**Clarification Questions & Answers**

**Contract Reference: C21-0262-1579**

**Title: Marine Recorder Redevelopment**

**Q1)** Are there any requirements for the system to be bilingual (with the Welsh language)?

**A1)** Whilst this was initially deprioritised by the project steering group for the redevelopment, we are in the process of reconfirming this with the relevant members of the project steering group.

As an answer to this may arrive too late to be considered in bids, we would ask that if you believe that developing the ability for bilingual interface to the system would add a significant cost, you provide a cost for both a mono-lingual and bilingual system, or otherwise indicate specific costs associated with this.

**Q2)** Are the terms and conditions fixed or is there possibility to negotiate on these?

**A2)** Yes, there is some room for negotiation. Any proposed clarifications to our terms should be submitted with your bid and will be considered by JNCC and the external reviewers and taken forward to negotiation should the bid be a leading candidate.

**Q3)** How close is the current MR infrastructure to that laid out in the specification

**A3)** The current infrastructure is outdated, no longer fit for purpose and built on an unsuitable technology platform. The new MR would be a completely new development, new software, new infrastructure and a new (revised) data model.

**Q4)** What are the requirements for data analysis?

**A4)** The data management and analysis elements are being considered as separate elements. Typical analysis elements might include those designed to improve the quality of the data including visualisation of points on a map, or automated logical checks for data matching the expected type. Future potential analysis might include the development of 'rules' to check data quality based on existing species distribution or additional information

**Q5)** What is the deadline for submission

**A5)** All detailed in the contract documentation (August 10th 2021)

**Q6)** Would the ability to show who has accessed what data be useful

**A6)** Yes. An "audit trail" for data would be a useful feature

**Q7)** Do you consider the sitemap / wireframes to be complete, i.e., cover all required functionality?

**A7)** The sitemap and wireframes were generated from a suite of user stories that were collated from the stakeholder groups, and these form a very good set. However, whilst every effort was made to gather a complete set of user stories, there may be assumed user-stories/functionalities which may have been missed as part of the initial gathering session.

The same stakeholders’ organisations will form the project working groups through the development of the system and individuals are will be available for discussion of user stories (and subsequent UAT), therefore we would encourage the successful bidder to engage with the project groups to further clarify and question the user stories, drive discussion and through UAT to test that developed areas of the application are fit for purpose.

The background material and current work on this system has been build from user stories up. However, a fully detailed audit has not been carried out.

**Q8)** We assume that you anticipate a bespoke development, rather than tailoring something off-the-shelf?

**A8)** Not necessarily. We do not want to be tied in to one approach. If an off-the-shelf system can take a proportion of the requirements, that is a viable approach.

**Q9)** Are there any preferred technologies (e.g., languages, framework, libraries) for the implementation - perhaps because the internal team at JNCC already have familiarity with certain technologies?

**A9)** There are no preferred technologies. Whilst JNCC is leading on the build and development, it is a partnership project with other organisations. JNCC will not be taking on the ongoing development, and therefore JNCC’s familiarity with underlying infrastructures or infrastructures is not of concern.

**Q10)** You mention that you have a "preference to move to open-source software due to the risk of the system becoming unsupported". This is in the context of the current system being on MS Access. Do you mean that the building blocks should be based on open-source tools/frameworks/libraries, or do you also mean that the implementation of the new system should be open sourced too?

**A10)** Being a government organisation, there is a lean towards the use of open-source code due to public funding, but the key requirement is to find the best technology. The main driver behind the open-source statement is to mitigate risk of underlying libraries/frameworks becoming unavailable for use for a variety of reasons mid-lifespan of the final system, leaving the system unusable. Similarly, whilst novel code developed as part of the project does not necessarily have to be open source (or partnership owned), the main concern is the novel code becoming again unavailable, for example if the IPR holding developer ceases to exist.

**Q11)** How much usage is anticipated for the public API? Number of accesses per day/week and are these accesses likely to return large quantities of data? How much usage will organisations generate?

**A11)** In the initial phase, whether there is a direct public API is a question of phasing that will be discussed with the successful bidder. That aside, this is a current uncertainty as the current system’s data is provided as a static download, with general public often downloading and reusing as many times as they wish. However, API induced costs (e.g., egress) would be incorporated in the system’s ongoing maintenance/hosting costs and should costs surpass available funds for hosting between the project partners, accessibility to the API and funding would be considered appropriately. As alternative could be to move the public accessibility to an environment that allows for free access to public data (e.g., AWS), thus moved to an external system.

**Q12)** What is the team that will be working on this from JNCC (and other’s) side? What are their capabilities, roles and what is the engagement that JNCC are expecting from this?

**A12)** In terms of hours provided, the two joint JNCC project leads have a large amount of hours prioritised towards this (the product manager should be able to be fully focused on these developments if required), so the successful contractor should expect full and speedy engagement from JNCC to directly answer questions and drive discussion as required. Time for stakeholders to engage in user-stories discussions and UAT is negotiable but the external stakeholders have committed to supporting this development and hopefully should be able to provide around 1 hour every two weeks at minimum, and as development ramps up this time should increase as the stakeholders are fully invested in the system.

Alongside the project will be two main groups. The technical working group will be formed mainly from users of the current system, generally data managers with a mixed data science “coding” skillset (not developers) but good eye for detail, and will be tasked with discussing user stories, technical questions, providing UAT and similar tasks. The governance group will be comprised of more senior staff and/or IT specialists and will handle raised difficult queries from the technical working group and also take strategic or political decisions, for example regarding reprioritisation of areas due to cost implications.

**Q13)** Is there an appetite for using this development to address wider concerns within the community? For example, if there are a number of systems used to manage similar data types, are there aspirations that this development could consolidate work done elsewhere, or is this more about replacing the current Marine Recorder system?

**A13)** The primary focus of the redevelopment will be to replace the current system, as that is the driver behind the funding of the project and the main driver of the current stakeholders (they need to manage their own data and the current system is past end-of-life). However, there is the appetite to grow the user base (in terms of custodian organisations), or the system could interact with other systems through linked data. We would like to work with existing systems, not necessarily replacing them, but helping to streamline the marine data environment.

**Q14)** What is the expectation of accessibility?

**A14)** As the initial system is an internal management system rather than public engagement, the concern is lowered. However, as staff within the organisations may have access issues to technology, we do have to consider accessibility as part of the system (as noted in the Annex A).

**Q15)** Different survey methodologies ideally have quite different interfaces for efficient data entry. Is this something that is covered somewhere – or are you wanting tenderers to look at this?

**A15)** Differing views of forms depending on sample/survey methods (e.g., simplified entry forms) is not something that has been prioritised as part of phase 1 / MVP of the system (the data entry indicative wireframes as seen in the additional documentation contain all fields within the related data model entities). However, it may prove of interest to the stakeholders, and whilst every effort to gather a full suite of user-stories has been undertaken, gaps and assumptions may remain. If such functionality is possible, it should be highlighted in the bid response as extra potential functionality along with associated cost estimates to allow fair comparison with responses that may not contain this added potential value. Even if not included as part of the initial MVP build, such functionality, after consultation with the other partner organisations that form the steering group, may then be added as potential feature development under future funding.

**Q16)** To help model costs and also in case it has a bearing on the solution architecture, how many custodian organisations are there at the moment? How many might there be in the future? And how many users within a custodian organisation, and how many within JNCC?

**A16)** There are currently 10 custodian organisations – an additional 10-20 custodian organisations by the end of the solutions life could be a reasonable maximum expectation. There is of course uncertainty around this, but hopefully the answer provides an indication of order of magnitude. Again, users with an organisation are highly variable. As a general indication, however, I would expect 1-2 “Data managers”, 5-10 with data entry roles, and 20-100 with view only access rights. Again, there may be considerable variability here depending on the size of the organisation itself. A single organisation would likely take the administration role with likely 5 staff maximum with access.

**Q17)** Section 5 in the Annex document, seems to suggest that for the initial £325k budget, it might only be feasible to deliver data management functionality, and not analysis and reporting. What would that mean in practice? Would the system not be able to go live without the analysis and reporting functionality?

**A16)** In this instance the management functionality of the system would “go live” enabling data managers to properly curate and govern existing data and enter new data as necessary. The partners and project groups would then work with the developers to determine a stop-gap method of data access to allow analysis and reporting functionality to be externalised from the application in the client’s machines (for example, but not only, a spatialite database download of current published data). Additional subsequent funding would then go towards the development of an online in-system analysis and reporting environment.

**Q18)** In the Annex document, the conceptual data model has clearly had much work put into it - does this mean that there are certain assumptions and constraints built into it that may not be obvious? I.e. should we aim to keep the structure as given because it will cause unforeseen problems to others if it is changed?

**A18)** JNCC (and the external partners) would welcome discussion with the successful bidder to work through possible alternative data models. The data model presented as part of the associated documentation has been developed with constraints on fields where suitable (including but not limited to range constraints, type constraints, and composite constraints where for example multiple fields must sum to 100). And some user stories contain these constrains in detail. One downside of the internal data model used by the **current** system is due to its normalised nature regarding e.g., measurements, all values are stored as text in a single field the “measurements” table, which leads to more difficult constraining of values and as a result certainly the current system has experience issues of unclean or inappropriate data values stored against various measurements. However, if a modern system is able to provide an alternative data model which is robust and resilient and can provide access to highly standardised data, this is a viable option.

**Q19)** We are asked to include “copies of appropriate risk assessments” with our proposal - was there anything specific in mind behind this?

**A19)** Standard company risk assessments regarding work undertaken by their employees.

**Q20**) For clarification questions, and for those asked by other bidders, how are the responses distributed?

**A20**) The questions will be published via the relevant JNCC website procurement and contracts finder (<https://www.gov.uk/contracts-finder>) webpages.

**Q21)** The Annex A says: *“In our new site map and other documentation, that we have supplied in this invitation to tender, we have retained this concept of a snapshot of the data for publication, but this is only for simplicity. We are aware that with a central web hosted database solution, the requirement for a user to be able to download a data snapshot could be fulfilled by direct access to the live database and that user only having access to the survey in question. To enable this a new system would require survey data to be updated through version control, so a currently saved and published record could be accessed and downloaded for a “snapshot” while another more up to date version of the record was being updated and edited. We are happy for either approach to be considered as a solution, depending on the features in each provider’s solution.”*

If I understand this right this means that a given survey must have the ability to be flagged as available for dissemination whilst additional data is being added / corrected that is not available for dissemination.

Clearly this considerably complicates any potential model probably requiring it to be duplicated (or historical versioning maintained – though not really sure that would work for adding new samples etc unless they were individually flagged as OK to disseminate.)

So, my question is:

1) Is this functionality really required – given normally data is input for a survey and then only correctional edits are made (for which there is an argument the sooner they are available the better)

2) If it is – what level of granularity is required – is it every field – or could it for example be OK for a correction to a sample to be immediately visible – but not data on a new sample?

**A21)** The functionality required here is preventing edits to data from “going live” immediately and allowing a QC workflow to take place where the data manager can check said edits are correct and confirm the revised data for publication. This is to prevent the case of erroneous edits (e.g., from junior staff) creeping through and being published, with implications for policy etc. In the case of a completely new survey, the entire dataset can be “held back” by the manager before release. In the case of a survey which is already released/published but to which subsequent edits must be made (e.g., a species record has to be redetermined), the revised data should not be immediately released/published until the senior staff member has signed off the changes. **This behaviour should be for both new and corrected data (Q2)**.

Whether this approval process is undertaken at the entire survey level (as indicated in the indicative workflows), or at a more granular level (i.e., data manger signs off each change where it occurs) is a discussion that the successful bidder can have with the project groups. The indicative model of a “management” data area and a “published” data area, which could be two distinct databases, should allow for this by preventing movement of data across the divide until the data manager has signed off revisions (the “published” version is retained if the “management” version of a survey is marked as “awaiting signoff”), but this relies on nightly (or other frequency) migration, rather than instant publication/release. This delayed release was deemed acceptable by the user group.

However, other methods may exist without requiring this split system (such as a single data storage solution) and may be equally applicable, acceptable and potentially more beneficial, though as mentioned in the question will likely then require version control. As such we welcome alternative viable suggestions.

If satisfying this edit and republish behaviour is deemed by the bidder to be a considerable source of cost, we would welcome this being quantified within the bid.