

AQUASERV



Research services
for **sustainable aquaculture**,
fisheries and **blue economy**

D3.2 – AQUASERV Data Landscaping

Project title	Research Infrastructure Services for Sustainable Aquaculture, Fisheries and the Blue Economy
Grant Agreement No	101131121
Programme	HORIZON Research and Innovation Actions
Call	HORIZON-INFRA-2023-SERV-01
DG/Agency	European Research Executive Agency
Deliverable number	D3.2
Deliverable title	AQUASERV data landscaping
Lead beneficiary	Flanders Marine Institute (VLIZ)
Contributing beneficiaries	Centro de Ciências do Mar do Algarve (CCMAR)
Author(s)	Katrina Exter, Joanna Goley, Cymon J. Cox, Andrzej Tkacz
Type of deliverable	Report
Work package No	WP3
Work package title	Open science policy, data management, and service customisation
Dissemination level	PU
Due date (in months)	M8
Delivery date (actual)	February 3rd 2025

Description of deliverable	Results of the landscaping and gap analysis of AQUASERV services
----------------------------	--

Legal Disclaimer



Funded by the European Union.

Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them.

Copyright

© AQUASERV. Copies of this publication and extracts thereof may only be made with reference to the publisher.

Document History

Version	Date (DD/MM/YYYY)	Created/Amended by	Changes
0.0	06/01/2025	WP3	Initial draft
1.0	10/01/2025	Katrina Exter and Joanna Goley (VLIZ)	First draft
1.1	20/1/2025	Cymon J. Cox, Andrzej Tkacz (CCMAR)	Second draft
2	29/1/2024	QC reviewers	Final

Table of contents

• List of Abbreviations	5
• Glossary	6
• Executive Summary	7
1. The AQUASERV TA services	8
1.1. The Transnational Access services on offer	8
1.2. Collecting information about the data produced from Transnational Access services	8
2. The Questionnaires	9
3. Results	11
○ 3.1 The quality of the responses	12
○ 3.2 Summary of data types	15
○ 3.3 Summary of data management practices	16
■ 3.3.1 Practices	16
■ 3.3.2 Vocabularies, file naming, protocols, and instrument IDs	18
○ 3.4 Training	19
4. Summary and follow-up	20
5. Opportunities to improve data management practices	21
• APPENDIX	23

- **List of Abbreviations**

Abbreviation	Definition
AQUASERV	Research Infrastructure Services for Sustainable Aquaculture, Fisheries, Ecological Restoration and the Blue Economy
ARIA	Access to Research Infrastructure Administration
DMP	Data Management Plan
FAIR	Findability, accessibility, interoperability, and reusability
SOPs	Standard operating procedures
TA	Transnational Access
WP	Work package

• Glossary

Name	Definition
Access Officer	The Access Officer (AO) is responsible for overseeing and coordinating the Transnational Access programme in AQUASERV
Access Manager	The contact person at the Access Provider is responsible for communicating with the Access Provider and the applicant/user. There may be one or more Access Managers (AMs) per Access Provider. See also “Liaison Officer”
Access Provider	The organisation where the TA project is performed. Sometimes referred to as “Host”
Access Provider’s facility	The premises and the platforms of the Access Provider where the TA project is performed. Sometimes simply referred to as “facility”, “service-providing facility”, “access-providing facility” or “host facility”
Beneficiary	A beneficiary of Aquaserv project funding
Liaison officer	This term might sometimes be used synonymously with Access Manager
Station	The individual organisations/labs/marine stations that provide TA service. One institute can have several stations, one beneficiary can have several institutes and/or beneficiaries
Service provider	The entity responsible for delivering services, maintaining infrastructure, or offering consulting and technical support
Transnational Access (TA)	Provision of access (by an Access Provider) to a User or User group whose home institution is located in a country other than the country where the Access Provider is located

- **Executive Summary**

This document provides a landscaping and gap analysis of the data and data management aspects of the Transnational Access services provided in AQUASERV. This work is based on two surveys sent to all service providers to assess instruments, activities, data types, metadata, workflows, and adherence to FAIR data principles.

The findings reveal variability in data management practices and highlight areas for improvement. A key observation is the lack of formal workflows, publicly-available standard operating procedures, and data management plans across many institutions. While data is often acquired locally, only 30% of respondents regularly publish data. Metadata management practices were inconsistent, with limited use of standardised vocabularies, impacting data interoperability and reuse. The survey also revealed that only half of the respondents use unique identifiers for instruments. There is a clear need for training in FAIR data principles, and many respondents indicated an interest in training on FAIR data and database management. As so many institutions expressed a need for capacity building, a self-paced online FAIR data course is planned to meet these needs.

In conclusion, the institutions involved in AQUASERV exhibit varying levels of data management maturity. The responses received represent fewer than half of the total service providers, and a very small percentage of the number of actual services provided in AQUASERV; in addition many of the answers were too casual to provide sufficient information". Nonetheless, we could extrapolate the results across the network to build a first version of a data landscaping and gap analysis. Follow-up actions will take place in 2025 to pursue the additional responses and additional detail identified as necessary for a fuller analysis. While many are actively working on improving their practices, targeted support is necessary to help formulate standardised workflows, improve metadata management, ensure data sharing practices are aligned with FAIR principles, and to reach the goals outlined in the Data Management Plan ¹. This will enhance data interoperability and foster open science across the network.

¹ <https://zenodo.org/records/13839416>

1. The AQUASERV TA services

1.1. The Transnational Access services on offer

AQUASERV consortium comprises a network of six European Research infrastructures (RIs). The consortium has 33 beneficiaries providing Transnational Access (TA), via their 63 individual stations, to research and innovation services for sustainable aquaculture, fisheries, ecological restoration, and the blue economy. AQUASERV's TA proposals are created and submitted via the ARIA² (Access to Research Infrastructure Administration) system. All the services that are offered by the stations in AQUASERV are described and can be selected from within the ARIA platform. Work conducted by work package (WP) 7 harmonised the categorisation and description of services that are offered by the individual stations. An introduction to the services, grouped by country, infrastructure, and expertise, and categorised according to the WP7 schema, can also be found on the AQUASERV³ website. See Table A1 for a listing of the beneficiaries, service providers, and service categories.

More information about the construction of the service catalogue can be found in AQUASERV D7.1 - Services catalogue⁴. For the purposes of this deliverable D3.2, it is useful to summarise the range and scope of services that are on offer, as this provided the input WP3 needed for the data landscaping and data management (DM) among all the participants in the TA services of AQUASERV. We have provided a detailed summary of these services, organised by category and subcategory, in the Appendix (Figures A1a and A1b).

1.2. Collecting information about the data produced from Transnational Access services

The collection of TA service information by WP7 for ARIA naturally focussed on what was necessary for the TA proposal process; this did not, unfortunately, include any information about the types of data, the FAIRness, and the data management processed at each TA provider station. Therefore, WP3 devised new forms which the TA service providers were asked to fill in:

- As input to an early WP7 meeting with the TA service providers (focused on constructing the ARIA service categories), we requested basic information about the services they offered on a google sheet⁵: the generic names of the types of instruments/devices/platforms they had (e.g. "Molecular Biology Platform"), details of the particular instruments/devices for each generic type

² <https://instruct-eric.org/help/about-aria>

³ <https://www.aquaserv-ri.eu/transnational-access>

⁴ <https://zenodo.org/records/13627443> (Survey of aquaserv TA providers.xlsx)

⁵ <https://zenodo.org/records/13627443> (Survey of aquaserv TA providers.xlsx)

(e.g. “Ultracentrifuge - Beckman Optima Max XP”), and Y/N whether those produced data as output (e.g. imaging devices, sequencing platforms) or the user had to collect and save the data themselves (e.g. Conductivity, temperature and depth device (CTD), sampling devices). The aim was to get an initial overview of the scope of the instrumentation and amount of data that the TA services could provide.

- Two questionnaires focussed on collecting more specific information about the standards, formats, vocabularies, databases, and DM activities that each individual service provider had at their station. One questionnaire⁶ was aimed at those instruments/platforms which produced data as output, which would then be accessed and analysed by the TA users, and the other questionnaire⁷ was aimed at the services that were activities in which data were collected by the TA users themselves. It was intended that each beneficiary in AQUASERV would send this to their service provider stations to fill in. These questionnaires form the main backbone of this deliverable.

The combination of these two approaches gave us the overview – the landscaping – of the data and DM practices in AQUASERV. Questions regarding their use (or not) of standards, formats, etc, allowed us to identify gaps. Each service provider has the opportunity to then plug those gaps via the AQUASERV internal, WP3-run, funding call that should start in the first year and continue until the funding runs out.

2. The Questionnaires

At the start of the project, WP3 did not have access to a finalised list of service providers or their associated services. Consequently, we conducted a pilot survey focusing exclusively on The Algarve Centre of Marine Sciences (CCMAR) beneficiary, whose available services were already known to us.

The pilot survey was created using Google Forms and distributed on June 11, 2024, with responses collected by June 23, 2024. Its design was informed by prior experience with the AQUAEXCEL project⁸ (AQUAculture infrastructures for EXCELlence in European Fish research). The survey link was shared with the local liaison officer at CCMAR, who then forwarded it to the service provider managers for completion. We received detailed and highly informative answers, in nearly all cases. Facility managers attempted to complete each question thoroughly, and all relevant facilities participated in the survey, and even organised meetings to decide a unified response in some cases. The responses revealed two distinct types of services based on how they generate data:

⁶ <https://zenodo.org/records/14781098> (AQUASERV D3.2 survey - instruments_v2 - Google Forms.pdf)

⁷ <https://zenodo.org/records/14781098> (Questionnaires and analysis data for D3.2)

⁸ <https://aquaexcel.eu/>

Data-Heavy Services – These facilities produce large volumes of data in complex and varied formats, such as microscopy facilities generating image files or sequencing facilities producing sequencing data outputs.

Data-Light Services – These facilities produce simpler datasets, often in tabular form, such as data from boat services or access to coastal areas.

Based on this distinction, we categorised these services as **instruments** (data-heavy) and **activities** (data-light). Consequently, we divided the updated survey into two distinct parts: “Instruments” and “Activities.” This better reflects the varied data formats and DM needs of these different types of services. For instance, access to boats for water sampling at sea is expected to produce minimal digital outputs that can be effectively stored in a spreadsheet. In contrast, a sequencing facility generates vast amounts of data requiring specialised storage, analysis, adherence to protocols, and metadata containing ontology terms.

To enhance clarity, we revised the two successive surveys and included explanatory sections at the beginning of each. These sections employ visual aids to guide respondents on how to accurately complete the surveys (Figure 1). These can be found on our Google drive: questionnaire for instruments, questionnaire for activities⁹.

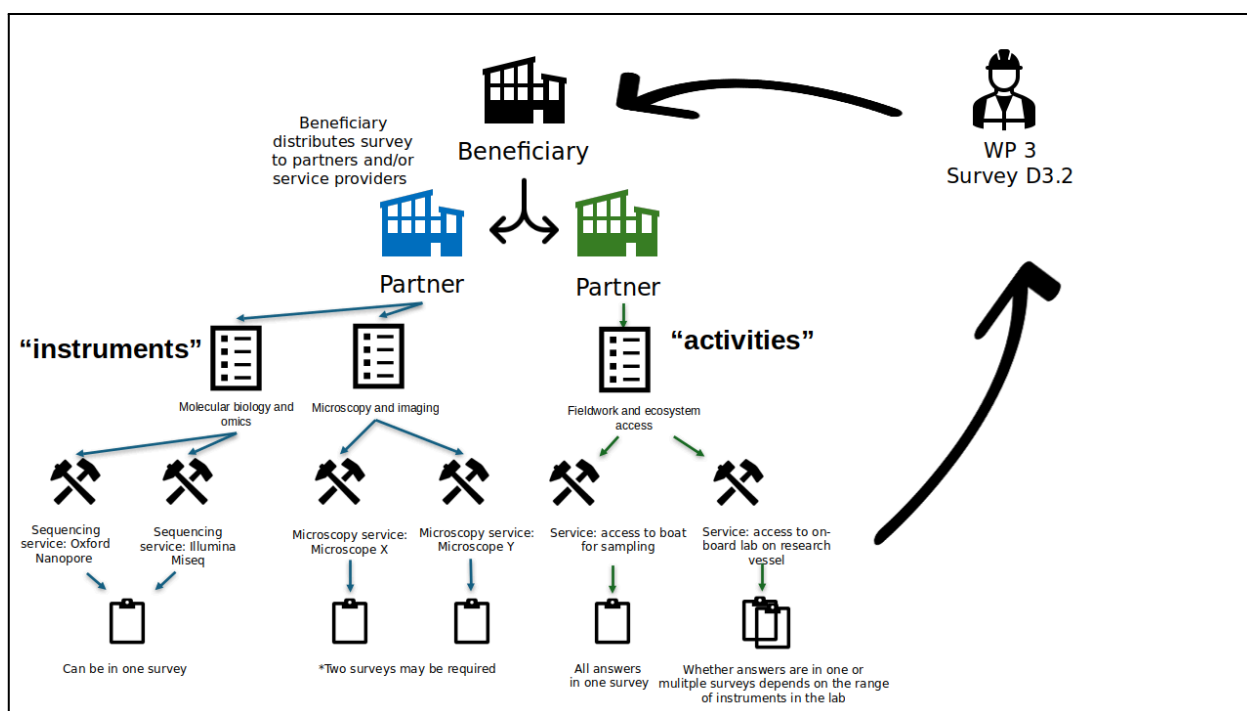


Figure 1: Visual guide provided in the questionnaires

⁹ <https://zenodo.org/records/14781098> (AQUASERV D3.2 survey - activities_v2 - Google Forms.pdf)

The categorisation of the services from all beneficiaries and their service providers was not completed by WP7 until Oct. 2025. Ideally, we would have sent out the questionnaire at that time, with one questionnaire per service subcategory, because then the respondents would have understood better to what instrument or activity each questionnaire was referring, and we could have personalised the emails based on the descriptions given for their individual services. However, the unfortunate, but necessary, timeline meant that the questionnaires had already been sent out. Indeed, during our analysis, it was clear that we did not have a uniform approach to what was considered an instrument, an activity, or even a service. The resulting reduction in the utility of the results will be compensated for in 2025 when we will hold one-on-one discussions with each service provider to fill in the gaps and to pursue follow-up questions. This deliverable will subsequently be updated when this action is completed.

3. Results

We summarised the questionnaires by grouping the questions into four categories: data types, DM practices, training requirements, and general (Table A2 and A3). By categorising the questionnaire, areas that may require more attention could be identified. Figure 2 describes the statistics of the responses to the questions related to DM, training, and standards.

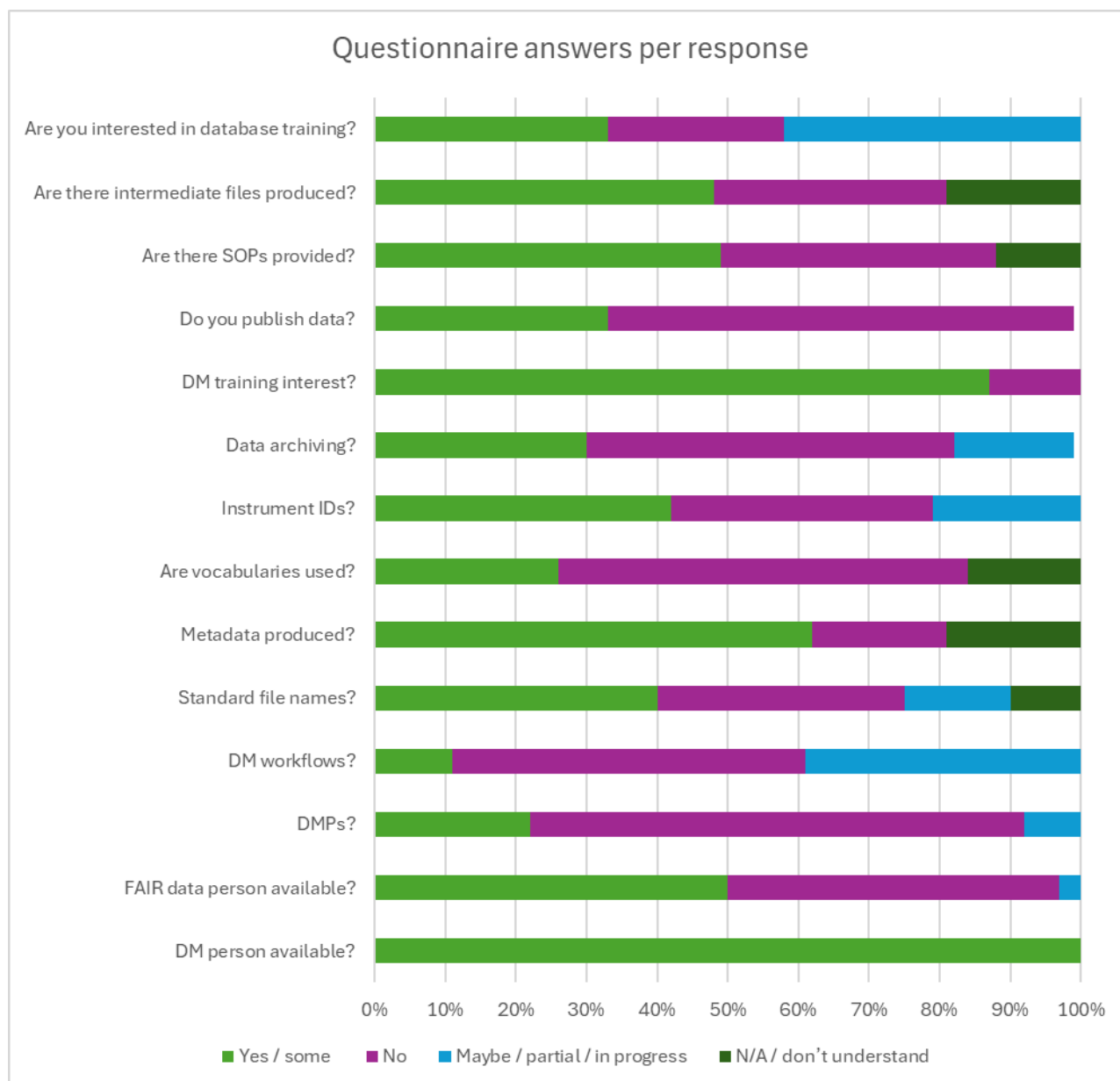


Figure 2: The answers (grouped as yes, no, maybe, NA) for the questions related to DM practises and training summarised over all questionnaire responses

3.1 The quality of the responses

Respondents to the questionnaire were asked to identify which service provider and for which service/instrument they were responding, but most did not provide the requested details. We received replies from 22 respondents, from 21 unique beneficiaries/service providers; compared to 63 service providers which between them provide over 500 unique services (see Figure 4).

Survey answers were classified as follows:

- **Adequate:** Answers were provided to all questions with responses beyond simple "yes" or "no."

- **Not adequate:** All answers not considered adequate.
- **N/A:** Assigned if, based on the other answers given, the service being described in the questionnaire is expected to not produce data that needs to be managed (e.g. an instrument that just gives a readout, a service that is just a physical activity).
- **No answer:** Used when, based on the subcategory (column D) and service description (column E), it was expected that digital data would be produced but no answer was given.

More adequate answers than inadequate answers were received (144 vs. 128). However, the number of responses received represented fewer than half of the total number of services in the AQUASERV catalogue (analysis can be found in quality of responses spreadsheet¹⁰). Figure 2 shows the number of questions with these different levels of response. In Figures 3 and 4 we show a summary of the quality of the answers categorised by service category as classified in ARIA. In Figure 5 we show the quality of the answers by service category, in order of the number of actual services provided in each category overall by AQUASERV service providers.

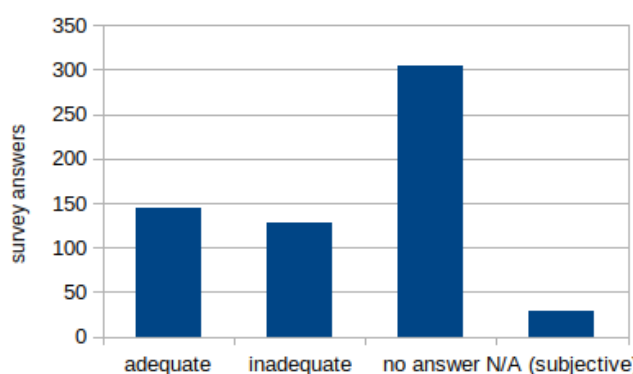


Figure 3: Quality feedback from both the Activities and Instruments questionnaires, plotted against the number of questions asked

¹⁰ <https://zenodo.org/records/14781098> (Quality of responses.xlsx)

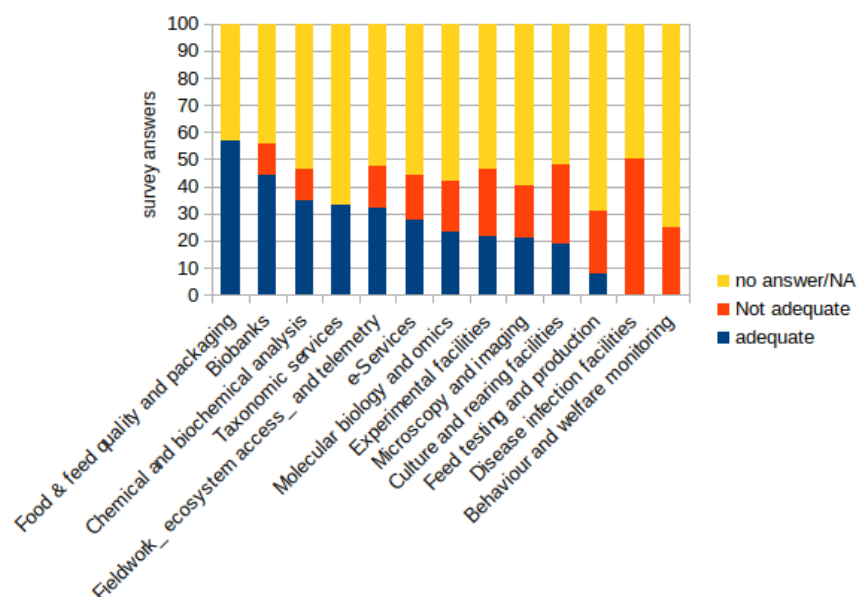


Figure 4: Quality of survey answer grouped by service category and ordered proportionally by survey answer quality

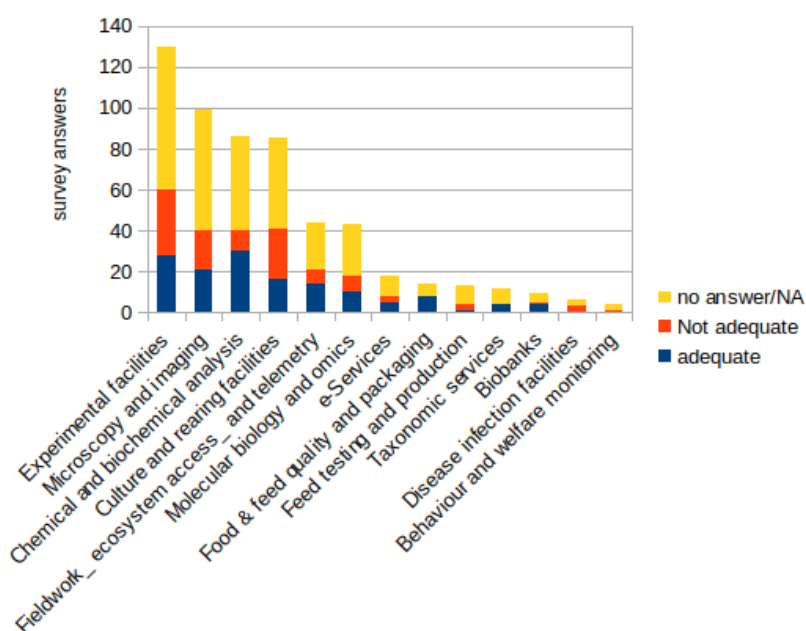


Figure 5: Quality of the survey answer grouped by service category and ordered by number of services in each category

The lack of adequate responses was likely, at least partially, due to a lack of data (management) literacy among respondents. Additionally, the quality of the survey responses appears to be more closely linked to the respondents' affiliation with specific beneficiaries rather than any technical categorisation (Figure

6). Beneficiaries who were heavily invested in the project may have been more likely to provide detailed responses, particularly those with a stronger focus on DM practices.

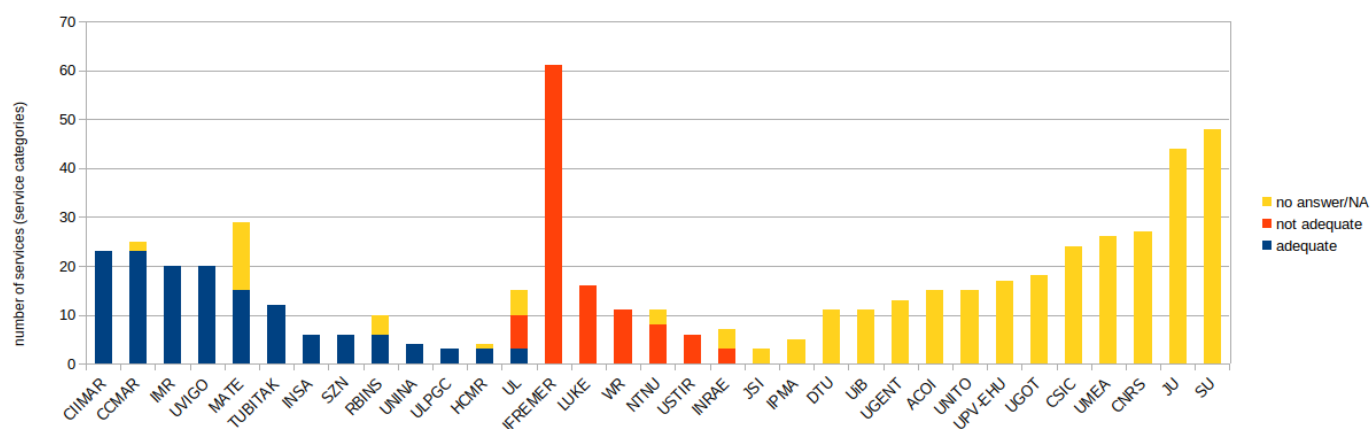


Figure 6: Number of services with survey answers grouped by hosting beneficiary

Key Findings

- **Follow up will be necessary:** We will need to follow up on the missing responses as fewer than half of the expected responses were received.
- **Accommodate lack of data literacy:** These questionnaires will need to be rephrased, or we will need to adopt a face-to-face interview approach, to accommodate the inadequate understanding of data management and hence the questions asked.

3.2 Summary of data types

The data types reported in the questionnaires are:

- Spreadsheet (EXCEL, CSV)
- Images (JPG, TIFF, PDF)
- Videos (MOV, AVI, MP4)
- Sequence data (FASTA, FASTQ)

Additionally, several instrument-specific and potentially proprietary formats were indicated. Many institutions also use specialised file types associated with their research equipment, including formats for mass spectrometry, chromatography, and microscopy. Some respondents highlighted challenges in managing and standardising these diverse data types, particularly when dealing with proprietary formats which can complicate data sharing and long-term preservation.

Key Findings

- **Standard data types:** Most of the data types reported are widely accessible and not proprietary. However, follow-up to check on this for the most “complex” instruments will be necessary.

○ 3.3 Summary of data management practices

A key aspect of effective data management is having established a data workflow, particularly for data generated by instruments. Setting these up and running them involves understanding the methods used to archive data and the mechanisms for data sharing, as well as the role that metadata and vocabularies should play in the data sharing. It is also crucial to assess whether metadata is created, properly linked to the data, and formatted according to (community-agreed) standards. Additionally, the interoperability of metadata should be examined, including the use of controlled vocabularies and standardised measurement units. Finally, ensuring that operating protocols — especially those related to instruments and software usage — are available is essential for maintaining robust DM and sharing practices. These aspects of DM were analysed from the questionnaires (see Tables A2 and A3) and our findings are summarised here.

■ 3.3.1 Practices

Data management practices were assessed across several areas, including data archiving, workflows, management plans, and adherence to Findability, accessibility, interoperability, and reusability (FAIR) data principles (referenced in Tables A2 and A3). All respondents reported having a designated individual responsible for DM, and nearly all confirmed their ability to assist TA users with data publication. However, 40-50% of the individual stations indicated a lack of staff with expertise in FAIR data, highlighting a significant skills gap (Figure A2). While data can be published, it is not always guaranteed to meet FAIR standards. Further follow-up is needed to evaluate the expertise of the designated data managers regarding FAIR data, clarify their responsibilities, and identify any training requirements.

More than half of the respondents indicated a lack of documented data workflows, although many mentioned having partial workflows or being in the process of developing them. Further follow-up is needed to understand the specifics of these developments. The responses regarding data archiving were mixed, with an equal split between stations that had "yes", "no", or "partial" archiving practices or were in the process of implementing them. However, based on follow-up questions about the type of archives and the storage of data and metadata, it became clear that many stations rely on file-based storage

systems (e.g. spreadsheets) rather than structured database systems. Additional follow-up is required to gather more detailed information on this aspect of data management.

The relevance of data publication varies significantly across stations, particularly where data collection is conducted primarily on behalf of third parties rather than by the service providers themselves. Only about 30% of beneficiaries publish data regularly or occasionally, highlighting a general lack of experience in data publishing (Figure A2). This gap may limit the level of assistance service providers can offer to TA users in this domain.

Key Findings

- **Variability in Data Management Practices:** Surveyed institutions demonstrate significant differences in adopting structured DM practices. Many are in the early stages of developing formal Data Management Plan (DMPs) and workflows aligned with FAIR principles. Some institutions are still defining their understanding of “data management,” which differs from the definitions outlined in WP3.
- **Inconsistencies in Data Publication:** Data publication practices vary widely. Half of the institutions either do not publish data routinely or only do so at the discretion of external parties. For data collected on behalf of third parties — particularly those with commercial interests — data-release policies are typically determined by those external agents.
- **Opportunities for Open Science:** Promoting open access to data collected through service providers would significantly benefit open science. This includes ensuring the provision of necessary metadata, standardisation, and provenance information to facilitate data sharing. While some service providers are actively improving their data sharing capabilities, many still lack clear, standardised processes for metadata management, naming conventions, and workflows.

To address the identified gaps, several follow-up actions will be undertaken. Service providers would benefit from targeted training on data archiving, workflows, and FAIR data management practices to build their capacity. In addition, conducting detailed assessments of specific service providers’ data management practices will help ensure alignment with global best practices in data sharing, storage, and publication. Identifying and promoting “FAIR data management champions” — service providers that excel in implementing best practices — will offer a peer-led model for improvement. These champions can provide actionable, practical advice to other providers by leveraging their firsthand experience in managing stations and laboratories. Together, these steps would foster a more cohesive and

FAIR-aligned data landscape.

No service provider indicated that they used electronic laboratory notebooks, which is unfortunate as they are a useful means for managing data arising from laboratory work and therefore potentially a way into FAIR data management.

■ 3.3.2 Vocabularies, file naming, protocols, and instrument IDs

The survey explored several additional aspects of data management: the use of vocabularies, metadata provision, standard operating procedures (SOPs) / protocols, file-naming conventions, the saving of intermediate or “unwanted” data files, and the use of permanent and unique identifiers for instruments. These practices are critical for ensuring data interoperability and reusability.

A key observation is the inconsistent application of standardised vocabularies across institutions in data reporting and metadata management. Survey responses indicate some confusion among respondents regarding the distinction between standards (e.g. International System of Units (SI units), ISO formats¹¹) and vocabularies (e.g. BODC¹², ENVO¹³, MIXS¹⁴, or the marine terminologies WoRMS¹⁵ and Marine Regions¹⁶). This highlights the need for follow-up actions, including clarification and training on this topic.

Approximately half of the respondents reported using unique IDs for instruments, but they could not give many examples. Follow-up is required to explain their understanding of what IDs are for and how to create and use them in FAIR metadata. While this practice may be more closely related to lab management than traditional DM, using instrument IDs allows for the unique and unambiguous identification of instruments, which greatly enhances data reusability and supports provenance tracking.

The survey also addressed the use of publicly-accessible/available SOPs for instruments and activities. However, upon reflection, the phrasing of this question could have been improved. In many cases, it is the user of the instrument or activity who determines the appropriate SOP. That said, certain instruments and activities — such as field sampling services — should ideally have established SOPs. Follow-up actions should assess whether this is relevant to AQUASERV services and ensure that relevant activities are adequately documented.

¹¹ <https://www.iso.org/standards.html>

¹² https://vocab.nerc.ac.uk/search_nvs/

¹³ <https://sites.google.com/site/environmentontology/>

¹⁴ <https://www.gensc.org/pages/standards-intro.html>

¹⁵ <https://www.marinespecies.org/>

¹⁶ <https://www.marineregions.org/about.php>

Key Findings

- **Lack of awareness of instrument Identifications (IDs):** Additional work will be needed to better understand how these are created and used by our service providers, and to recommend how they should be used in FAIR metadata. This will make the resulting data more reusable and reproducible.
- **An incomplete understanding of vocabularies:** A need to clarify the difference between standards and vocabularies, to identify which standards and vocabularies are most useful to the service providers, and to improve their use of these in the (meta)data. This can be achieved largely via training courses.
- **The documentation and sharing of protocols:** Greater understanding is needed of the importance of openly publishing SOPs for instruments or activities. This improves the reusability and reproducibility of resulting data.

○ 3.4 Training

Two key questions were posed to assess the training needs of AQUASERV stations: whether they would benefit from training in setting up an internal database for data archiving and whether they were interested in training on FAIR data principles. A clear interest was expressed in both areas.

- For database training, 6 stations indicated they would benefit from this support, while 4 noted they would need to consult their management before committing
- Interest in FAIR data training was strong, with 7 responding positively. Additionally, 11 stations indicated a potential interest in FAIR training.

Furthermore, several stations identified specific gaps in DM practices that could be addressed with targeted support. These stations should be encouraged to apply for internal WP3 funding to meet their needs. To respond to this demand, WP3 will review its resources and plans to develop an online, self-paced FAIR data course, which will be made available to all AQUASERV TA users and participating stations.

Key Findings

1. **A desire for training is expressed:** An Aquaserv FAIR data online, self-paced training course will be developed for service providers and service users both to use.
2. **Data management training:** Focussed training for the service providers on data archiving, metadata management, and data management, would find a ready audience. WP3 will see if they can accommodate this. The WP3 internal funding call should also be used by the stations to plug

gaps they have already identified in these areas.

4. Summary and follow-up

The review of the questionnaire responses indicates an absence of raw data types specific to the instrument. It is expected that unprocessed data, collected directly from the instrument, to be reported without any subsequent manipulation (such as organization or analysis). Clarifying this is critical for understanding the data landscape and ensuring comprehensive DM planning. Where data have been indicated as being instrument-linked, further investigation is needed to assess the proprietary nature of these formats. Additionally, it is essential to explore how such proprietary formats could be converted to open access (OA) formats without compromising metadata integrity, as this will support data interoperability and accessibility. It is important to emphasize that both the raw or primary data and its immediate interpretation must be properly managed and made FAIR. This concept may appear to be a novel idea for some facility managers.

There is also ambiguity surrounding the origin of spreadsheet data. It remains unclear whether spreadsheet formats, which are predominantly noted as Excel, are automatically generated by the instruments or manually created by users as a summary interpretation of the raw instrumental data. This distinction is vital, as the appropriate DM advice depends heavily on the source of the data. Resolving this uncertainty will enable the development of more precise and relevant guidance for users.

To better understand the current state of data practices, obtaining and inspecting example datasets is a priority. These datasets must be evaluated for their adherence to FAIR principles to ensure that they meet the standards for good DM and interoperability. While the diversity of data formats is to be expected given the specific requirements of different research domains, there is significant commonality that can simplify the creation of templates for TA DMPs.

Further clarification is needed on the processes for handling, exporting, and sharing raw data and metadata, especially as many responses regarding these aspects were either vague or incomplete. Only with this clarity can effective recommendations be made to improve the standardisation of data formats across instruments and services. Ensuring consistency in these practices will facilitate better data interoperability and alignment with FAIR data principles. Special attention is required for spreadsheet

data, given their widespread use, to provide targeted advice on creating interoperable spreadsheets that align with the goals of TA activities.

To address the gaps identified as our **Key Findings**, we will prioritise training programmes focused on FAIR data principles which aligns with the strong interest expressed by many institutions. Additionally, efforts should focus on providing guidance for: creating structured workflows, data publishing, data archiving, and improvements in metadata use. Collaborative initiatives should also be encouraged to support institutions actively working to enhance their DM practices.

As the response rate to our questionnaire was very poor, we will also need to supplement the existing questionnaires so that they are easier for the service providers to understand and fill in for their data-producing services. At a minimum these will need to be answered by the non-responsive service providers. It could be the survey may not have been distributed to the individuals operating the individual services. While there is certainly room for improvement in the survey itself, the primary focus should be on ensuring it reaches the appropriate individuals.

It would be helpful for WP3 to provide a guide to data management that the stations can use as a check-list to compare their data management activities to.

In conclusion, this review has highlighted key areas where further follow-up is necessary. Addressing these gaps will enhance DM practices, ensure adherence to FAIR principles, and support the development of standardised approaches across instruments and services. This effort is crucial for achieving robust and interoperable data-sharing practices in support of research and collaboration.

5. Opportunities to improve data management practices

The results of our survey reveals a diverse landscape of DM practices among services offered by the AQUASERV partners. By assessing the survey responses we have identified existing gaps in DM practices and areas requiring remedial enhancement to increase data interoperability among consortium members and, more generally, within the wider Life Science data spaces. Funding opportunities within AQUASERV (administered by WP3) will be made available to partners to provide solutions to existing problems in DM practices based on this (D3.2) understanding of the current DM landscape. This funding will be provided In consultation with the service providers, and proposed actions will be triaged with respect to urgency and feasibility of the proposed remedial actions. Partners that have provided detailed responses to our survey will be placed in an advantageous position for funding opportunities (having participated in the data landscaping will be a requirement for application for funding), but we intend to

consult with all partners to more fully understand their needs and thereby enable a more beneficial distribution of the funded enhancements to the project as a whole.

• APPENDIX

Sunburst plots giving an overview of the services provided in AQUASERV.

Figure A1a: Sunburst chart summarizing the range and scope of services that are on offer in AQUASERV organised by category and subcategory

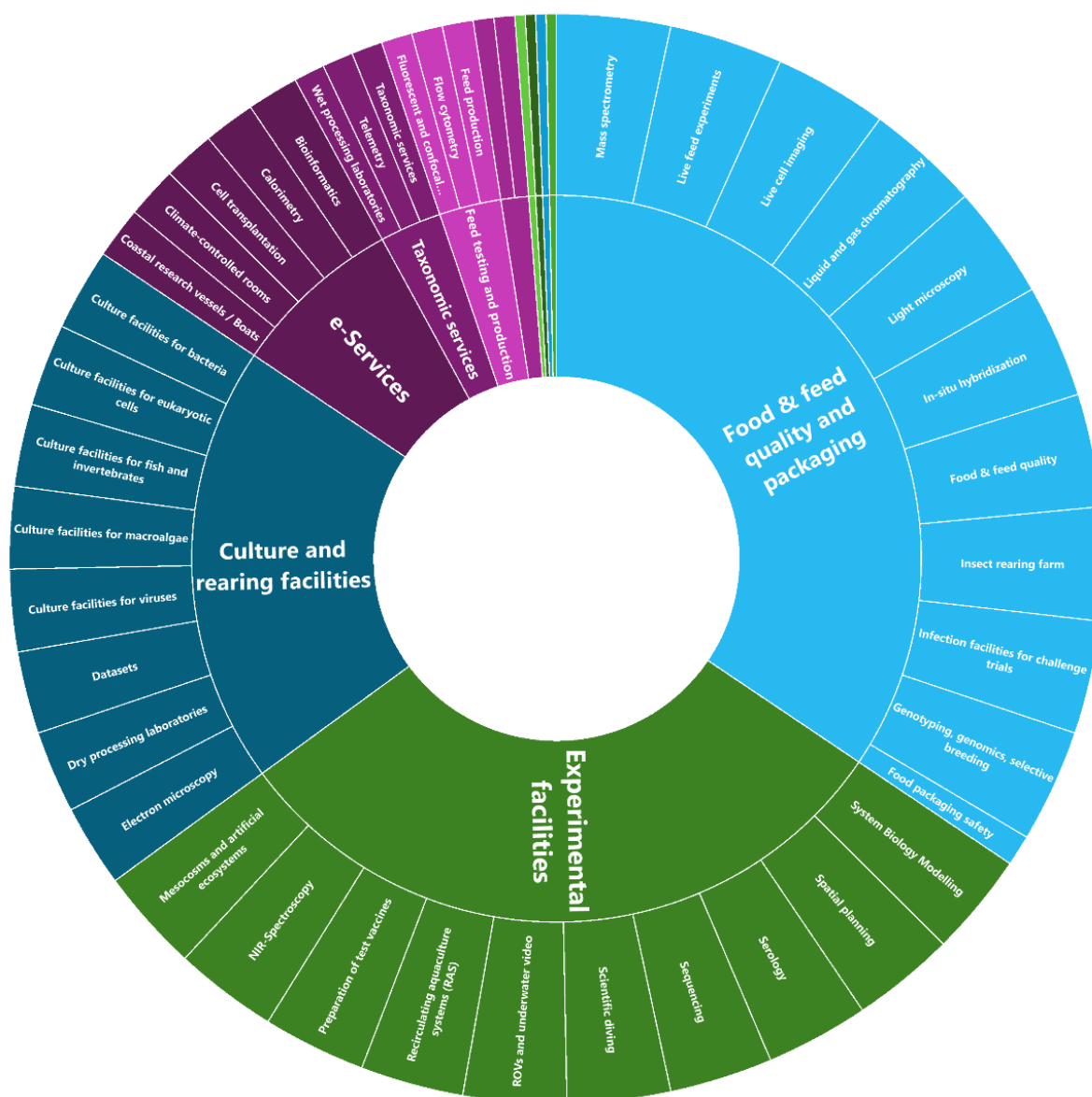


Figure A1b: A expanded view of the unlabeled sections at the top of Figure A1a



Summary of services providers, service categories, and data types

The table lists the beneficiaries and their service providers (by acronym as used in ARIA), along with the service categories they provide and the data types identified in the questionnaires. Many questionnaire responses did not specify the service provider or the service/instrument type, as the ARIA service categories were not established at the time. We have subsequently mapped the available information to the service categories as accurately as possible. Service categories likely to include instruments that generate data files (rather than simple readouts) are highlighted in italics in Table A1. These categories are based largely on the descriptions provided by service providers in ARIA.

Table A1: Summary of service categories from the providers and the data types arriving from those services (null indicates no answer to the question/questionnaire)

Beneficiary	Service Providers	Service Categories provided	Data types
CNRS	Metatron, Planaqua	Culture and rearing facilities; Experimental facilities; <i>Microscopy and imaging; Chemical and biochemical analysis; Molecular biology and omics; Fieldwork; ecosystem access, and telemetry; Taxonomic services</i>	null
CSIC	IATS EXP, IATS ANA, IEO,	Culture and rearing facilities; Experimental facilities; <i>Microscopy and imaging; Behaviour and welfare monitoring; Chemical and biochemical analysis; Molecular biology and omics; e-Services; Fieldwork, ecosystem access, and telemetry; Disease infection facilities</i>	null
DTU	DTU	Culture and rearing facilities; Disease infection facilities; <i>Experimental facilities; Microscopy and imaging; Chemical and biochemical analysis; Molecular biology and omics</i>	null
IFREMER	MET, MEB, PEMP, BMSA, FMSP	Culture and rearing facilities; Disease infection facilities; Experimental facilities; <i>Microscopy and imaging; Chemical and biochemical analysis; Fieldwork, ecosystem access, and telemetry; Molecular biology and omics; e-Services; Feed testing and production</i>	.xlsx, .csv
INRAE	IERP, LPGP, STPEE, PEIMA	Culture and rearing facilities; Fieldwork, ecosystem access, and <i>telemetry; Food & feed quality and packaging; Disease infection facilities; Experimental facilities; Microscopy and imaging; Molecular biology and omics; Behaviour and welfare monitoring; Feed testing and production; Chemical and biochemical analysis; Biobanks;</i>	null
INSA	INSA	Food & feed quality and packaging; <i>Chemical and biochemical analysis</i>	ICP-OES: raw data in the format ldf; and spreadsheet xlsx, DMA: raw data in format M80
IPMA	IPMA	Food & feed quality and packaging; <i>Chemical and biochemical analysis</i>	null
JU	LIFD, IAPW, IFA, ICS, GRC	Culture and rearing facilities; Disease infection facilities; <i>Microscopy and imaging; Fieldwork, ecosystem access, and telemetry; Feed testing and production; Food & feed quality and packaging; Experimental facilities; Chemical and biochemical analysis; Behaviour and welfare monitoring; e-Services; Molecular biology and omics</i>	null

JSI	JSI	<i>Food & feed quality and packaging; Chemical and biochemical analysis; e-Services</i>	null
LUKE	KFRS, Enonkoski, LARS, PMRS, SelBreed, FINFARMGIS, Otolab	<i>Culture and rearing facilities; Fieldwork, ecosystem access, and telemetry; Experimental facilities; Feed testing and production; Molecular biology and omics; e-Services; Microscopy and imaging</i>	.txt, excel
MATE	POND, RAS, NUTRI, DANIO, GENO, CBL	<i>Culture and rearing facilities; Experimental facilities; Microscopy and imaging; Chemical and biochemical analysis; Molecular biology and omics; Feed testing and production;</i>	null
ULPGC	PCTM	<i>Culture and rearing facilities; Experimental facilities</i>	null
TUBITAK	TUBITAK	<i>Fieldwork, ecosystem access, and telemetry; Food & feed quality and packaging; Experimental facilities; Microscopy and imaging; Chemical and biochemical analysis</i>	spreadsheet xlsx, instrument specific file formats such as (.raw) digital image (.jpg, .png, or .tiff).
UL	UL EPA	<i>Culture and rearing facilities; Feed testing and production; Experimental facilities; Microscopy and imaging; Chemical and biochemical analysis; Molecular biology and omics</i>	null
UNINA	UNINA	<i>Food & feed quality and packaging; Chemical and biochemical analysis</i>	null
UNITO	Aqua, Insects	<i>Culture and rearing facilities; Feed testing and production; Chemical and biochemical analysis; Molecular biology and omics; e-Services; Experimental facilities</i>	null
USTIR	MERL, NBFUR	<i>Experimental facilities</i>	.xlsx, .jpg, .csv, .fastq, .avi, .mpg
WR	WR	<i>Culture and rearing facilities; Biobanks; Experimental facilities; Chemical and biochemical analysis; Molecular biology and omics; Behaviour and welfare monitoring</i>	Spreadsheets, digital image formats
CCMAR	CCMAR	<i>Culture and rearing facilities; Fieldwork, ecosystem access, and telemetry; Food & feed quality and packaging; Taxonomic services; Experimental facilities; Microscopy and imaging; Chemical and biochemical analysis; Molecular biology and omics; e-Services</i>	spreadsheet (.csv), digital images (.tif, .jpg), Instrument (.raw)
CIIMAR	CIIMAR	<i>Culture and rearing facilities; Fieldwork, ecosystem access, and telemetry; Taxonomic services; Biobanks; Experimental facilities; Microscopy and imaging; Chemical and biochemical analysis; Molecular biology and omics; e-Services</i>	.jpg, .csv, .fastq, .xlsx, etc.
UVIGO	ECIMAT UVIGO	<i>Culture and rearing facilities; Fieldwork, ecosystem access, and telemetry; Taxonomic services; Biobanks; Experimental facilities; Microscopy and</i>	Spreadsheets (.xlsx, .csv),

		<i>imaging; Chemical and biochemical analysis; Molecular biology and omics; e-Services</i>	digital images mostly (jpg)
HCMR	HCMR IMBBC	Culture and rearing facilities; Fieldwork, ecosystem access, and <i>telemetry</i> ; Taxonomic services; <i>Experimental facilities</i>	spreadsheets (.xlsx)
SU	IMEV, OOB, SBR	Culture and rearing facilities; Fieldwork, ecosystem access, and <i>telemetry</i> ; Biobanks; <i>Experimental facilities</i> ; <i>Microscopy and imaging</i> ; <i>Molecular biology and omics</i> ; <i>Taxonomic services</i> ; <i>Chemical and biochemical analysis</i>	null
IMR	IMR	Culture and rearing facilities; Fieldwork, ecosystem access, and <i>telemetry</i> ; Feed testing and production; Taxonomic services; Biobanks; <i>Experimental facilities</i> ; <i>Microscopy and imaging</i> ; <i>Chemical and biochemical analysis</i> ; <i>Molecular biology and omics</i> ; <i>e-Services</i>	Spreadsheets (csv, excel)
NTNU	NTNU	Culture and rearing facilities; Fieldwork, ecosystem access, and <i>telemetry</i> ; <i>Experimental facilities</i> ; <i>Microscopy and imaging</i>	null
UPV-EHU	PiE UPV EHU	Culture and rearing facilities; Fieldwork, ecosystem access, and <i>telemetry</i> ; <i>Taxonomic services</i> ; Biobanks; <i>Experimental facilities</i> ; <i>Microscopy and imaging</i> ; <i>Chemical and biochemical analysis</i> ; <i>Molecular biology and omics</i>	null
ACOI	ACOI	Culture and rearing facilities; Fieldwork, ecosystem access, and <i>telemetry</i> ; Food & feed quality and packaging; <i>Taxonomic services</i> ; Biobanks; <i>Experimental facilities</i> ; <i>Microscopy and imaging</i> ; <i>Chemical and biochemical analysis</i> ; <i>e-Services</i>	null
RBINS	RBINS	Fieldwork, ecosystem access, and <i>telemetry</i> ; Taxonomic services; Biobanks; <i>Experimental facilities</i> ; <i>Chemical and biochemical analysis</i> ; <i>e-Services</i>	null
SZN	SZN	Fieldwork, ecosystem access, and <i>telemetry</i> ; <i>Taxonomic services</i> ; <i>Experimental facilities</i> ; <i>Microscopy and imaging</i> ; <i>Chemical and biochemical analysis</i> ; <i>e-Services</i>	null
UGENT	UGENT	Culture and rearing facilities; Feed testing and production; Taxonomic services; <i>Experimental facilities</i> ; <i>Microscopy and imaging</i> ; <i>Chemical and biochemical analysis</i>	null
UGOT	KMRS, TML	Fieldwork, ecosystem access, and <i>telemetry</i> ; <i>Experimental facilities</i> ; <i>Microscopy and imaging</i> ; <i>Molecular biology and omics</i> ; <i>Chemical and biochemical analysis</i>	null
UiB	UiB	Culture and rearing facilities; Fieldwork, ecosystem access, and <i>telemetry</i> ; Taxonomic services; <i>Experimental facilities</i> ; <i>Microscopy and imaging</i> ; <i>Chemical and biochemical analysis</i>	null
UMEA	UMEA	Culture and rearing facilities; Fieldwork, ecosystem access, and <i>telemetry</i> ; <i>Taxonomic services</i> ; <i>Experimental facilities</i> ; <i>Microscopy and imaging</i> ; <i>Chemical and biochemical analysis</i> ; <i>Molecular biology and omics</i> ; <i>e-Services</i>	null

Questionnaires: categorisation of the questions

As part of our analysis, a number of the questions were grouped into four categories: data management, training, data types, and general/other. This is shown in Table A2 for the activities questionnaire and Table A3 for the instruments questionnaire.

Table A2: Questions asked in the questionnaire: **Activities**. Question grouping: Data types, Data management, Training, General/ Other

Question grouping: Data types, Data management, Training, General/ Other	Questions
General/ Other	1. Which station are you, and which service are you describing here? Example answer: "CCMAR-Faro, Sequencing platform: Oxford Nanopore Minion"
General/ Other	2. Who is compiling this information? (position, name and email address)
Data management	3. Person who will be responsible for data management at the service provider (name and email address, ORCID if available)
Data management	4. Is there someone at the service provider who understands FAIR data, Open data, and data management and could help TA users with this? If yes, give their contact details (name, email, ORCID).
Data types	5. What type of digital object(s) (e.g spreadsheet, digital images) and what type of file(s) (e.g xlsx, jpg, csv, fastq) are typically produced by people using this service? (e.g. "photographs of type jpg", "spreadsheets of various types").
Data management	6a. Are there (Standard) Operating Procedures (SOPs) followed for this activity? SOPs are agreed protocols (i.e. not protocols used by someone and modified at will by others) allowing for repeatable measurements. They can be as simple as a small note next to a machine (switch on, place the sample in a holder, click on the green button, wait, write down the value, remove sample, switch off) or multi-page explanation - here an example of nitrogen concentration measurement (https://www.sfwmd.gov/sites/default/files/documents/sfwmd-lab-sop-3090-003.pdf). You may store your SOPs on a piece of paper, as an internal database or online.
Data management	6b. If yes to question 6a: please give the URLs (links) to those. If they are not available online, give the contact details of the person who can provide these to us.

Data management	7. Can anyone at the station assist/train/advise the AQUASERV TA users in publishing their data from the service?
Data management	8. Does your station have a/any data management plan(s) that you follow for particular services or activities? If yes, provide the link to that DMP or indicate how we can see a copy of it/them.
Data management	9. Does your station have organised data management workflows, to turn collected/raw data into FAIR data?
Data management	10. Are the instruments and devices used in experimental/field work in your station given unique IDs or names (to which they can be referred to for re-usability purposes in data)? For example, a code or a name that only that instrument has and which is recorded somewhere.
Data management	11. If yes to the question above: how are these IDs/names constructed (it is easier to just give some examples).
Data management	12. Is there an internal database where the data produced by the service is archived?
Training	13. If no to the question above: if AQUASERV could offer advice and training, would the station like to set up/use a database for this service?
Data management	14. If yes to the question on "internal database": is this database accessible?
Data management	15. If yes to the question on "internal database": what type of database it is (e.g. MongoDB, MS Access files, MS Excel files, google sheets, relational database) Your instrument may not allow for storing any data, requiring you to export and store it in a dedicated database. However, many modern instruments can store data on the internal/connected computer or cloud. You might be using such a database directly through the instrument's software without knowing its type. To check, look for files like config.php, database.yml, or those with a .env extension. There are also many more advanced methods to identify the database type. If you struggle with this, please let us know if you would like our assistance.
Data management	16. If yes to the question on "internal database": how are the data and the metadata archived?
Training	17. If AQUASERV offered data management and FAIR data training, would anyone at your station be interested (please be honest!).
Data management	18. Is additional data collected/produced by your service that is typically discarded (not published) that in other contexts could be valuable? Possibly "null" data, or "noise". If yes, why are these data typically discarded?

General/ Other	19. Do you offer the use of electronic laboratory notebooks that TA users could use?
Data management	20. Does your station commonly publish data produced by your services to community repositories? If yes, please indicate for which services and which repositories are used.
Training	21. Can you identify any areas where you believe the current procedures for the management of data could be improved in your station?
General/ Other	22. Any other information or discussion you would like to provide...

Table A3: Questions asked in the questionnaire: **Instruments**. Question grouping: Data types, Data management, Training, General/ Other

Question grouping: Data types, Data management, Training, General/ Other	Questions
General/ Other	1. Which station are you, and which service are you describing here? Example answer: "CCMAR-Faro, Sequencing platform: Oxford Nanopore Minion"
General/ Other	2. Who is compiling this information? (position, name and email address)
Data management	3. Person who will be responsible for data management at the service provider (name and email address, ORCID if available)
Training	4. Is there someone at the service provider who understands FAIR data, Open data, and data management and could help TA users with this? If yes, give their contact details (name, email, ORCID).
Data types	5. What type of digital object(s) (e.g spreadsheet, digital images) and what type of file(s) (e.g xlsx, jpg, csv, fastq) are typically produced by people using this service? (e.g. "photographs of type jpg", "spreadsheets of various types").
Data management	6a. Are there (Standard) Operating Procedures followed for this activity? SOPs are agreed protocols (i.e. not protocols used by someone and modified at will by others) allowing for repeatable measurements. They can be as simple as a small note next to a machine (switch on, place the sample in a holder, click on the green button, wait, write down the value, remove sample, switch off) or multi-page explanation - here an example of nitrogen concentration measurement

	(https://www.sfwmd.gov/sites/default/files/documents/sfwmd-lab-sop-3090-003.pdf). You may store your SOPs on a piece of paper, as an internal database or online.
Data management	6b. If yes to question 6a: please give the URLs (links) to those. If they are not available online, give the contact details of the person who can provide these to us.
Data management	7. Can anyone at the station assist/train/advise the AQUASERV TA users in publishing their data from the service?
Data management	8. Does your station have a/any data management plan(s) that you follow for particular services or activities? If yes, provide the link to that DMP or indicate how we can see a copy of it/them.
Data management	9. Does your station have organised data management workflows, to turn collected/raw data into FAIR data?
Data management	10. Does the naming of the data files follow a standard format -- either created by the instrument or created by your organisation -- or is the user of the service free to create their own filenames? If there is a standard, please give examples (for example Nanopore_JohnSmith_26062024.fastq) and a short description, i.e. equipment_username_data.extension)
Data types	11a. Do the digital objects produced by the instrument have associated metadata. Metadata may be placed in an independent file, separate spreadsheet or may even be a part of the data. By metadata we understand any type of data describing the main (raw) data. For example instrument settings, experiment time and duration, etc.
Data management	11b. If yes to the question above: is that metadata also made available for the TA user to access/copy/take home?
Data types	12a. Do the digital objects produced by the instrument have associated intermediate files (e.g. raw as well as processed images, analysed or semi-analysed data) that are produced separately.
Data management	12b. If yes to the question above: are those additional data also made available for the TA user to access/copy/take home?
Data management	13a. In the data produced by the instrument, are standard vocabularies used ? These vocabularies are sets of standardised terms with persistent and unique identifiers (called PIDs, normally a link to a website with a description of such a term, a list of related terms, etc.). Using standard vocabularies greatly improves interoperability of data (i.e.

	<p>researchers from other places/institutes are able to quickly understand what has been measured/sampled in your lab)</p> <p>Examples of vocabularies:</p> <p>ISO standards - e.g. SI units as length in meters (m) rather than inches (in)</p> <p>WoRMS - e.g. Oomycota rather than Oomycetes (https://www.marinespecies.org/aphia.php?p=taxdetails&id=172230)</p> <p>MarineRegions - e.g. Latvian part of the Baltic Sea rather than "Baltic" or "Latvian Baltic" or "20 km NE from a city of X". Such a region comes with a unique identification number (here 25669) and a dedicated website link - i.e. PID (http://marineregions.org/mrgid/25669)</p> <p>ENVO - e.g. concentration of organic nitrogen anion in soil rather than "soil N" or "N conc". Such term has its own PID (website link) of https://www.ebi.ac.uk/ols4/ontologies/envo/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FENVO_3200066</p> <p>There are other vocabularies lists that may be more relevant to your research field.</p> <p>Also, to check whether your terms/description can be considered as standard vocabulary you can either use automated tools and libraries like:</p> <p>OpenRefine: A tool for working with messy data, it supports reconciliation against various vocabularies.</p> <p>SHACL (Shapes Constraint Language): Used for validating RDF data against vocabularies.</p> <p>PySHACL: A Python library for SHACL validation.</p> <p>RDFLib: A Python library to work with RDF, useful for parsing and validating against vocabularies.</p> <p>B. Manual Inspection</p> <p>or manually inspect a sample of your data and metadata:</p> <p>Extract Terms: Identify the terms used in your metadata fields.</p> <p>Compare Terms: Compare these terms with the terms in the known vocabularies (e.g. WoRMS, MarineRegions)</p> <p>Consistency Check: Ensure that the terms are used consistently according to the definitions in the vocabulary.</p>
General/ Other	13b. If yes to question 13a: please provide the names and home page (web address) of those vocabularies.
Data management	13c. If no to question 13a, how are the data described/labelled so that they can be understood by the data user?
General/ Other	13d. Can you provide us with some example data files for this instrument upon request?

Data management	14a. In the metadata produced by the instrument, are standard vocabularies (e.g. ISO standards, WoRMS, MarineRegions, ENVO...) used?
General/ Other	14b. If yes to question 14A: please provide the names and home page (web address) of those vocabularies.
Data management	14c. If not to question 14a: how are metadata described/labelled so that they can be understood?
General/ Other	14d. Can you provide us with some example metadata files for this instrument upon request?
Data management	<p>15. Are the instruments in your station given unique IDs or names (to which they can be referred to for re-usability purposes in data)?</p> <p>For example, a code or name that only that instrument has and is recorded somewhere—e.g., you may call the microplate reader "blue" or "red" or "Thermo," "the one with luminescence," but such terms would mean little to others. We are looking for IDs such as (e.g.) BMG Labtech SPECTROstar Nano 601-101-12 (catalog name). Your instrument may also have a unique ID given by the manufacturer, which will allow you to know which machine produced the data in case you have multiple copies of the same model.</p>
Data management	16. If yes to the question above: how are these IDs/names constructed (it is easier to just give some examples)
Data management	17. Is there an internal database where the data produced by the service is archived?
Training	18. If no to the question above: if AQUASERV could offer advice and training, would the station like to set up/use a database for this service?
General/ Other	19. If yes to the question on " internal database": is this database accessible?
Data management	<p>20. If yes to the question on " internal database": what type of database it is (e.g. MongoDB, MS Access files, MS Excel files, google sheets, relational database).</p> <p>Your instrument may not allow for storing any data, requiring you to export and store it in a dedicated database. However, many modern instruments can store data on the internal/connected computer or cloud. You might be using such a database directly through the instrument's software without knowing its type. To check, look for files like config.php, database.yml, or those with a .env extension. There are also many more advanced methods to identify the database type. If you struggle with this, please let us know if you would like our assistance.</p>
Data management	21. If yes to the question on "internal database": how are the data and the metadata archived?

Training	22. If AQUASERV offered data management and FAIR data training, would anyone at your station be interested (please be honest!).
Data management	23. Is additional data collected/produced by your service that is typically discarded (not published) that in other contexts could be valuable? Possibly “null” data, or “noise”. If yes, why are these data typically discarded?
General/ Other	24. Do you offer the use of electronic laboratory notebooks that TA users could use?
Data management	25. Does your station commonly publish data produced by your services to community repositories? If yes, please indicate for which services and which repositories are used.
General/ Other	26. Can you identify any areas where you believe the current procedures for the management of data could be improved in your station?
General/ Other	27. Any other information or discussion you would like to provide...