Towards a Global Public Repository of Community Protocols to encourage Best Practices in Biomolecular Ocean Observing and Research

Robyn M. Samuel[1, 2*], Raissa Meyer[3], Pier Luigi Buttigieg[4], Neil Davies[5], Nicholas W. Jeffery[6], Christopher Meyer[7], Christina Pavloudi[8], Kathleen Johnson Pitz[9], Maxime Sweetlove[10], Susanna Theroux[11], Jodie van de Kamp[12], Alison Watts[13]

[1]School of Ocean and Earth Science, University of Southampton, Southampton, UK

[2]Ocean Technology and Engineering Group, National Oceanography Centre, Southampton, UK

[3]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

[4]Helmholtz Metadata Collaboration, GEOMAR, Kiel, Germany

[5]Gump South Pacific Research Station, University of California Berkeley, Moorea, French Polynesia

[6]Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada

[7]Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA

[8]Hellenic Centre for Marine Research, Institute of Marine Biology, Biotechnology and Aquaculture, Heraklion, Greece

[9]Monterey Bay Aquarium Research Institute, Moss Landing, California, USA

[10]Royal Belgian Institute of Natural Sciences, Brussels, Belgium

[11]Southern California Coastal Water Research Project, Costa Mesa, California, USA

[12]Commonwealth Scientific and Industrial Research Organisation, Oceans and Atmosphere, Hobart, TAS, Australia

[13]University of New Hampshire, Durham, NH, USA

* Correspondence:
Robyn M. Samuel
r.m.samuel@soton.ac.uk

Abstract

Biomolecular ocean observing and research is a rapidly evolving field that uses omics approaches to describe biodiversity at its foundational level, giving insight into the structure and function of marine ecosystems over time and space. To achieve a global ocean biomolecular observing network (OBON) for the UN Decade of Ocean Science for Sustainable Development and beyond, research groups will need a system to effectively share, discover, and compare omic practices and protocols. While numerous informatic tools and standards exist, there is currently no global, publicly-supported platform specifically designed for sharing marine omics (or any omics) protocols across the entire value-chain from initiating a study to the publication and use of its results. Towards that goal, we propose the development of the Minimum Information for an Omic Protocol (MIOP), a community-developed guide of curated, standardized metadata tags and categories that will orient protocols in

38 the value-chain for the facilitated, structured, and user-driven discovery of suitable protocol suites on the Ocean
39 Best Practices System. Users can annotate their protocols with these tags, or use them as search criteria to find
40 appropriate protocols. Implementing such a curated repository is an essential step towards establishing best
41 practices. Sharing protocols and encouraging comparisons through this repository will be the first steps towards
42 designing a decision tree to guide users to community endorsed best practices.

43 **1   Introduction**

44 The term "omics" generally means studying anything holistically, and here we take a broad view of
45 biomolecular omics that includes, but is not limited to: quantitative target gene amplification (e.g.
46 qPCR, qNASBA etc.), (meta)barcoding, (meta)genomics, (meta)transcriptomics, (meta)proteomics,
47 and metabolomics; and field collection approaches that target organisms or parts thereof, including
48 single-celled organisms (microorganisms), as well as environmental DNA (eDNA). In the marine
49 realm, omic techniques are used to assess and monitor biodiversity, reveal population structure and
50 gene flow, and discover new compounds with applications in medicine and industry. Rapid advances
51 in omics research, and the declining cost of high-throughput sequencing technologies (Wetterstrand,
52 2020) support the increasing application of omics in marine microbiome research.

53 The recent expansion in marine omics has led to a proliferation of protocols specific to multiple
54 applications. However, these protocols are rarely shared publicly with sufficient detail to reliably
55 reproduce a study (Dickie et al., 2018). While the omics community has already achieved high
56 standards for sharing sequence data through the International Nucleotide Sequence Database
57 Collaboration, these data often lack sufficient metadata and provenance information on the protocols
58 used (Dickie et al., 2018), undermining efforts to implement the Findable, Accessible, Interoperable
59 and Reusable (FAIR) data principles (Wilkinson et al., 2016). These limitations create challenges for
60 marine microbiome research and operations from individual labs up to global (meta)data analysis
61 efforts such as MGnify (Mitchell et al., 2020), which must identify data collected using comparable
62 methods, in order to integrate and re-use data for meta-analysis (Berry et al., 2020). Moreover, a lack
63 of protocol-sharing impedes the identification of comparable methods needed for global monitoring
64 efforts aiming to understand, and sustainably manage the changing marine ecosystem (Berry et al.,
65 2020).

66 Many projects are looking to develop best practices for omics research: standards organisations, such
67 as the Genomic Standards Consortium's (GSC) Genomic Biodiversity Interest Group, the
68 Biodiversity Information Standards (TDWG) and the Biocode Commons are working collaboratively
69 towards standards specifications for genomic observatories (Davies et al., 2012; Davies et al., 2014).
70 Large campaigns, such as the Earth Microbiome Project (Thompson et al., 2017), TARA Oceans
71 (Sunagawa et al., 2020), and the Australian Microbiome Initiative (AM; Bissett et al., 2016; Brown et
72 al., 2018; DOI:10.4227/71/561c9bc670099), have already developed standardised practices, and
73 innovative software enterprises, such as protocols.io, are providing powerful solutions for sharing
74 protocols. Yet there is currently no global, publicly-supported infrastructure developed explicitly for
75 encouraging the exchange and harmonization of omic protocols, so these valuable contributions
76 remain fragmented and underutilized.

77 For marine ecosystems, the Intergovernmental Oceanographic Commission's Ocean Best Practices
78 System (OBPS) provides a public repository for all ocean research methodological documentation
79 that can interlink protocols, standard specifications, and other guidelines. The OBPS seeks to support
80 continuous convergence of methods as they undergo community refinement to become best practices
81 (Hörstmann et al., 2021). In collaboration with the broader omics community, through the Omic
82 BON initiative (Buttigieg et al., 2019), we propose to develop a best practice system specific to

83 marine omics research, leveraging the framework of the OBPS to curate a global repository for
84 marine omics protocols.

85 As part of the omics/eDNA session at the 4th OBPS workshop, we discussed recommendations and
86 community needs for an omics/eDNA specific best practices system. Recognizing an urgent need for
87 the ocean omics community to get organized as the UN Decade of Ocean Science for Sustainable
88 Development starts, we identified the demand for publishing protocols into a user-friendly decision
89 tree framework. With such a framework we would aim to support protocol selection, increase
90 protocol findability and improve recognition for protocol developers. In a series of focused follow-up
91 meetings, we identified that an omics decision tree would require a library of constituent parts (the
92 protocols) and framework to: (1) locate where the protocol fits within the entire omics workflow
93 (outlined in section 2), and (2) organise protocols using focused descriptive terms (metadata tags),
94 based on what the protocol does and how/why it is used (outlined in section 3).

2 Ocean Omics Methodology Categories

96 The typical omics workflow involves a series of protocols, which take a project from ideation,
97 through to publication, and on to societal use. Protocols from each step in the omics workflow hold
98 valuable information for different groups. For example, sample collection protocols may be most
99 relevant to scientists/technicians in the field, whereas local stakeholders and indigenous communities
100 may primarily engage with aspects of how the project and resulting data address and impact
101 important ethical, legal, and societal issues. Documenting details and provenance for the entire
102 marine omics workflow requires input from multiple parties, as each step of the workflow may be
103 conducted by different individuals or groups. The omics OBPS therefore needs to identify these key
104 methodological categories, to allow protocols and accompanying metadata to be uploaded in modules
105 that link together to form the entire workflow.

106 We propose twelve protocol categories (Figure 1A) for ocean omics research and operations.
107 Protocols and guidelines are assigned into these categories according to the purpose they serve[1].
108 Categories 5-12 outline methodological categories for operational activities used in the AM Initiative
109 (van de Kamp et al., 2019). Categories 1-4 were identified to additionally cover cross-cutting
110 documentation in the omics workflow: 1) Society, 2) Sampling/observational design, 3) Ethics and
111 law, and 4) Data management.

112 1) Society **-** All workflows should begin and end with society; societal needs inform the question
113   or purpose behind the research, and societal impacts show the value in the research once it has
114   been completed.
115 2) Design and logistics - This category covers the practical logistics for implementing ocean
116   omics research and operations, including the experimental/observational design formulated to
117   address the societal priorities outlined in 1.
118 3) Ethics and law - A survey of workshop participants highlighted a need for guidance on
119   sharing data and complying with important ethical and legal requirements (Simpson et al.
120   2021). This category will include information on permits and permission required to obtain
121   samples and release data. Collating and publishing this information will firstly provide
122   examples for how previous projects have adhered to legal requirements/ethical principles and

---

[1] Currently, the protocol categories focus on genomics and transcriptomics but we expect this list to expand with further input from the broader omics community, particularly in areas such as proteomics and metabolomics.

123   secondly stimulate discussion on how to facilitate adherence to these requirements and
124   principles, perhaps through checklists, templates, or training materials.
125   4) Data management - The data management plan (DMP) is designed to support all the
126   downstream steps according to the ethics, legalities and societal needs identified in 1) - 3),
127   while making sure that the (meta)data flows to the right stakeholders in society that we need
128   to interface with. DMPs should be drafted prior to data collection and referred to throughout
129   the workflow to ensure that quality assurance and quality checks take place, and that detailed
130   information on (meta)data requirements for both short and long-term (meta)data storage is
131   given. There is a growing body of tools and best practices surrounding DMPs, including
132   principles for making them more machine-actionable, that should be leveraged in omic
133   protocols and associated infrastructure (see Miksa et al., 2019). Publishing documentation on
134   omics specific DMPs will increase transparency for funders by providing direct links to the
135   protocols they refer to. Furthermore, collating examples of omics specific DMPs will provide
136   insight into what the community needs from omics specific data management tools.

137   In Figure 1B, we give an example of a DNA metabarcoding workflow, where the colour of each step
138   corresponds to a methodology category in Figure 1B. Protocols uploaded to OBPS can be assigned
139   (tagged) to the relevant omics categories. The granularity of protocols uploaded to the OBPS may
140   include individual uploads for sub-stages (i.e. Tagging/Enrichment within 4. Omics sequencing
141   procedures), or single documents spanning multiple methodology categories (i.e. 3. Sample
142   extraction & purification, through to 5. Bioinformatics). To accommodate these levels of granularity,
143   each upload could be tagged with single or multiple methodology category and linked to those
144   protocols pre- and succeeding it. The granular use of methodology categories will increase
145   modularity within the omics workflow and facilitate the mixing and matching of methods from
146   various projects.

147   The interplay between the activities within and across the steps within a workflow - and how they
148   bring value to the community and society - is complex and beyond the scope of this article; however,
149   we have provided an initial perspective on this using the Porter's value chain approach (Porter, 1985;
150   Supplementary Figure 1).

A

1. Society	Documenting societal inputs and outputs (e.g. mission statements (linked to broadly accepted societal priorities e.g. SDGs EOVs, EBVs. etc), stakeholder consultation outputs, associated publications/media, links to relevant data repositories, educational materials associated with the workflow).
2. Design and logistics	Documenting plans for the experimental/observational design (e.g. cruise schedules/reports, deployment regimes, statistical replication, logistics, costing/resources/material lists)
3. Ethics and law	Documenting permits, permissions and guidelines for obtaining samples and release data. (e.g., Internationally Recognized Certificate of Compliance (IRCC) code for sampling, or relevant Due Diligence documentation if an IRCC code is not required, for countries that have signed and ratified the Nagoya protocol (Nagoya Protocol, 2010)) checklists showing compliance to ethical principle (e.g. OCAP) for ethical sampling on indigenous lands, etc
4. Data management	Protocols for data and quality management. This will include data management plans with quality assurance and quality check measures to be undertaken throughout the workflow, as well as checklists and guidance documents for the compliance with FAIR/CARE data standards; and links to all data repositories to which the omics study (meta)data should be uploaded.
5. Sampling Collection Procedure	Protocols for the physical collection of samples to be used in omics research. Including both environmental samples (e.g. water, sediments, biofilms) and organismal samples (e.g. gut contents, bulk plankton, sponge).
6. Contextual Metadata Analytical Procedures	Protocols used for the measurement of additional parameters (e.g. temperature, salinity, pH, chlorophyll etc.)
7. Sample Extraction & Purification	Protocols for the extraction and purification of biological molecules (e.g. DNA, RNA, Proteins) from environmental and/or organismal samples.
8. Omics Sequencing Procedures	Protocols for sequencing nucleic acids. This stage may contain sub-stages that occur during the sample extraction and purification stage, if so, links between the two stages should be provided.
9. Bioinformatics	Bioinformatics pipeline for processing sequence data. This can include links to github, docker etc.
10. Quantitative amplification	Protocols for the quantitative amplification of DNA and RNA. Including quantitative PCR (qPCR), reverse transcriptase quantitative PCR (rt-qPCR), nucleic acid sequence-based amplification (NASBA), Recombinase Polymerase Amplification (RPA) etc.
11. Data Analysis	Protocols for analysis of data. This will include scripts for statistical tests, modelling and data visualisations.
12. Sample archiving/biobanking	Protocols for archiving physical samples at all stages of laboratory processing (e.g. whole environmental/organismal samples, DNA/RNA/protein extracts and/or amplicons)

B

151
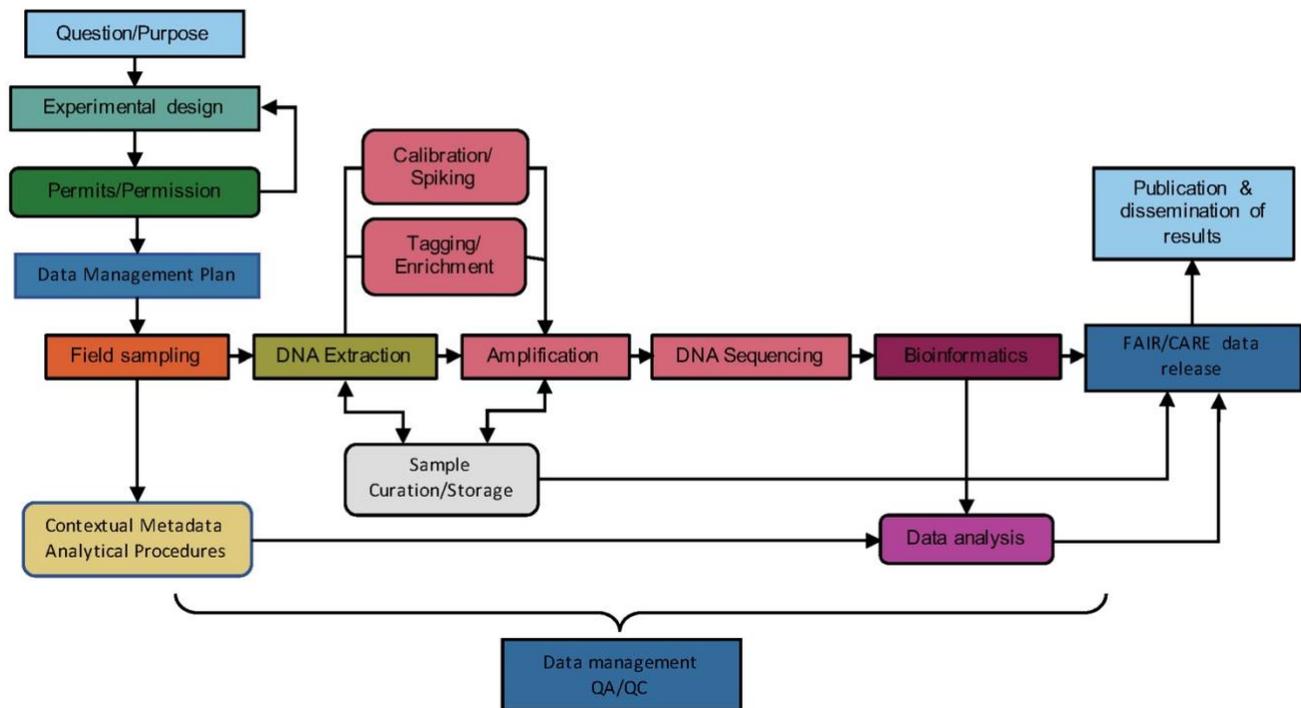
Figure 1. A. Proposed methodology categories to enhance exchange of ocean omics analysis knowhow. Protocols, guidelines, and other methodologies in some of these categories (such as Sample archiving/biobanking, Data Management, and Society) are cross-cutting and may apply at multiple points in the workflow. B. Example workflow for a DNA metabarcoding project. Colours correspond

155 to the methodology categories outlined in A and arrows indicate the order of the workflow. Square boxes show essential steps in a
156 metabarcoding workflow, whereas rounded boxes indicate non-essential steps. Data management and QA/QC are required throughout
157 the entire workflow.

3   Essential Metadata for Omics Protocols

159 The targeted discovery and reuse of protocols can be improved if protocols are effectively described
160 using standardised metadata terms on upload to OBPS and other platforms. Terms and checklists to
161 standardise metadata about primary sequence or biodiversity data already exist (GSC's Minimum
162 Information about any (x) Sequence checklist (MIxS; Yilmaz et al., 2011) and TDWG's Darwin
163 Core standard (Wieczorek et al., 2012)); however, no such standards have thus far been published for
164 metadata about omics protocols.

165 Here we present initial suggestions for the Minimum Information for an Omic Protocol (MIOP), a set
166 of ten metadata categories which could correspond to ten key decision tree questions asked to
167 identify the relevant protocol for any project. The ten MIOP categories (Table 1) consist of five novel
168 categories (methodology category, purpose, resources, analysis, target) and five categories already
169 used in the GSC's MIxS (project, geographic location, broad-scale environmental context, local
170 environmental context, and environmental medium). Each category is linked to a set of predefined
171 keywords (metadata terms) from existing vocabularies or ontologies; except for the 'project'
172 category, which contains project names, affiliations, and contact details and the 'methodology
173 category' outlined in section 2 (Figure 1A). Omics users would then select the most appropriate
174 keywords for each category, assigning the terms as metadata for the protocol. This will improve the
175 FAIRness of our protocol data, by allowing consequent users to search the protocol database using
176 the same set of keywords; thereby, limiting the proliferation of descriptive keywords (e.g., mapping
177 synonyms) and increasing the findability of protocols.

178 **Table 1.** Description of keyword categories for protocol metadata and the terminologies (controlled vocabularies, thesauri, ontologies)
179 containing the relevant keywords. Terms would be added at upload and additional metadata would accumulate as the protocols are used
180 in different settings (e.g., Geographic Locations; see 4.2 learning from failed practices). EFO: Environmental Factor Ontology, OBI:
181 Ontology for Biomedical Investigations, NCIT: NCI Thesaurus, GAZ: Gazetteer, ENVO: Environment Ontology, UBERON: Uber-
182 anatomy ontology, NCBITaxon: NCBITaxon ontology.

Categories	Terminology/Ontology	Description
Methodology category	Methodology category (see Figure 1A.)	Methodology category which the uploaded protocol belongs to. This links to the associated methodology categories which precede and succeed it in the workflow, to facilitate the linking of protocols into entire workflows, while keeping granularity and flexibility. This will enable the mixing and matching of protocol modules from various uploaded workflows.
Project	N/A	Details about the project (e.g. Name, Affiliation, website). May also includes a field for tagging any projects that protocols are compliant with (e.g. Earth Microbiome Project/TARA Oceans). Once submitted the relevant PI may be notified and could choose to endorse or reject the protocol as compliant with their project.
Purpose	EFO, OBI	Terms to describe the purpose of the omics research. (e.g. time series design [OBI:0500020] or taxonomic diversity assessment by targeted gene survey [OBI:0001960])
Resources	EFO, NCIT	Terms to identify the key resources needed to complete the protocol (e.g, Illumina MiSeq [EFO:0004205], centrifuge [OBI:0400106])
Analyses	EFO, OBI, NCIT	Terms to describe the types of analyses used in the protocol (e.g. amplicon sequencing assay [OBI:0002767] or polymerase chain reaction [OBI:0002692])
Geographic Location	GAZ	Geographic location/s in which the protocol has been used (e.g. Hawaii Ocean Time-series Site [GAZ:00187530], Western English Channel Sampling Stations [GAZ:00187525])

Broad-scale environmental context (former Biome)	ENVO	Biome/s in which the protocol was successfully used (e.g. oceanic epipelagic zone biome [ENVO:01000033])
Local environmental context (former Feature)	ENVO, UBERON	Environmental feature/s targeted using the protocol (e.g. seasonal thermocline [ENVO:01000107])
Environmental medium (former Material)	MIxS environmental packages; ENVO	Identify the environmental or organismal material from which the biological molecule (e.g. DNA/RNA/Protein) was extracted (e.g. ocean water [ENVO:00002151])
Target	NCIT, NCBITaxon, EFO	Identify the target taxa, gene and/or molecule for the protocol (e.g. Polaribacter [NCBITaxon:1642819], 16S Mitochondrial Ribosomal RNA [NCIT:C131261]).

183

4   Discussion

185 OBPS provides a neutral, global public repository for ocean community practices. It is a stable and
186 persistent foundation that can host protocols themselves, or link to other protocol tools and
187 functionalities that can (and should) continue to be developed by other organizations including the
188 private sector. The primary function of Omics OBPS would be to publish and archive omics
189 protocols to enhance their global visibility and discoverability, and provide stable links to the entire
190 workflow of protocols. Expanding and improving the functionality of the OBPS for omics protocols
191 will help the community mature by providing a structured system in which context-based best
192 practices can be discovered and identified. A transparent and structured process for handling our
193 omics protocols will be an essential step towards operationalizing omics observing.

194 Increasing protocol transparency, through detailed publication on OBPS, also means that simple cited
195 protocol strings can become a core component of methods sections in publications. Those strings can
196 then be harvested by machines to generate a graph of "what came before" and "what came after".
197 When used with the decision tree recommendations this process could point out the most recent
198 protocol development to users and would essentially provide the decision-tree resource we are aiming
199 for. Such an approach enables 'practices' (which might be defined as 'protocol strings') to emerge
200 from how protocols are actually being used in the community. Assessment of which of these
201 practices represent a "best" practice in a given context is a distinct challenge, but not a unique one in
202 knowledge sectors. Peer endorsement and citation metrics are two commonly employed ranking
203 mechanisms that could also be applied here.

4.1   Learning from community preferences

205 Community-use metrics offer a way to capture the community's preference for certain protocols. We
206 suggest that metrics such as times cited, user upvotes, and number of associated data records all be
207 recorded and used to rank lists of relevant protocols. Combined with the MIOP-based grouping into
208 methodology categories, this process will help accelerate the identification of potential best practices
209 within each category. Narrowing down the list of relevant protocols will additionally provide the
210 basis for more targeted and rigorous scientific comparisons between multiple potential best practices
211 for a given scientific endeavour. Outputs of such comparisons may offer further information about
212 the superiority of certain protocols, and could be considered in addition to the more general

213   community-use metrics[2]. Furthermore, focusing on these community driven best practices will help
214   to reveal protocols that are effective and convenient for a broad range of research facilities. This in
215   turn can reduce literature biases toward novel state of the art practices, which may not be feasible for
216   mainstream use.

4.2   Learning from failed practices

218   During the initial workshop, participants outlined a desire for a best practice system to include 'failed
219   practices' and flag when a protocol may limit or eliminate a range of downstream applications. While
220   this type of functionality would not be immediately addressed by implementing MIOP metadata,
221   there would be potential for users to provide feedback for protocols using MIOP metadata and
222   boolean operators. For example, if a protocol, originally designed for seawater, was used with
223   freshwater samples, the user could upload additional MIOP metadata using 'AND freshwater' if the
224   protocol was successful or 'NOT freshwater' if unsuccessful. Thereby, broadening the findability of
225   successful protocols and documenting potential limitations to be aware of. Documenting these failed
226   attempts has the potential to save both time and resources.

4.3   Promoting collaborative omic networks

228   MIOP may additionally promote collaboration between groups. For example, the 'Project' category
229   is an administrative metadata field that will describe the project (study or program) for which the
230   protocol was developed, including contact details and affiliated institution. To create links between
231   similar projects and facilitate collaboration, it would be possible to introduce an option to tag a
232   protocol as compliant with pre-existing projects. In such cases, a notification could be sent to the PI
233   of the lead project, allowing them to add or reject the protocol to their list of compliant protocols.
234   Protocols linked this way could form overarching protocol concepts, which may contain a variety of
235   versions and accepted, cross-comparable protocols that include minor adaptations to make them
236   suitable in different circumstances.

237   An endorsement process for a global observation network has already been developed by Global
238   Ocean Observing System (GOOS) in cooperation with OBPS, to encourage standardised methods for
239   global observations and for reporting on GOOS' Essential Ocean Variables (EOVs) (Miloslavich et
240   al., 2019; Hermes, 2020). To gain this endorsement, protocols will have to undergo a rigorous
241   community review process that will be strengthened if there is a large source of omics protocols to
242   compare with on the OBPS. Standardised practices and official endorsements are likely to become
243   increasingly valuable as countries begin to use legislation to make biodiversity targets legally
244   binding. Any omic method used to measure biodiversity impacts will need to undergo legal scrutiny
245   if it is used as evidence of a country/organisation meeting or failing to meet biodiversity targets.
246   Therefore, protocols officially endorsed through international programmes, such as GOOS, are likely
247   to hold more sway legally. Broad participation from the omics community in open sharing and
248   reviewing of protocols on the OBPS will help to ensure that community endorsed best practices are
249   representative of the wider community needs and not only focussed on expensive state of the art
250   methodologies.

4.4   Machine readability

8

[2] In certain cases (e.g., for contributing to a standardised global sampling scheme) it may not be about which method is "best", but about which method delivers reliable results while being applicable throughout all regions of the ocean and inclusive of lower capacity research activities.

252 Machine readable tracking of protocol versions presents an opportunity to visually map the
253 progression of protocols by linking all versions to a "concept", as implemented in Zenodo and
254 GitHub. Like software, omic protocols may be updated, corrected, and improved necessitating forms
255 of version control and tracking, such as the use of semantic versioning (Preston-Werner, 2021;
256 Hörstmann et al. 2020). Implementing this would help to increase recognition for the
257 scientists/technicians/students involved in protocol development through citable documentation of
258 their contributions.

259 Machine-readable and machine-actionable protocols are becoming more important as autonomous
260 technologies evolve. Devices such as the Environmental Sample Processor (ESP) and the Robotic
261 Cartridge Sampling Instrument (RoCSI) are currently being used and developed for autonomous
262 collection, preservation, and *in-situ* analysis of omics samples (Yamahara et al., 2019; National
263 Oceanography Centre, 2021). Eventually, smart sensing platforms using these technologies will be
264 able integrate data from various sensors and satellites to implement adaptive sampling regimes or
265 extraction protocols based on real-time environmental observations (Witt et al. 2020). To reach this
266 goal a variety of protocols will need to be translated into a machine actionable format using common
267 workflow language. A systematic review of protocols will help to devise such machine actionable
268 formats and protocol templates may help to bridge the gap between lab-based protocol development
269 and *in-situ* autonomous use.

270 **5   Conclusion**

271 Multiple groups within the omics community are actively developing best practices for their field. To
272 ensure that all these efforts are effectively utilized, a concerted and community wide effort will be
273 needed to gather and organise these practices. By harnessing the OBPS infrastructure and further
274 developing the MIOP metadata we can: 1) allow protocols to be searched for within a decision tree
275 framework; 2) establish a system that encourages the systematic review of protocols; and 3) reveal
276 community preferences through the accumulation of community use data. Taking these steps toward
277 a structured and global public repository of omics protocols will increase transparency and streamline
278 biomolecular ocean observing research to foster the collaborative networks needed to achieve global
279 scale biodiversity observations.

280 **6   Conflict of Interest**

281 The authors declare that the research was conducted in the absence of any commercial or financial
282 relationships that could be construed as a potential conflict of interest.

283 **7   Author Contributions**

284 All authors contributed to the discussion and wrote the manuscript. AW, CM and RS constructed the
285 main text figure with input from all authors. RM, PLB, and RS developed the supplementary figure.

286 **8   Funding**

9 Acknowledgments

10 References

316 Aylagas, E., Borja, A., Pochon, X., Zaiko, A., Keeley, N., Bruce, K., Hong, P., Ruiz, G., Stein, E.,
317 and Theroux, S. et al. (2020). Translational Molecular Ecology in practice: Linking DNA-based
318 methods to actionable marine environmental management. Science of The Total Environment 744,
319 140780. doi:10.1016/j.scitotenv.2020.140780.

320 Berry, O., Jarman, S., Bissett, A., Hope, M., Paeper, C., Bessey, C., Schwartz, M., Hale, J., and
321 Bunce, M. (2020). Making environmental DNA (eDNA) biodiversity records globally accessible.
322 Environmental DNA 3, 699-705. doi:10.1002/edn3.173.

323 Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P., Reith, F., Dennis, P., Breed, M., Brown, B., Brown,
324 M., and Brugger, J. et al. (2016). Introducing BASE: the Biomes of Australian Soil Environments
325 soil microbial diversity database. GigaScience 5. doi:10.1186/s13742-016-0126-5.

326 Brown, M., van de Kamp, J., Ostrowski, M., Seymour, J., Ingleton, T., Messer, L., Jeffries, T.,
327 Siboni, N., Laverock, B., and Bibiloni-Isaksson, J. et al. (2018). Systematic, continental scale
328 temporal monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity
329 Initiative. Scientific Data 5. doi:10.1038/sdata.2018.130.

330 Buttigieg, P. , Appeltans, W. , Canonico, G. , Chavez, F. P. , Deck, J. , Glöckner, F. O. , Goodwin, K.
331 D. , Hunter, M. , Kotoulas, G. , Janssen, F. , Lanfear, J. , Meyer, C. P. , Muller-Karger, F. , Obst, M. ,
332 Pitz, K. J. , Provoost, P. , Thompson, L. R. , Walls, R. and Davies, N. (2019). Building a strategy
333 towards an Omic Biodiversity Observation Network (Omic BON) , OceanObs'19 conference,
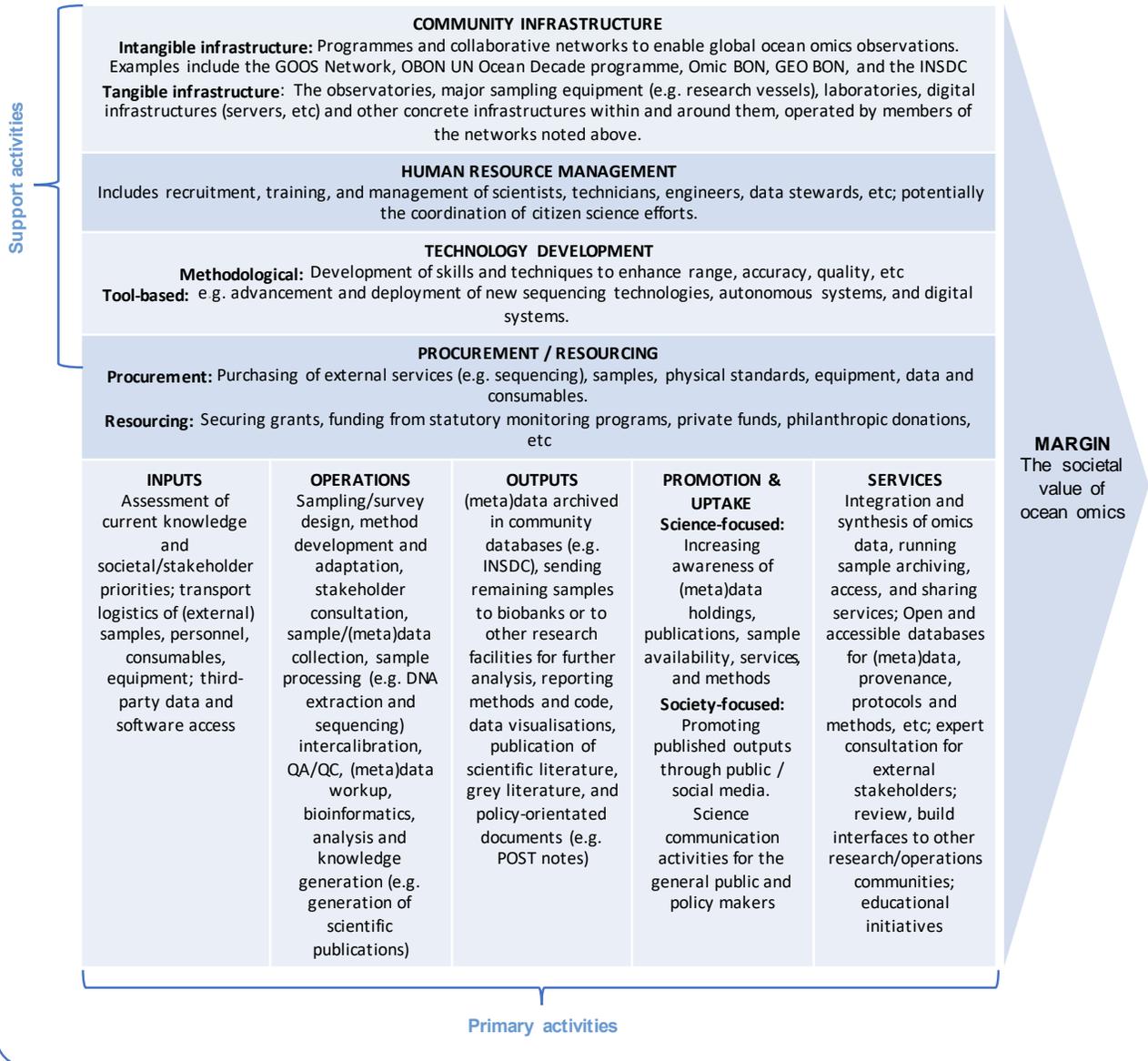334 Honolulu, Hawaii, USA, 16 September 2019 - 20 September 2019.

335   Carroll, S., Garba, I., Figueroa-Rodríguez, O., Holbrook, J., Lovett, R., Materechera, S., Parsons, M.,
336   Raseroka, K., Rodriguez-Lonebear, D., and Rowe, R. et al. (2020). The CARE Principles for
337   Indigenous Data Governance. Data Science Journal 19. doi:10.5334/dsj-2020-043.

338   Davies, N., Field, D., Amaral-Zettler, L., Clark, M., Deck, J., Drummond, A., Faith, D., Geller, J.,
339   Gilbert, J., and Glöckner, F. et al. (2014). The founding charter of the Genomic Observatories
340   Network. GigaScience 3. doi:10.1186/2047-217x-3-2.

341   Davies, N., Meyer, C., Gilbert, J., Amaral-Zettler, L., Deck, J., Bicak, M., Rocca-Serra, P., Assunta-
342   Sansone, S., Willis, K., and Field, D. (2012). A call for an international network of genomic
343   observatories (GOs). GigaScience 1. doi:10.1186/2047-217x-1-5.

344   Dickie, I., Boyer, S., Buckley, H., Duncan, R., Gardner, P., Hogg, I., Holdaway, R., Lear, G.,
345   Makiola, A., and Morales, S. et al. (2018). Towards robust and repeatable sampling methods in
346   eDNA -based studies. Molecular Ecology Resources 18, 940-952. doi:10.1111/1755-0998.12907.

347   Gilbert, J., Jansson, J., and Knight, R. (2014). The Earth Microbiome project: successes and
348   aspirations. BMC Biology 12. doi:10.1186/s12915-014-0069-1.

349   Hermes, J. (ed.) (2020) GOOS Best Practices Endorsement Process. Version 1. Paris, France, Global
350   Ocean Observing System, 7pp. DOI: http://dx.doi.org/10.25607/OBP-926

351   Hörstmann, C., Buttigieg, P., Simpson, P., Pearlman, J., and Waite, A. (2021). Perspectives on
352   Documenting Methods to Create Ocean Best Practices. Frontiers in Marine Science 7.
353   doi:10.3389/fmars.2020.556234.

354   Hörstmann, C., Buttigieg, P.L., Simpson, P., Pearlman, J., Karstensen, J. and Waite, A.M. (2020)
355   Towards a Best Practice for Developing Best Practices in Ocean Observation (BP4BP): Supporting
356   Methodological Evolution through Actionable Documentation. Paris, France, UNESCO, 33pp.
357   (Intergovernmental Oceanographic Commission Manuals and Guides No. 84). (IOC/2020/MG/84).
358   DOI: http://dx.doi.org/10.25607/OBP-781

359   Makiola, A., Compson, Z., Baird, D., Barnes, M., Boerlijst, S., Bouchez, A., Brennan, G., Bush, A.,
360   Canard, E., and Cordier, T. et al. (2020). Key Questions for Next-Generation Biomonitoring.
361   Frontiers in Environmental Science 7. doi:10.3389/fenvs.2019.00197.

362   Miksa, T., Simms, S., Mietchen, D., and Jones, S. (2019). Ten principles for machine-actionable data
363   management plans. PLOS Computational Biology 15, e1006750. doi:10.1371/journal.pcbi.1006750.

364   Miloslavich, P., Bax, N., Simmons, S., Klein, E., Appeltans, W., Aburto-Oropeza, O., Andersen
365   Garcia, M., Batten, S., Benedetti-Cecchi, L., and Checkley, D. et al. (2018). Essential ocean variables
366   for global sustained observations of biodiversity and ecosystem changes. Global Change Biology 24,
367   2416-2433. doi:10.1111/gcb.14108.

368   Mitchell, A., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M., Kale,
369   V., Potter, S., and Richardson, L. et al. (2019). MGnify: the microbiome analysis resource in 2020.
370   Nucleic Acids Research. doi:10.1093/nar/gkz1035.

371   Nagoya Protocol. (2010). Nagoya Protocol on Access to Genetic Resources and the Fair and
372   Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological
373   Diversity: (Chapter XXVII: Environment). United Nations Treaty Collection, Nagoya.

374   National Oceanography Centre (2021). Auto-sampler device and method of autosampling.

375   Open Data Watch (2021). The Data Value Chain: Moving from Production to Impact. Open Data
376   Watch. Available at: https://opendatawatch.com/publications/the-data-value-chain-moving-from-
377   production-to-impact/ [Accessed August 4, 2021].

378   Porter, M. (1985). Competitive Advantage: Creating and Sustaining Superior Performance. New
379   York: Free Press.

380   Preston-Werner, T. (2021). Semantic Versioning 2.0.0. Semantic Versioning. Available at:
381   https://semver.org/ [Accessed July 28, 2021].

382   Simpson, P., Pearlman, F. and Pearlman, J. (eds) (2021) Evolving and Sustaining Ocean Best
383   Practices Workshop IV, 18; 21-25 & 30 Sep 2020 [Online]: Proceedings, Volumes 1 & 2. Paris,
384   France, UNESCO, 66pp. & 115pp. (IOC Workshop Report No. 294, Vols. 1 & 2). DOI:
385   https://doi.org/10.25607/OBP-1036

386   Sunagawa, S., Acinas, S., Bork, P., Bowler, C., Eveillard, D., Gorsky, G., Guidi, L., Iudicone, D.,
387   Karsenti, E., and Lombard, F. et al. (2020). Tara Oceans: towards global ocean ecosystems biology.
388   Nature Reviews Microbiology 18, 428-445. doi:10.1038/s41579-020-0364-5.

389   Thompson, L., Sanders, J., McDonald, D., Amir, A., Ladau, J., Locey, K., Prill, R., Tripathi, A.,
390   Gibbons, S., and Ackermann, G. et al. (2017). A communal catalogue reveals Earth's multiscale
391   microbial diversity. Nature 551, 457-463. doi:10.1038/nature24621.

392   van de Kamp, J., Brown, M., Ostrowski, M., Seymour, J., Smith, M., Raes, E., Karsh, K., Holmes,
393   B., Bodrossy, L. and Bissett, A. (2019). The Australian Microbiome Initiative-towards microbial
394   omics at the continental scale. In Geophysical Research Abstracts (Vol. 21).

395   Wetterstrand, K. (2020). The Cost of Sequencing a Human Genome [WWW Document]. URL
396   https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost (accessed
397   3.28.21).

398   Whitt, C., Pearlman, J., Polagye, B., Caimi, F., Muller-Karger, F., Copping, A., Spence, H.,
399   Madhusudhana, S., Kirkwood, W., and Grosjean, L. et al. (2020). Future Vision for Autonomous
400   Ocean Observations. Frontiers in Marine Science 7. doi:10.3389/fmars.2020.00697.

401   Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., and
402   Vieglais, D. (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard.
403   PLoS ONE 7, e29715. doi:10.1371/journal.pone.0029715.

404   Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N.,
405   Boiten, J., da Silva Santos, L., and Bourne, P. et al. (2016). The FAIR Guiding Principles for
406   scientific data management and stewardship. Scientific Data 3. doi:10.1038/sdata.2016.18.

407   Yamahara, K., Preston, C., Birch, J., Walz, K., Marin, R., Jensen, S., Pargett, D., Roman, B., Ussler,
408   W., and Zhang, Y. et al. (2019). In situ Autonomous Acquisition and Preservation of Marine
409   Environmental DNA Using an Autonomous Underwater Vehicle. Frontiers in Marine Science 6.
410   doi:10.3389/fmars.2019.00373.

411   Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J., Amaral-Zettler, L., Gilbert, J., Karsch-
412   Mizrachi, I., Johnston, A., and Cochrane, G. et al. (2011). Minimum information about a marker gene
413   sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications.
414   Nature Biotechnology 29, 415-420. doi:10.1038/nbt.1823.

415   **11   Supplementary Material**

Ocean Omics Value Chain

Applying Porter's value chain concept to ocean omics research can help to systematically identify activities that will increase the value of ocean omics research to society at large: A well-functioning ocean omics value chain will generate more downstream value than the costs needed to operate it. This is the basis of justifying increases or decreases in investment in any activity along or across the chain. We believe such mechanisms will help the omics community coherently reflect on the activities needed to advance our methods while ensuring a healthy cost/benefit ratio, especially as we interface with other ocean communities and their own value chains. For example, when considering the uptake of new technologies (e.g. new sequencing technologies), value chain analysis helps to consider whether uptake is worth the disruption to the continuity of long-term ocean omics observatories. The process helps to evaluate what value is compromised (i.e. Are the time series broken? Are the POST notes valid anymore?) and what needs co-investment to mitigate it (e.g. intercalibration across space and time). Essentially answering "is it worth it?" at scale, or should it be a more isolated test by a smaller community (e.g. a scientific team/pilot study).

Support activities

COMMUNITY INFRASTRUCTURE
Intangible infrastructure: Programmes and collaborative networks to enable global ocean omics observations. Examples include the GOOS Network, OBON UN Ocean Decade programme, Omic BON, GEO BON, and the INSDC
Tangible infrastructure: The observatories, major sampling equipment (e.g. research vessels), laboratories, digital infrastructures (servers, etc) and other concrete infrastructures within and around them, operated by members of the networks noted above.

HUMAN RESOURCE MANAGEMENT
Includes recruitment, training, and management of scientists, technicians, engineers, data stewards, etc; potentially the coordination of citizen science efforts.

TECHNOLOGY DEVELOPMENT
Methodological: Development of skills and techniques to enhance range, accuracy, quality, etc
Tool-based: e.g. advancement and deployment of new sequencing technologies, autonomous systems, and digital systems.

PROCUREMENT / RESOURCING
Procurement: Purchasing of external services (e.g. sequencing), samples, physical standards, equipment, data and consumables.
Resourcing: Securing grants, funding from statutory monitoring programs, private funds, philanthropic donations, etc

MARGIN The societal value of ocean omics

INPUTS	OPERATIONS	OUTPUTS	PROMOTION & UPTAKE	SERVICES
Assessment of current knowledge and societal/stakeholder priorities; transport logistics of (external) samples, personnel, consumables, equipment; third-party data and software access	Sampling/survey design, method development and adaptation, stakeholder consultation, sample/(meta)data collection, sample processing (e.g. DNA extraction and sequencing) intercalibration, QA/QC, (meta)data workup, bioinformatics, analysis and knowledge generation (e.g. generation of scientific publications)	(meta)data archived in community databases (e.g. INSDC), sending remaining samples to biobanks or to other research facilities for further analysis, reporting methods and code, data visualisations, publication of scientific literature, grey literature, and policy-orientated documents (e.g. POST notes)	**Science-focused:** Increasing awareness of (meta)data holdings, publications, sample availability, services, and methods **Society-focused:** Promoting published outputs through public / social media. Science communication activities for the general public and policy makers	Integration and synthesis of omics data, running sample archiving, access, and sharing services; Open and accessible databases for (meta)data, provenance, protocols and methods, etc; expert consultation for external stakeholders; review, build interfaces to other research/operations communities; educational initiatives

Primary activities

416

Supplementary Figure 1. We frame part of our perspective in an adaptation of Porter's (1985) value chain, where a product passes through all portions of the chain, gaining value from each activity. The language used in Porter's value chain relates to commercial business activities, this adaptation gives examples for the types of omic research activities that could apply to each category. Category headings have been adapted to fit omics operations and research, as follows: Firm infrastructure → Community infrastructure, Procurement → Procurements / Resourcing, Inbound logistics → Inputs, Outbound logistics → Outputs, Marketing & Sales → Promotion & Uptake (note that original headings would be appropriate for omics-focused businesses). It should also be noted that

423   value within omics operations and research does not only refer to monetary transactions (for example, procurement may be facilitated
424   by credit on scientific publications).

425