

High impact Good suitability

Note: This is not a complete overview of all the types of ocean science – but only features the areas which were mentioned in the sample group (67 individuals).

SUPPORTING INSIGHT 4 & 5

National capability to develop digital infrastructure, and under-represented skillsets and gaps

Currently, digital infrastructure is developed on a 'ad hoc' basis, depending on individual research grants within Organisations. Levels of digital maturity differ across these Organisations, and departments. Digital strategies are being put in place but are at different stages.

There is a wider need for national capability to develop digital infrastructure – and to introduce skillsets which are required to build and maintain a cohesive infrastructure.

Aside from this – certain skillsets within Research Science are in high demand and need more focus.

Skillsets more in demand/required for further digital transformation:

- AI / Machine learning specialists
- Entry level: numerical modelling / data handling
- Digital Research & Design (UXR / UX / UI)
- Digital Product Owners
- Technical Architects
- Quality Assurance (in digital product development pipelines)
- QC / Data quality with metadata focus (vessels)
- Benthic Data Analysis
- Genomics

NOTE: Although not specifically mentioned as a topic in the discussion guide, IoT technologies were not mentioned in any of the 22 contextual interviews, or the Survey Q5. The assumption has been made that this is an unexplored area of knowledge.

AI / Machine learning
"Under resourcing… up-skilling Scientists to use these new processes…[22:00]

"Theres quite a shortage of skills of people who are Machine learning AND marine, and an even bigger shortage of skills if that becomes ML and a particular species of fish…[42:36]

"AI team…they're swamped, they're busy, under resourced, getting anything done requires writing a grant…the will is there, but more resource required…. they'll be really enthusiastic but there won't be time to do it…we need experts, we need collaborations"

"In that respect, its not necessarily that all of us have to be trained in how to use it, or be experts in it - but more funding for getting those cross links - for getting people to talk to each other - I like this idea of having an AI fellow coming into a group - and potentially vice-versa, Scientists going to work with AI teams coming up with solutions - that would enable those discussions happening a lot more…" [25:51]

Digital Product Design *
"Design of decision support tools - not a well represented skillset in marine research." [29:30]

"It's very hard to design infrastructure set ups or datasets if you don't understand the ways it will be used - the end user… if you get that correct - everything else falls into place - but it's also the most challenging piece." [41:37]

"The ux we tend to need is at technician level - its functional, the outward function stuff - the event logging will be different, because it does require end-users to be involved with it. *But most of the ux is for technician / administrator type - but it does not need to wow people."*

***NOTE**
This capability exists in some, but not all Organisations. Digital maturity is inconsistent and varied across Oceanography institutions [see Insight 1].

"And previously, when we've been running at capacity, which we're not at the moment - we've had 2 devs, a QA, and a 50% of a design guy doing the platform. **What we refer to as a Squad… we chuck a problem at them, and they are able to solve it**" [09:39]

Technical Architecture
"There is no time for infrastructure thinking - it ends up being more reactive."

Quality Assurance Testing
"Testing should be planned in." [Workshop - RISC Project: Cyber Physical Infrastructure]

Autonomy Piloting
"Horrifically people poor. We can get bits and pieces and little gadgets, but we don't have people to run them. That's probably the case with digital infrastructure as well. It's the staff time that's most concerning.

Passive acoustic equipment and operators required (but will introduce new complexity)
"There is no trained MMO onboard the ship… courses are not expensive."

Numerical Modelling
"We're really struggling to recruit the right skilled people in numerical, physical oceanography, numerical modelling especially…" [44:00]

Environmental genomics
"Importance of, and data governance, of environmental genomics"

Excerpt: "To date, no well-established model species that have been sequenced and for which standardised 'gene-chips' are available"

INSIGHT 6

To enable innovation, room for short term experimentation, collaborative project opportunities & training need to be fostered.

“Two approaches - MFP is too big to fail internationally, and it’s used massively - but we need to claw back some ‘developer’ aspects to areas where we can slot in new features, where we don’t need to go back and forth with the external... or, we know that the MFP is what it is. And NERC considers what it’s got in house - as expertise - a software engineering pool at NOC, there’s mapping experts at BAS, there’s other software engineers at BAS - **having a consortium where people work together**."

Interview – British Antarctic Survey

INSIGHT 6

To enable innovation, room for short term experimentation, collaborative project opportunities & training need to be fostered.

"There was a lot of trouble shooting – by the end of the deployment it was working smoothly – but it took a lot of work to set it up. **It wasn't an off the shelf solution. It certainly wasn't at the beginning, but it's there now for others to use.**" [11:11]

"Tried to get into using radars on different kinds of vessels, especially on Autonomous Surface vessels - that can carry a radar. **We've bought one of those, we just haven't had the funding to move into doing that... but it's still an idea... that was around 2020....**" [04:47]

"**We consider our physical instrumentation a critical pieces of the digital infrastructure continuum** - they are what take the physical sample and turn it into something that is digital..." [08:17]

"Difficult for us to have cost and scope for use of cloud services - more a process than a technology thing. But if you're using Amazon - you get a hard bill in pounds - which has to go into a project proposal. **So if it's on site - somehow, it's a bit softer - there's a bit overhead… somehow some other funding turns up - and it somehow links - it's a less of a 'hard' cost - other ways of managing it.** [17:55]

Note: Systemic - currently it's easier for Research institutions to get funding for on-prem over Cloud, because of the way grants are provided.

"I think it's being used by a very small group of scientists. **If the aspiration is to draw more people in... a small grant type thing, specifically to use for small grant projects, that's the way to draw people in. That's a resource the community need...**" [22:55]

"You pretty much have to have funding to get anything done - **there isn't any slack in the system - so doing things like Pilot studies is really hard - there isn't really time to do a little practice project...**" "Proof of concept", short terms study to say "I can do this / use the tech / done some work to integrate - something to prove your idea is feasible… wiggle room is gone - 10 years ago - had room in funds to scope out additional pilot projects. Now not possible. [26:59]

"What I would actually like to do – **I haven't found a project to do it with - convolutional neural networks...** what I want to be able to do is to throw my satellite images at a CNN, and for that to determine what areas are permanently land, permanently sea, and anything in-between....I think I'd be able to do it.... [21:49]

"**a whole lot of science that needs to be done on a shorter time scale - but if the capability isn't there** – you can't really do it." [15:00]

"PhD students can get plenty of exposure. At UEA we always ensure students get an opportunity to go to sea. **It's harder at undergraduate level… that's tricky. We're really struggling to get people on field courses.**" [44:43]

"I wanted to know what it can actually do - I've used other types of gliders, but **as soon as you have something different - you don't know what it can do. So you don't want to start proposing something - if you don't know what it can do.** " [21:11]

"You need to know the real-world experience of using this stuff. **There needs to be mechanisms for getting people to try stuff.**" [21:52]

"We have what we need and its pretty straightforward...that's once something is routine - **but when it's new it takes a while to set up.**" [13:25]

"Proof of Concept studies often need lower-level resources at shorter time periods. They reduce risk - ensure no practical issues, it's plausible…need a small grant type system to test ideas / equipment (no longer available).

At University it's "very hard to do that"

"To increase scientists in leading more research - opportunity to 'try' gliders, slocums in a 'scientifically smaller' way…" [20:00]

"In general, enable partners: focusing on developing the governance, training rather than the applications on behalf of the organisation, nurture our ability to create and innovative our own apps." [Survey, Q5]

"Not a lot of people trained out of University... it doesn't exist, you **need to be trained on the job - especially with the newer instruments...**" [28:07]

"Potential researchers who might be interested in Robotics, **but be too afraid to look into it, maybe not quite realise the potential with it...** it's an aspect that is underdeveloped."

Software / SaaS costs

"Software subscriptions are stifling/hampering science discovery & innovation = prohibitive costs to Organisations which don't have the same funding /costing models as scientific research" [Workshop - RISC Project: Cyber Physical Infrastructure]

"This is one of the first things we'd put at risk - because it's not part of the core delivery - unless we can find other ways to support it - through commercial revenue or... we are very aware...but it's less than ideal what we have at the moment." [40:00]

"There may be **more opportunities for collaboration on AI and spatial modelling techniques and infrastructure, storage, and also on data collection.** " [Survey, Q5]

"I think there is **scope for more and better use of machine learning in all sorts of ways in what we do -** probably choosing where to send gliders would be part of that... it's really quite early days - **and we're still experimenting...**" [07:10]

INSIGHT 7

Thinking styles and existing constraints need to be clearly understood and respected, to manage change during digitisation.

Start with service design: when introducing new processes or infrastructure – this needs to take human behaviour into account, and not work against it.

Organisational moves from MATLAB to Python have not factored in the move of past research and training / upskilling

"Also, I have a huge amount of work done in Matlab for data display and analytics. I can work efficiently and well in Matlab. I can't write well in python or web languages, so it's only helpful when offered assistance within that frame: **"We can help, you just have to rewrite your 27 years of code in python first" is not helpful (and is something I have encountered).**" [Survey, Q5] |

"I don't have all the coding knowledge yet, **something I need to learn, and lack of time...it's such a big job. I'm sure I'll manage to do it, but I just don't have time right now... I still need to do it - we need to transition from MATLAB...**" [14:46]

Data values differ across different teams, and organisations - incentivise metadata input. Make it part of the process.

"We don't all have the same data values. **A lot of people don't want to share. We need to win them over, if we want to have people sharing, and being honest about their data, and logging things. I don't think it can just be a top-down** – this is how we do thing – because then what happens is people will play a little bit, but tuck some stuff down the sofa…"

"It has to be about building data culture – unfortunately**, this project is hampered by the fact it's NOC centric, because a lot of these data culture things are driven by things outside of the buildings.**" [NOC DAWNS Project Prototype Testing feedback]

Consideration for differing levels of data security required

"Sometimes the embargoed data still has some metadata available (eg: will be available in X time), sometimes it's kept completely restricted…"

Desire / need for policies on AI use

Policy on Artificial Intelligence. Decision to switch off ChatGPT - legal, power usage etc… but there has been a realisation that there's a balance to be made - **we have a policy on that... people are starting to realise if you use them well, they're a really good thing...**" [25:18]

Clarity of definition of roles and responsibilities (digital project governance/ product ownership)

"At the moment, critical "elements like data governance and standards are eating into our staff's limited time. This reduces time available to our limited resourcing to engage with the real value proposition…"

"We had a lot of meetings – different, conflicting objectives… Oxford - proving they can pilot glider more effectively, NOC trying to create UI for Scientists to control the glider, Scientists trying to do interesting things, and MetOffice had a contractual obligation to deliver certain work… that functionality is almost complete - but not quite there. "

Mental Model consideration – AI & Robotics

That's not how Autonomy people view it… what happened with both Organisations, is if the glider went out of the box, they stopped sending instructions... apparently that's part of tradition of robotics space.

What I was assuming, the glider would try to navigate from transect, move to trying to navigate back into the zone. This was not addressed in scope of either project. Big potential for future work. In reality, Pilots had to intervene (timely & labour intensive) - but it could have been handled by Autonomy…" [MARS Interview]

Example of two users with opposing 'data sharing' mental models, working together collaboratively:

USER 1: SCIENTIST

JTBD / Mental model

- Collects marine imagery & data
- Provide / enter metadata identification for images (when time)
- Everything is peer reviewed. Things are only exposed to "public view" once absolutely ready. (Feeling: sensitive to criticism)
- Lack of trust as people can steal each others research / work.
- Embargoes prevalent.
- Culture of collaboration differs (some people do more collaboration within NOC, some do more work externally)
- Potentially less exposure to digital product development & design

USER 2: DEVELOPER

JTBD / Mental model

- Want to use imagery for developing pattern recognition model using AI
- Culture of using GIT. Open source is an accepted format. "It's normal to share."
- Open source: drafts in public domain, people can comment on drafts / code reviews etc.
- Embargoes not seen as desirable (want to get the information as soon as possible – embargoes prevent this).
- Culture of collaboration differs (some people do more collaboration within NOC, some do more work externally, more familiar with digital design & product processes.

INSIGHT 8

Increasing need to connect outside of Organisations with “Non-Ocean Science” communities – and balancing this with security needs.

We’ve been running workshops, **and it was quite interesting to find what people want from data platforms… they are interested in API’s - lot of interest… developing new flood risk mapping. There’s a close community there - pre-formed community – from events – researchers, flood consultancies, DEFRA…”** [CEH]

“One area that I think is worth highlighting further is **how we can improve inclusivity and the overall user experience in digital environments.** This includes **identifying and removing barriers for users with a wide range of skill levels. Engaging more closely with the community—through co-design and regular feedback—can ensure that tools and platforms truly meet users' needs.**

While data and tools may exist, it’s **crucial to ask whether they are easily discoverable, findable, and accessible in ways that align with how users actually want to engage with them.** Efforts to improve onboarding materials, tutorials, and community guidance can make a big difference. “ [Survey, Q5]

“I have a lot of graphics that would be great on NTSLF.org, to improve the availability of real time, archived and processed data to the public and other research scientists.”

[See example]

“It's a great automated graphic, I use it regularly to check on surge risks and would like to refine it, but it's not even properly linked from the main pages. It would be useful to have someone with significant resources and web expertise to develop NTSLF in partnership with sea-level and surge experts. We could make it far more accessible (eg visual tools with screen readable text) and have a lot more information we could be displaying. Yet it's essential that the site remains robust through fallow digital resource years, when there is no funding available**. It mustn't go the way of the anytide app, which was a brilliant project for a couple of years but failed for lack of maintenance funding to keep up with operating systems**. NTSLF has survived largely due to its simplicity. [Survey, Q5]

Note: See Insight 6 - Software / SaaS costs]

Although possibly already covered by some of the themes highlighted in the survey, it is **worth explicitly highlighting the requirement for international and cross-discipline interoperability and standardization of our digital infrastructure.**

The **requirement (scientific and sometimes legal) to feed UK data into international repositories needs to be taken into account when developing new tools/systems.** [Survey, Q5]

”One of the main areas we would like to improve **is that better relationship between Governance and Academia -** to understand where that wider science evidence base can exist beyond our immediate field of view…” [04:26]

“Lot of value… bridging the gaps between science outputs, data outputs, technology **and people who actually make decisions (the “so what” side of things).** [29:03]

“Find the data they need, access the data they need, more importantly maximise the impact of the data they create… so it doesn’t only live on their hard-drive and only they know what it is - but it actually is integrated into other teams in PML, **and also externally to UK partners, different Research Orgs throughout Europe, throughout the World… can collaborate with other researchers to drive science forward…”** [02:60]

“We haven’t touched on **- the international perspective – make sure we can interact with other programmes going on internationally… “**

“How do we then make sure whatever infrastructure we’re putting in place **isn’t siloing us from working with others…”** [35:54]

“The UX we tend to need is at technician level - its functional, the outward f`unction stuff - **the event logging will be different, because it does require end-users to be involved with it. “**

“I think security is going to play a bigger role in the future. And I think having a system where you can ‘hide’ assets - and a system that can run, and also leverage information from an academic standpoint is interesting as well…”

[35:16]

“However, it is **the next step that will make the real difference to non-expert users.** This is **when a national, regional or global hub sets up a user-friendly interface or portal to support the user.** Both steps need to be completed before life is easier for the dat`a user.” [14:00]

“It’s a separate community to Research community – policymakers, CEFAS, N.Ireland and Scotland (devolved)

Nutrification assessment - traditionally used ship data (legally required for UK and EU) **“Can’t use Sensor data from CTDs / Gliders – there’s no processes in place”**…the group is set up to identify solutions, most likely a digital solution is required… `

“Develop improved public visualisation tools of NRT data…”

[Survey, Q5]

ON NATIONAL CAPABILITY AND UNDER-REPRESENTED SKILLSETS / GAPS

“**There’s very few people in place across BAS specifically... that can do that.** The one aspect is within the Mapping team. This is the Sea Ice Information Service... a toolkit that’s built that works on board the SDA. There is expertise within mapping...but those expertise have a wider impact across aspects like FMRI. **And systems like these should not be thought of in isolation, especially if they are being adopted already.**”

“We need to have better digital strategy; to maintain people in place - for example - our software engineering teams here are based on research grants, not national capability.

So having National capability to develop digital infrastructure is key.”

Interview – British Antarctic Survey

INFRASTRUCTURE ALREADY BEING INVESTED IN

Existing investments in infrastructure mentioned during interviews

University of Plymouth

PML has invested in a special 5G network which covers the area (Weston Channel Observatory / Smart Sound Channel) - allows us to have good connectivity from the vessels and buoys.

Connectivity

University of Southampton

“University has invested in heavily in high performance computing, to support researchers. The University has invested a lot in it’s HPC facilities. Top 10 in UK for capacity…”

[preference of local servers over Cloud compute]

High Performance Computing

SAMS

”IT team is pushing our entire system to cloud… just before Christmas we had some power-cuts, all the servers went down - at midnight, the full security team and IT had to reboot… we’re thinking more and more… we need (Cloud)”

Cloud compute

NOC – MARS / Gliders

“Everything we do pretty much is digitised and stored online - Sharepoint. We’re using GIT for some elements of work - we’re kind of slowly migrating to moving more processes to GIT…”

Digital tools Cloud compute

BAS – BIOPOLE project digital tools

- SharePoint centralised system
- Centralised Risk register
- Centralised disk storage

“It’s a big enough project, that we have resource for this at BAS… even though there’s this structure in place - it doesn’t mean everyone engages with it…”

[BIOPOLE NERC] Digital tools Cloud compute

MetOffice

“MetOffice has shifted a lot of their internal employees into AI training / upskill (training courses, etc) due to competition in market… drive to not get left behind….some parts of process (of climate modelling) are doing this better with AI…”

Training

UKCEH

“We are starting to work with internal object storage at UKCEH…

- AWS - Object store only
- Not using Llambda side (decision)
- Use python scripts
- Budgeted up to £80,000 a year

“Just what we’ve got running on there at the moment… we haven’t scaled it up to test what the costs could be. We looked at doing a proper procurement through the Govt. portal for proper credits - but a lot of them expect you to know exactly what you want…”

Storage

NOC – PAP-SO/Moorings & Data flows

- Automated data flow processes from Moorings now being implemented as part of PAP-SO project workstream.
- Currently only one Mooring but will be replicable for others (in future).

Connectivity Training

CURRENT PROJECTS UNDERWAY

Some notable projects already “making waves”

BAS AI Route Planning Carbon Reduction Tool

“Currently UKRI doesn’t see a strategic need for AI decision support and decarbonisation aspects, using digital… so we’re pushing on alternative funding sources… we’re doing all right on.

FMRI funding: “We have had a few funded through FMRI, for the Decisions Logistics Tool for planning of fleets of Autonomous Vehicles…” [01:49]

The MVP tool was built to optimize the Sir David Attenborough – the team set out to build something which could effectively carry out “route planning for the SDA that reduces carbon cost.”

However, it has great potential in supporting autonomy operations too:
“If we move to the FMRI future of 100+ autonomous vehicles, and it takes a month to plan 2 vessels. And there’s 12 here… and here there’s a 100. It’s infeasible. So systems like this are needed – to reduce the complexity of the problem enough - so planning expert can come in and say “oh yeah that makes sense, that works” - it reduces the complexity of the problem down to ‘completely unfathomable’ to something that you can dive into and correct.”

IMMEDIACY OF PLANNING:
“It takes months to create the initial plan for marine planners. It takes multiple days to then reorganise to just to find carbon savings. And **it takes 30 minutes for this AI planner.** We got it down to 35 seconds for one year ahead…”

“**Saving potential of £2.1 million in costs**…”
“This is first stages, we haven’t got all the constraints in the problem like equipment…”

The Crown Estate / MEDIN

”We’ve started doing some stuff, particularly around metadata creation. Nothing groundbreaking, Basically using a large language model to try and help key metadata fields on MEDIN metadata. That was step 1 for us, when as soon as we have a functioning form of a chat model, based on a large language model. Applicability of that is the next step…right let’s automatically fill in the description fields, date ranges - and run over it with the human eye… that’s sort of the steps we’ve gone down at this moment.”

[I explained it’s about setting up teams to do this cross-org]

“In terms of that then, what I’d say is we have 50% of a (contractor) AI guy. And previously, when we’ve been running at capacity, which we’re not at the moment - we’ve had 2 devs, a QA, and a 50% of a design guy doing the platform. What we refer to as a Squad… we chuck a problem at them and they are able to solve it” [08:38]

Egeria (Linux Foundation Project)

Open metadata and governance for enterprises - automatically capturing, managing and exchanging metadata between tools and platforms, no matter the vendor.

MAS-DT

Initially began as the SOAR (Squad of Autonomous Robots) project – a project with Royal Holloway University & NOC, to allow various autonomous sensors to work together within a demarcated area, using autonomy – to carry out (for example) backup procedures if one of the robots failed, others would automatically cover the area – without human intervention from Pilots required.

The defined backbone of the messaging comms platform then helped shape MAS-DT. MAS-DT was more about the reliability of the autonomy and creating a UI to allow a user to define objectives for the autonomy to execute – including:

- Added Constraints (eg: stay in this work area)
- Exclusion zones (eg: oil rig, lighthouse, offshore wind farm, hazard zones)

Technically the project is finishing, but the Glider is carrying on (MOWGLI project) and it is “running itself at the moment” – so NOC are speaking to MetOffice to continue the data collection.

“**Currently the project focuses on “Route planning” objectives, but we’d like to make it more about Science objectives. There is no automated way of us supplying them training data.** ”

The concept would be to create a ”marketplace” within the existing platform whereby we could:

- Advertise what capabilities are available
- What models are compatible / accepted
- What platforms they apply to (eg: Oxford robotics only works for Slocums)

There is currently no automated way of supplying training data.

ATLANTIS

AtlantiS is a new UK ocean observing and marine science research programme that includes the continuation and evolution of critical observational time-series in the Atlantic Ocean.

Website

“These technological developments are really exciting. Are they going to be available to other users once ready via the marine equipment pool?”

[AtlantiS Working Session Attendee]

Partners:
- NOC
- SAMS
- Plymouth Marine Laboratory
- Marine Biological Association

Workpackages on the project cross a variety of subjects* and are aimed to deliver the continuation and evolution of critical observational time-series in the Atlantic Ocean – ultimately delivered through various end-points, including data visualisation dashboards and online tools.

*Work Packages include Ellet Array moorings and gliders, North Atlantic current time series, Go-Ship, Continuous Plankton recording data, MarineMet, Sea level observations from tide gagues, near present day simulations, tide gauge data and other oceanography data (BODC, PSMSL and GLOSS data centres), as well as a focus on innovations in observational infrastructure and data handling, early warning systems and sensing systems for evaluating fluxes and processes (Air Sea Flux, Benthic Flux and Subsea cable technology).

CURRENT PROJECTS UNDERWAY

Some notable projects already “making waves”

“Drone Metadata Project” – UK CEH

Amplified project looking at sensor data standards, collaborative, data re-use. “Building a platform to enable data sharing, a data model for our sensor data, we’re making sure that can be extended to other sites… we’re hoping that will deliver an extended metadata catalogue. At the moment, for any monitoring network dataset, if you’re lucky, they’ll list the names of bounding sites… explicit sets of terms…now it’s all done by dumping words in an abstract…”

“Were taking the stuff that we need - making sure that it COULD meet wider requirements.”

Using Research Data Alliance drone metadata standards: “It’s not really an agreed standard – but it’s an international standard.”

(Co-Lead Alice Fremand – BAS)

CURATE portal
“They’ve created a web-portal - **designed to be used internally… but there are a few Clients, similar to us, Nationally funded facilities - wanting to make the data more broadly visible – if not more accessible…**” [12:10]

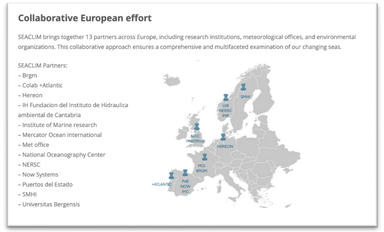
EO Data Hub
We provide a single point of access to EO data and processing infrastructure.

“Survey on the UK Directory of Marine observing systems. What sustained marine observations are going on around the UK: 150 responses, 22 people individually, plus ran workshops – past 3 months, currently analysing. “**Conclusion was - yes the community still does want some form of database of what sort of monitoring is going on…**” [36:58]

[MEDIN – Clare Postletwaite]

The **AI team at BGS** has recently started predicted seabed mapping. **A short project was also undertaken with the Alan Turing Institute on detecting shallow gas in scans of legacy seismic data records.**
[Survey, Q5]

“**MERMAN exists - but there’s no equivalent to MERMAN for (ocean) sensor data**”
[21:45]



SeaClim project
13 partners: Pilot project for Marine climate projections for Europe. Drive regional biogeochem projections (North Sea, N orth East Atlantic, Arctic)

Work Packages led by:
• Management and coordination: MercatorOcean
• Regional ocean decadal predictions: NERSC & UiB
• 21st century regional ocean projections: SMHI & MercatorOcean
• Ocean climate indicators: IMR & NOW
• Coastal & ocean climate service demonstrations: HEREON
• Communication, dissemination, stakeholder engagement, uptake & exploitation: CoLab +ATLANTIC
• Later WP’s are closely tied to end-uses of research (Paula Salge)

Floods and Droughts Research Infrastructure Programme – UK CEH

An innovative long-term programme which, for the first time in the UK, will monitor the whole hydrological system, to improve resilience to floods and droughts:

• Open sensor monitoring infrastructure - recommendations and guidance for [land-based] research groups / scientists on how to deploy monitoring infrastructure to allow streaming into FDRI / NERC data systems. Remit yet to be entirely defined but may include open-source package for management of Campbell data loggers, examples of code for managing image devices on Raspberry Pis and transmission of data, etc.

• Sensor data management platform - cloud-based platform for receiving sensor data via MQTT and other protocols, automated processing, QC, infilling, user interfaces, APIs, etc. Developed for FDRI sensor data but able to be opened to other research groups / programmes and will be extended to other UKCEH monitoring.

• Open-source tools for QC, in particular of rainfall and river flow data, but lots of generic methods for other sensor types

• Sensor metadata management systems (primarily for FDRI data, potentially more widely usable) - data model for sensor data descriptions, deployments and datasets (see https://fdri.org.uk/news/sharing-sensor-metadata-models) - already in discussion with NOC, BGS, others about this and making use of existing standards including NOC's iADOPT, and use of vocabularies

• Sensor metadata management systems making use of this data model, based on linked data stores and services, primarily for FDRI data but potentially usable more widely

• Mechanisms for automated publication and update of published sensor datasets into NERC data centres (EIDC in the first instance)

• Exemplar systems for storage and management of other data types including images, videos, UAV data (based on standards in development through RDA group and involving BAS and others)

• Stores and patterns for large gridded time series datasets based on Pangeo community approaches, likely on Jasmin, again, in discussion with NOC and others

• Notebook and analytical platforms - based on NERC funded DataLabs running on Jasmin

TRIANGULATING OUR FINDINGS

Quantitative survey

We used qualitative research to help expand and deepen our understanding of the needs and requirements of the current users of the digital infrastructure. *Methodological triangulation is the use of at least two methods, usually qualitative and quantitative, to guide the same research problem**.

Having formed our Insights from the qualitative findings, we then wanted to validate our findings across a wider user group – to check credibility, and to provide an opportunity for anyone else we had not spoken to in a 1-2-1 listening session to provide feedback.

We used the survey to achieve this…

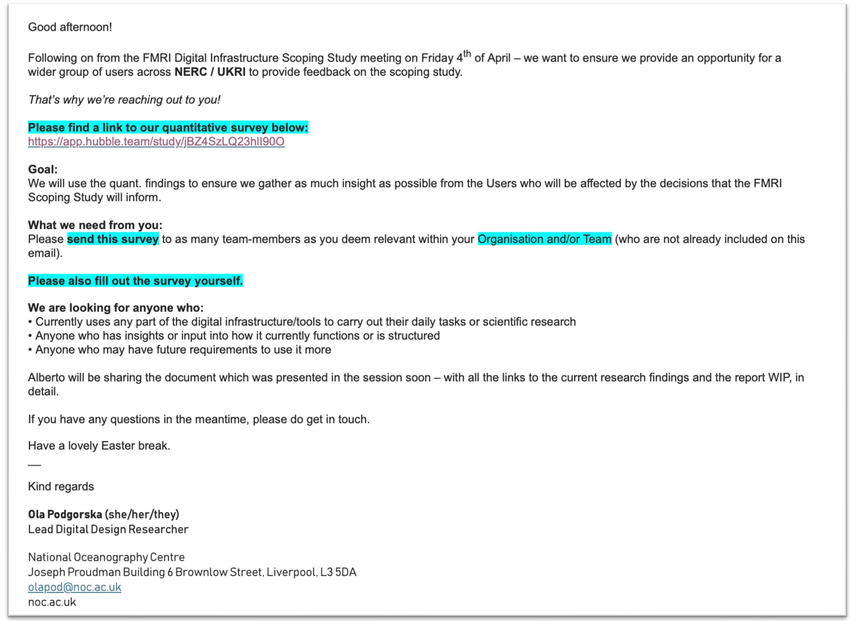
*Source

SURVEY

Survey sample – who was consulted?

Survey respondents

We sent the survey via email. The first email was sent in March 2025, the last reminders were sent towards the end of April, closer to the survey close date.



Good afternoon!

Following on from the FMRI Digital Infrastructure Scoping Study meeting on Friday 4th of April – we want to ensure we provide an opportunity for a wider group of users across **NERC / UKRI** to provide feedback on the scoping study.

That's why we're reaching out to you!

Please find a link to our quantitative survey below:
https://app.hubble.team/study/jBZ4SzLQ23hIl90O

Goal:
We will use the quant. findings to ensure we gather as much insight as possible from the Users who will be affected by the decisions that the FMRI Scoping Study will inform.

What we need from you:
Please **send this survey** to as many team-members as you deem relevant within your Organisation and/or Team (who are not already included on this email).

Please also fill out the survey yourself.

We are looking for anyone who:
• Currently uses any part of the digital infrastructure/tools to carry out their daily tasks or scientific research
• Anyone who has insights or input into how it currently functions or is structured
• Anyone who may have future requirements to use it more

Alberto will be sharing the document which was presented in the session soon – with all the links to the current research findings and the report WIP, in detail.

If you have any questions in the meantime, please do get in touch.

Have a lovely Easter break.
—
Kind regards

Ola Podgorska (she/her/they)
Lead Digital Design Researcher

National Oceanography Centre
Joseph Proudman Building 6 Brownlow Street, Liverpool, L3 5DA
olapod@noc.ac.uk
noc.ac.uk

Who we sent the survey to:

- NOC
- BAS
- MetOffice
- UKCEH
- CEFAS / DEFRA
- JNCC
- MEDIN
- NERC / UKRI (shared through the UK National Climate Science Partnership)
- SAMS
- Universities: Southampton, East Anglia, Plymouth
- Natural England
- The Crown Estate

Who responded:*

This is based on responses from Q1: "What is your role / which organisation do you work at".

41 responses, 4 left blank

- NOC
- UKCEH
- BAS
- CEFAS / DEFRA
- ABPmer
- Historic Environment Scotland
- University of Southampton
- The Crown Estate
- MEDIN
- SAMS
- MetOffice
- British Geological Survey (UKRI)
- The Marine Biological Association
- East Anglia University

SURVEY

Survey method

Preparation

Through the process of qualitative research, a large group of areas of opportunity were identified. Through the affinity mapping exercise, these were initially broadly grouped into 7 themes*. These helped us form **Question 3** & **4**.

Question 1 & **2** were optional – name and Organisation you work for. Question 5 was an open question (see Page 41).

The survey was set up in the Hubble app, and sent out as a link in an email.

* During further analysis, 8 key insight areas were finally settled on, but the survey was sent out before the full analysis was complete to ensure we gave participants enough time to provide feedback.

Question 3

We used max-diff analysis – rank choice voting (asking participants to rank the 7 themes from best to worst) for the survey. This constraint was intentional to help us understand what trade-offs or choices the participants would make.

Why we used this method

Through the question we obtain greater discrimination between all of the potential choices, and the degrees to which they are seen as important. This helped us triangulate the qual feedback and provide reassurance to the FMRI for potential resource allocation.

Question 4

We used the same max-diff method, but this time to understand how individuals rank what digital tools could help them improve and gain more value from their work.

There were 8 options to rank - and these were based on our qual research findings, and were aligned to the pan-organisational model (Customer/Strategy/Technology/ Operations/Organisation & Culture) - Delottie Digital Maturity Model (Page 11).

SURVEY

Question 1

You likely rely on digital tools to help you at work. This includes anything and everything from software, systems and networks infrastructure - applications that help you do your work more efficiently, and effectively.

Scenario 1: Imagine you have the opportunity of improving your day-to-day digital environment and infrastructure to help you do your work. Please rank the following options…

1 = needs most improvement
7 = needs least improvement

*__Please note__: The Hubble app ensures the questions are randomised to each participant – to reduce bias / preference based on existing ranking.

Options Participants were asked to rank*

A. **Tech environment systems, platforms**: Linux/Windows, Object store, Python etc.

B. **Consistency**: FAIR data, standardised sensor network data, embedded metadata best practice processes, digitisation of historic data

C. **Interoperability**: connecting systems internally, API's, reducing internal "knowledge silos", new cross-Org software etc.

D. **Storage**: intentional, managed use of Cloud and Servers (where and when data is stored depending on use case, increased capacity for large data volumes eg: imagery, audio, video etc.)

E. **More data from marine environments – incl. under observed areas** (increased autonomy, autonomous platform control, lab-on-chip, "green" ships etc.)

F. **Insights from data**: to demonstrate research & funding value, create reports, analysis, linking by context (eg: location, expedition, errors & anomaly detection).

G. **Digitise workflows**: from manual to automated task flows (e: reduce entering information into excel spreadsheets, moving files from hard drives manually etc.)

SURVEY RESULTS

Question 1

Consistency, Digitise workflows and **Interoperability** ranked in the top 3 positions, respectively.

This confirms our findings from the qualitative research – where the biggest challenges have been reported across disconnected network systems (between the autonomy & vessels infrastructure at sea, as well as the planning & data platforms (up and downstream).

Tech environments ranked last – which potentially suggests the audience of the survey did not correlate the importance of a robust technical infrastructure to the top 3 choices (but Tech environments need to be fully functional and connected to make the first 3 possible).

The participants ranked the options as follows:

Summary | By tester

	1	2	3	4	5	6	7	Weighted average
Consistency: FAIR d	10	12	11	7	2	2	1	2.76
Digitise workflows: f	11	9	7	7	5	6	0	3.09
Interoperability: cor	7	14	6	5	6	5	2	3.27
More data from ma	11	4	3	7	7	5	8	3.93
Insights from data:	3	2	7	8	12	8	5	4.51
Storage: intentional	3	1	9	6	6	9	11	4.82
Tech environments	0	3	2	5	7	10	18	5.62

Hubble Weighting

Hubble use a scoring method where lower “ranks” indicate higher preference. Weighting is applied based on how many times each item is ranked in each position (e.g. Item A: Ranked 1st x 6 times, Ranked 2nd x 3 times, Ranked 3rd x 1 time). Then, the total score is divided by the total responses (in our case, 45).

SURVEY

Question 2

Digital maturity refers to an Organisation's ability to effectively use digital technology to create value.

Based on your own experiences - what digital tools could FMRI develop / provide that will help you improve and gain more value from the work you carry out?

Please rank according to:
1 = needs most improvement
8 = needs least improvement

*__Please note__: The Hubble app ensures the questions are randomised to each participant – to reduce bias / preference based on existing ranking.

Options Participants were asked to rank*

A. **Delivery governance** (**resource** & **process management**)

B. **Standards** & **governance automation**

C. **Security**

D. **Applications** (**software, apps**)

E. **Interoperability** (sensors and autonomous platforms, IoT, integrated services etc)

F. **Network** (wifi, satellite etc.)

G. **Tech. architecture**

H. **Data** & **analytics** (real time, insights etc.)

SURVEY RESULTS

Question 2

In the second question – which focused more on digital maturity – **Interoperability, Tech Architecture and Standards** & **Governance** ranked in the top 3 positions, respectively, supporting the qualitative feedback around the need for a more connected Oceanography research network.

Tech Architecture moved up to second place from last place here – potentially because the question explicitly mentioned the word ‘digital’ (whereas the first question was more open-ended, and from a user point of view, could refer to, for example, the scientific infrastructure).

Security came in last. This is potentially because it’s not ultimately a user need, but a business requirement (security steps often introduce friction / additional steps in usability).

Summary | By tester

	1	2	3	4	5	6	7	8	Weighted average
Interoperability (s	13	10	9	5	3	1	4	0	2.87
Technological arc	7	6	6	9	4	7	2	4	4.02
Standards & gove	5	8	4	10	8	1	4	5	4.16
Applications (soft	9	6	6	5	2	4	8	5	4.2
Delivery governan	6	4	5	6	8	11	3	2	4.36
Data & analytics (	4	3	10	5	9	7	4	3	4.42
Network (wifi, sate	1	8	1	1	2	6	11	15	5.93
Security	0	0	4	4	9	8	9	11	6.04

Hubble Weighting

Hubble use a scoring method where lower “ranks” indicate higher preference. Weighting is applied based on how many times each item is ranked in each position (e.g. Item A: Ranked 1st x 6 times, Ranked 2nd x 3 times, Ranked 3rd x 1 time). Then, the total score is divided by the total responses (in our case, 45).

SURVEY RESULTS

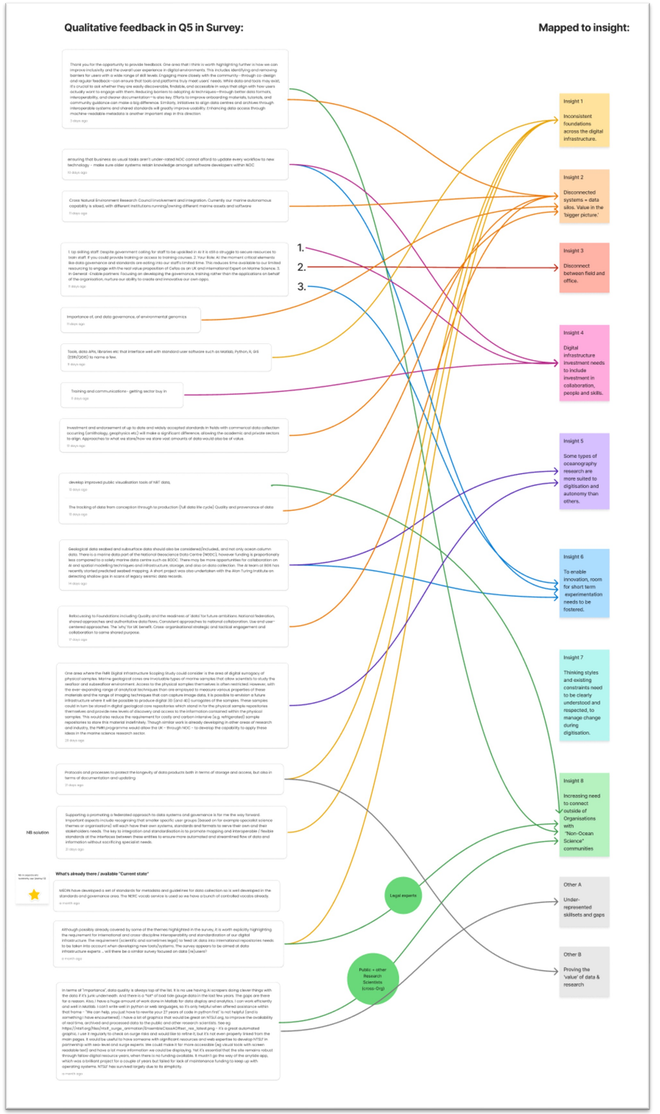
Question 3 – open feedback

Question

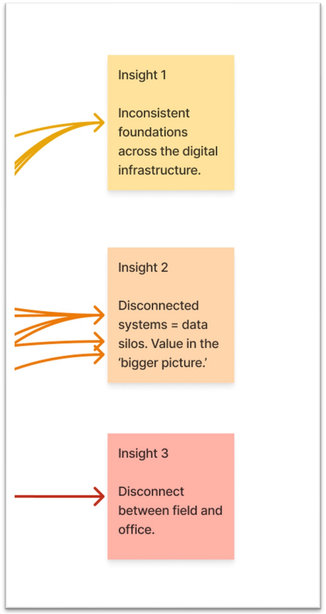
Is there anything we've not covered in the survey which you feel is important to note? (optional) **Please provide as much context as possible.**

- We received 24 responses.
- These were consolidated through the affinity mapping exercise and formed part of the qualitative Key (Pg 3 -7) and In-depth insights (Pg 19-38).

Feedback…



Mapped to Insights…



Section 3:

RECOMMENDATIONS

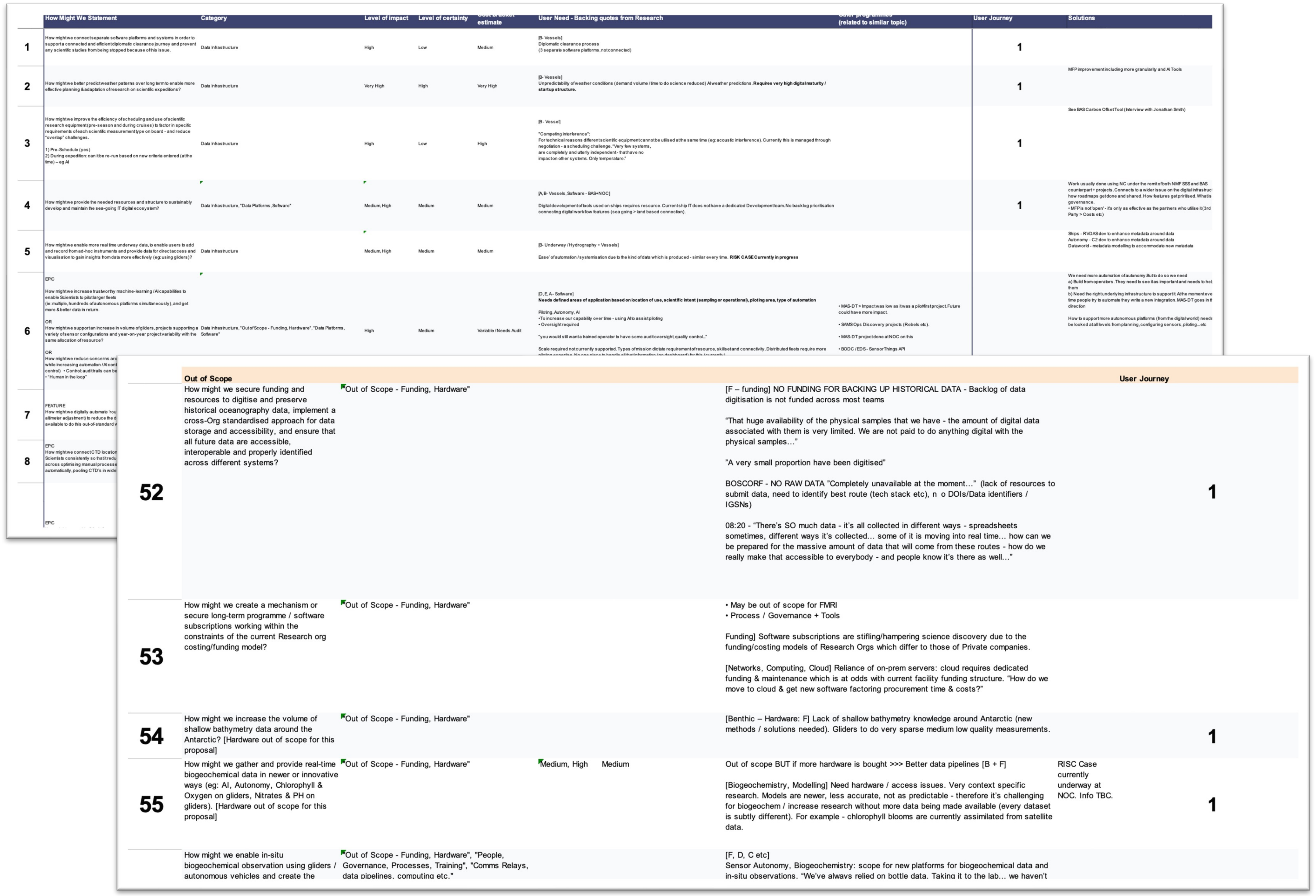
HOW MIGHT WE STATEMENTS

Please refer to the HMW spreadsheet for details

Based on the user research, we have created How Might We Statements to provide recommended areas and gaps which would benefit from solutions the FMRI should focus on. These HMW's would need to be ideated on and prototyped by digital teams who can then deliver on them (depending on solutions proposed – see diagram on Page 12 of this Report). These are numbered and are organised across two spreadsheets:

- There are 51 How Might We's (1-51)
- There are 17 Out of Scope HMW's (52-68)

The first sheet includes all the HMW's we consider to be in-scope. The second sheet is anything we considered out-of-scope, or that may be 'changed' to in-scope if the solution turns into a digital one.



	Out of Scope						User Journey
52	How might we secure funding and resources to digitise and preserve historical oceanography data, implement a cross-Org standardised approach for data storage and accessibility, and ensure that all future data are accessible, interoperable and properly identified across different systems?	"Out of Scope - Funding, Hardware"			[F – funding] NO FUNDING FOR BACKING UP HISTORICAL DATA - Backlog of data digitisation is not funded across most teams "That huge availability of the physical samples that we have - the amount of digital data associated with them is very limited. We are not paid to do anything digital with the physical samples…" "A very small proportion have been digitised" BOSCORF - NO RAW DATA "Completely unavailable at the moment…" (lack of resources to submit data, need to identify best route (tech stack etc), n o DOIs/Data identifiers / IGSNs) 08:20 - "There's SO much data - it's all collected in different ways - spreadsheets sometimes, different ways it's collected… some of it is moving into real time… how can we be prepared for the massive amount of data that will come from these routes - how do we really make that accessible to everybody - and people know it's there as well…"		1
53	How might we create a mechanism or secure long-term programme / software subscriptions working within the constraints of the current Research org costing/funding model?	"Out of Scope - Funding, Hardware"			• May be out of scope for FMRI • Process / Governance + Tools Funding] Software subscriptions are stifling/hampering science discovery due to the funding/costing models of Research Orgs which differ to those of Private companies. [Networks, Computing, Cloud] Reliance of on-prem servers: cloud requires dedicated funding & maintenance which is at odds with current facility funding structure. "How do we move to cloud & get new software factoring procurement time & costs?"		
54	How might we increase the volume of shallow bathymetry data around the Antarctic? [Hardware out of scope for this proposal]	"Out of Scope - Funding, Hardware"			[Benthic – Hardware: F] Lack of shallow bathymetry knowledge around Antarctic (new methods / solutions needed). Gliders to do very sparse medium low quality measurements.		1
55	How might we gather and provide real-time biogeochemical data in newer or innovative ways (eg: AI, Autonomy, Chlorophyll & Oxygen on gliders, Nitrates & PH on gliders). [Hardware out of scope for this proposal]	"Out of Scope - Funding, Hardware"	Medium, High	Medium	Out of scope BUT if more hardware is bought >>> Better data pipelines [B + F] [Biogeochemistry, Modelling] Need hardware / access issues. Very context specific research. Models are newer, less accurate, not as predictable - therefore it's challenging for biogeochem / increase research without more data being made available (every dataset is subtly different). For example - chlorophyll blooms are currently assimilated from satellite data.	RISC Case currently underway at NOC. Info TBC.	1
	How might we enable in-situ biogeochemical observation using gliders / autonomous vehicles and create the	"Out of Scope - Funding, Hardware", "People, Governance, Processes, Training", "Comms Relays, data pipelines, computing etc."			[F, D, C etc] Sensor Autonomy, Biogeochemistry: scope for new platforms for biogeochemical data and in-situ observations. "We've always relied on bottle data. Taking it to the lab… we haven't		

THANK YOU

If you have any questions, or would like to book in a presentation session of the research summarized in this document, please do get in touch:

- **FMRI Team** – info@fmri.ac.uk
- **Alvaro Lorenzo Lopez** - Senior Software Engineer, Marine Autonomous and Robotic Systems (MARS) - alvaro.lorenzo@noc.ac.uk
- **Ola Podgorska** – Lead Digital Design Researcher, NOC – olapod@noc.ac.uk