We thank the reviewers for their constructive comments and suggestions. We have thoroughly addressed all the points raised and implemented the suggested change. Importantly, following the insightful suggestions made by the reviewers we have:

- Clarified throughout the text that the ecosystem elements on which we chose to elaborate are merely representative examples.
- Changed the approach for interactive representation of the ocean data ecosystem, using a comprehensive knowledge graph instead of an ontology.
- Added information that was lacking in the previous version of the manuscript.
- Elaborated our vision for the future of the ocean data ecosystem.

Our detailed point-by-point response to the reviewers' comments and suggestions is found below. All page/line/reference/figure numbers refer to the clean version of the revised manuscript. Reviewers' comments are in regular text, **our responses are in bold** and ***new text from the manuscript is in bold italics***.

Reviewer #1

This manuscript is much improved with a slightly clearer structure and the addition on online tool is fantastic for a resource that can be kept up-to-date to ensure relevance of this manuscript and tool into the future. However, my fundamental concerns regarding the length and breadth of the paper remain.

The large and detailed list of "examples" in this manuscript mean the overall aim of the publication is lost and the map of the data ecosystem isn't clear. Whilst the authors have made it clearer these are "examples", the detail on these examples makes it appear more like an exhaustive list, which it isn't. Further, these examples are included in the online tool but no where does it make this clear these are "examples only" therefore again it appears to be an exhaustive list, which it isn't.

Answer: We thank the reviewer for the thorough reading and the constructive comments. We agree that there is somewhat a discrepancy between the length of the paper, and the fact that it does not give an exhaustive description of the ocean data ecosystem and all its components. We do think however that expanding the discussion on a limited number of selected examples, <u>while clarifying the fact that they are such</u>, does provide valuable information that will help the reader navigate his way in the world of ocean data, thus facilitating data-driven ocean research. Below we explicitly address the reviewers suggestions and comments.

My suggestions to make this manuscript clearer and more impactful:

1) Ensure it is very obvious that the description of data sources, tools and frameworks are examples only and not an exhaustive list.

Answer: we agree with the reviewer that it is essential to clarify that the different ecosystem elements discussed in detail are only, and do not constitute an exhaustive

list. Accordingly, in the current version of the manuscript this is mentioned explicitly a number of times:

Line 182: "Rather than presenting an exhaustive and potentially quickly outdated inventory…".

Line 190: "The methodology we used is based on a thorough literature and website review of over 90 scientific articles and over 100 websites. Articles and examples selected to illustrate the elements…".

Line 1138: "To maintain long-term relevance, the ecosystem model presented is aimed to be used as a tool in further characterization of the ocean data ecosystem. The examples given are not exhaustive…".

To further clarify this point, in the revised manuscript we have added the following text to section 3.4. (lines 510-514):
"The description of data sources, tools, and frameworks is not exhaustive. Due to the very broad nature of the ocean data ecosystem, rather than covering the large number of elements it contains, we give several representative examples. The readers may use these examples, presented in detail, to broaden their knowledge of the data ecosystem, and may perform further review of other data sources within their respective fields. We used a framework for analyzing the data sources, including key characteristics composing a data source, namely: Organization details such as number of partnering organizations, oceanographic domain, geographic region, number of data sets, data catalogue and product offering, main uses by the specific user community and interoperability strategy. Other characteristics may be selected depending on the reader's focus and interests."

2) Ensure all figures state that the list are examples (e.g., Figure 4 caption to include wording similar to Figure 11).

Answer: we have made sure that the relevant figure captions clearly state the exemplary nature of the figures, as follows:

"Figure 2. A diagram summarizing the types of societal elements and the representative examples discussed in Sec. 3.2"

"Figure 3. A diagram summarizing the types of interoperability tools and the representative examples discussed in Sec. 3.3."

"Figure 4. A diagram summarizing the examples for ocean data sources discussed in section 3.4. The exemplified data sources are organized according to their infrastructural approach (namely federated, distributed and centralized)."

"Figure 11. A diagram summarizing the examples for emerging solutions discussed in Sec. 3.5, along with brief descriptions of the approach they represent."

3) Significantly strip back example sections e.g., 3.4.1-3.4.10 to be summary paragraph or two in section 3.4 and instead include the detail in a supplementary section if the authors wish to still make this available

Answer: **Despite its length, we do not think that stripping back the example sections will improve the paper. The selected examples are sufficiently representative to enable readers to develop a product-oriented perspective, and provide valuable information for readers who wish to understand and characterize key products in this domain. While a different selection of examples may have been made, our evaluation suggests that similar conclusions would emerge. Importantly, the examples reveal a connecting thread across the different systems: how information is organized into product catalogs, how dataset searches are supported, how information remains distributed across disciplines and platforms, and what gaps still persist. Moreover, the organization of the selected examples in consequent subsections allows readers who are more familiar with the domain to skip their description without influencing the paper's consistency and readability. We therefore believe that retaining the examples in the main text adds significant value.**

I also think some of the explanatory text in the response to reviewers could be useful to include in the manuscript e.g., emerging solutions explanatory text.

Answer: **we have added explanatory text as suggested in section 3.4. (see details in our response above).**

As the manuscript stands, whilst it is factually correct and could be published as-is, I would worry on the impact the paper will have give its length and breadth. A stripping back of the paper and focus just on the key message of the easy to navigate map of the ocean data ecosystem would greatly improve the usability and impact of this tool, which is a very useful contribution to the ocean data community

Answer: **We thank the reviewer for the thoughtful comment regarding the length and breadth of the manuscript. We acknowledge that the paper is relatively long. However, it has been deliberately organized into clearly delineated sections and subsections so that readers can readily focus on areas of specific interest or skip sections that are less relevant to them while still following the overall narrative. We believe this structure preserves readability and flow, while allowing the manuscript to serve both readers seeking the key message (i.e. a map of the ocean data ecosystem) and those wishing to explore additional context and details.**

Thank you for producing a revised version. The new version addresses some of my feedback on specific points raised it also introduces or highlights new issues to be addressed before I can recommend the manuscript for publication.

Answer: We thank the reviewer for the thorough evaluation process, and appreciate the thoughtful and knowledgeable comments. As detailed below we have carefully addressed the reviewer's concern, thus substantially improving the quality of the paper.

On second review and after revisions the context behind and underpinning many of the entities described in the paper seems ambiguous or confused. Three examples that stand out are:

1. Of the principles presented (FAIR, TRUST, CARE); FAIR is given precedence with TRUST and CARE also existing. The reality is that each is designed to serve a different purpose and applicable in different contexts e.g. FAIR is about data accessibility and utility, TRUST is about the sustainability and practices used in data curation with the connection between CoreTrustSeal being an accreditation for TRUST missing later in the manuscript, and CARE is critical when handling data when indigenous communities are an interest-holder. Thus, each has a place and a role and value in its own right in their contexts.

Answer: we thank the reviewer for drawing our attention to this important issue. To give a more representative picture of the three principles and their roles, the relevant paragraph was modified as follows (lines 232-247):

"By and large, ocean data management is guided by the FAIR, TRUST and CARE principles, each is designed to serve a different purpose and applicable in different contexts. The FAIR principles are mainly concerned with scientific data management and stewardship and are meant to ensure reusability of data, with an emphasis on enabling the automation of data findability and usability (Tanhua et al., 2019b; "The FAIR Data Principles – FORCE11," n.d.; Wilkinson et al., 2016). The TRUST principles of data repositories (Lin et al., 2020), deals with data curation, providing guidance to demonstrate transparency, responsibility, user focus, sustainability, and technology. The CARE principles ("CARE Principles — Global Indigenous Data Alliance," n.d.) were defined for ensuring proper handling of data associated with indigenous communities, defining measures for Collective Benefit, Authority to Control, Responsibility, and Ethics.

In our context of modeling the ocean data ecosystem with a focus on data interoperability, we find the FAIR principles a useful tool, as they are meant to provide

data producers and publishers measurable guidelines for ensuring their data implementation to be Findable, Accessible, Interoperable, and Reusable, to overcome the barriers for large scale data utilization. The need to follow the FAIR principles stems from the fact that gathering the data required for answering research questions is often a tedious and time-consuming task, largely due to lack of attention to how the data assets are preserved when they are created (Tanhua et al., 2019b). The FAIR principles answer to the step-by-step process by which machines will be able to process the data, identifying the relevant data within a given context, determining if it is useful, if it is usable in terms of license or other accessibility and taking action (Tanhua et al., 2019b).”

In addition, we clarify in the introduction that our ocean data ecosystem model was developed in the context of our focus on data interoperability (lines 66-70):

“*To this end, this paper presents an overview of the ocean-data ecosystem and proposes a model that maps its key actors and their roles, with particular attention to data platforms and interoperability. The discussion is framed from the perspective of the marine researcher as end user, with the overarching goal of facilitating synergetic work between the actors involved in different aspects of ocean data, thus improving the ability to utilize the vast amount of available ocean data to address important scientific questions”*.

2. A key subtlety lost in the manuscript is that each data entity or service described has a different user community with duplication of data widespread across the ecosystem. For example, in this context many of the data assets in WOD are an aggregation of subsets of data from SDN/PANGAEA/NOAA/etc but they have been curated into a product for a specific purpose and user community where WOD is the modern equivalent of a hydrographic atlas. This duplication of assets is ubiquitous in the ecosystem e.g. EMODNet aggregates data form many sources (SDN, ICES, CMEMS, etc) to build products for specific applications and users.

3. The role of DataONE and SeaDataNet overlap significantly with one being a US entity bringing together American data assets and the later being a European entity. Albeit they have different funding, governance and collaboration agreements but the similarity in their high-level functions for end users is not clear in the revised manuscript.

Answer: We thank the reviewer for drawing our attention to these important points. To address the missing information identified in these two comments, we added the following text to our general description of the data sources and referenced the relevant sections where these sources are discussed (lines 492-507):

"Notably, different data sources may exhibit significant similarity, often with overlap between their contents. In general, each data source serves a distinct user community, and datasets may be duplicated throughout the ecosystem. For example, many records in the World Ocean Database (WOD, see below Sec. 3.4.1) are compiled from subsets of data originating in datasets such as PANGAEA (see below Sec. 3.4.5.), but are curated into a product designed for a specific purpose and audience. Overlap and similarity between data sources can be exemplified for the cases of SeaDataNet (see below Sec. 3.4.3) and DataONE (see below Sec. 3.4.7), which although differing in their funding structures, partnership frameworks and infrastructural approach, align in their high-level functions for end users."

In my initial review I raised the lack of vision presented in the manuscript for the next 20 years based on the special issue scope, This feedback does not appear to be addressed significantly by the authors with the manuscript presenting current trends in the summary section.

<u>Answer</u>: **We have elaborated our vision for the future of the ocean data ecosystem by adding the following text to the summary and conclusion section (lines 1107-1137):**

"Looking ahead over the next two decades the ocean data ecosystem is set to undergo further transformations, driven primarily by dramatic growth in the amount and diversity of oceanic data, and by rapid technological developments. The expected increase in data availability and diversity is a natural continuation of the growing use of autonomous and remote sensing platforms, expansion of global observation networks, and improved ability to collect and analyze new data types suchas environmental DNA and underwater imagery. Advances in data collection methods results in an unprecedented influx of ocean data each day, often in real-time, propelling ocean research into the era of big data, characterized by vast volumes, diverse formats, and widely dispersed datasets (Tanhua et al., 2019).

The synthesis presented in this review points to a decisive transition: the ocean data ecosystem is evolving from a patchwork of independent repositories into a globally connected, service-oriented network. Interoperable standards, shared vocabularies, and federated cloud infrastructures are dismantling the old "portal–download" paradigm and enabling machine-actionable data flows across disciplines and borders. The next decade will likely see near-real-time discovery, access, and fusion of ocean observations - from autonomous sensors to satellite archives - within a seamless digital environment.A defining feature of this emerging landscape is the integration of advanced artificial intelligence (AI) and machine-learning methods. As the volume, velocity, and variety of ocean data continue to grow, AI is becoming indispensable for automated quality control, feature extraction, and pattern recognition across heterogeneous data streams. Deep-learning models can already detect mesoscale

eddies, track marine heatwaves, and identify biodiversity "hot spots" in vast image libraries. Looking ahead, AI-driven digital twins of the ocean will couple observation networks with predictive models to deliver near-instant forecasting and scenario testing, transforming both basic research and operational decision-making.

This technological leap will also reshape the social and governance dimensions of the ecosystem. Trust, transparency, and inclusivity, embodied in the FAIR, CARE, and TRUST principles, remain critical as AI systems begin to make or recommend management choices. International initiatives such as the UN Ocean Decade, the Ocean Data Action Coalition, and IOC/IODE programs are fostering open access and shared stewardship, while highlighting the need for clear policies on data provenance, algorithmic accountability, and equitable participation. Ethical AI frameworks, together with persistent identifiers and accreditation schemes, will help maintain confidence in automated analyses and encourage collaboration across nations and institutions.

Ultimately, the ocean data ecosystem is poised to become an active, intelligent engine for discovery and policy. By linking high-resolution observations, interoperable standards, and AI-powered analytics, it will enable rapid synthesis of knowledge for climate adaptation, biodiversity conservation, sustainable fisheries, and the broader blue economy. In this envisioned future, ocean data are not merely archived, they are continuously analyzed, interpreted, and applied, allowing scientists, governments, and society to anticipate and respond to a changing ocean with unprecedented speed and precision."

The revised version introduces a significant issue into the manuscript. The revised version attempts to build on the previous version by describing the eco-system as an ontology. I have a good awareness of ontologies but am not an expert on the subject so sought advice from three experts in different contexts; specifically global ocean data brokering, and industry. I sought feedback based on the openly available published ontology excluding the manuscript to preserve manuscript confidentiality. A summary of the feedback from these specialists is below. This feedback is critical to address for the work to be of publication quality. From my perspective as a reviewer the value and interesting element of the original submission was in its broad context rather than the ecosystem model. The ecosystem model provided the structured context as a wrapper for the rest of the manuscript making a complex system understandable to the reader. Based on the expert feedback publishing the model as ontology in its current form adds little value unless it builds existing domain knowledge and is technically sound in its formulation. Additionally, A mature relational graph between entities has the potential to address points 2) and 3) raised above.

I also disagree with the final conclusion in the revised in manuscript. The original manuscript was an analysis of the existing data ecosystem using the relationship model to help the

reader readily navigate and understand what is an inherently a complex collection of data nodes and brokered data because of numerous dependencies including geopolitical requirements, requirements of varied user communities and collaborations, different funding environments, and human behaviour. The revised manuscript presents the pre-existing ecosystem as a new concept which is a questionable conclusion with the ecosystems having to exist already for this review paper to be possible (with the cited literature demonstrating pre-existence of an ecosystem). To me as the reviewer novelty and value in the first version was in helping a researcher navigate the complex data ecosystem that exists in the ocean community which has been somewhat lost in the revised manuscript.

Ontological feedback based solely on the openly published ontology at .

Global ocean data brokering context

Foremost, this isn't really an ontology: the class-subclass relationship is not ontological, in that subclasses do not inherit all properties of superclasses. In fact, they're often completely different things.

For example, a stakeholder can't be an OWL subclass of an "ocean digital ecosystem". That's like saying a chair is a type of house. An industry stakeholder can be a type of stakeholder, so that's right. It's a little patchy, and if this was submitted to me by a student in an ontology class, I wouldn't pass them.

Another thing that's missing is any consistent and logically coherent definitions. So most of the semantics here are roughly implied. This is more a mind map.

• Are there strong or excessive overlaps with existing ontologies?

As this is not an ontology, I don't think this is the right question. However, there are strong overlaps with resources like ODISCat and Ocean Expert which already map many components of the ocean digital ecosystem. This is especially true as they now release their content in JSON-LD using schema.org semantics, which gives them more semantic rigour than the resource you're reviewing.

OE and ODISCat are definitely not a perfect resources, so I would see merit in a research group grabbing its JSON-LD/RDF and creating a curated / improved product.

One could also point to regional maps of organisations and stakeholders, like EDMO and EDMERP, even MEDIN's offerings as an ODIS Node or independently.

• Is there alignment with concepts in similar or related ontologies needed in this ontology?

They should not attempt an ontology construction here - I don't think it will help them. It seems that they would make more progress with building a relaxed knowledge graph as an example of how to map a digital ecosystem.

However, I note again, much of this is redundant with ODISCat and Ocean Expert, especially when combined with ORCID, ROR, and other systems. It would be distracting and counterproductive for another group to claim they are running an authoritative map of the ecosystem.

More broadly, if the paper claims to be about some "new" concept of digital / data ecosystems, please do remind them this is not new and many actors have been working on it.

Industry context

What you've shared is not an ontology, it just happens to use an ontology modelling language to do something hopefully useful for them. Then there are the fundamental issues in the misuse of the relations; they have used subclass more like a partOf relation and then failed to produce any descriptions of the classes they have defined. From a FAIR perspective it fails on a number of fronts.

There are some quite major issues in understanding what they expect to get from this work as well. I would suggest looking at something like Kumu if they have not already done so to deliver research / community value, however it sounds like we would also have concerns in this also from a perspective of being novel vs other efforts. Other products are available, but it's the one I've seen The Alan Turing Institute use successfully. They should look towards a concept map to understand who is doing what and using what in the landscape with questions in mind of what they would like to answer rather than graphs for technology and visual interest.

<u>Answer</u>: We thank the reviewer for taking the pains to obtain additional expert reviews from Ontology experts and apologize for the unexpected time and effort this incurred. The incentive for development of the ontology came from the suggestion of reviewer #1, and was meant to provide an online tool that can be kept up-to-date to ensure relevance of the manuscript. Thanks to this thoughtful comment, we realize that it was presumptuous of us to elevate our model to an ontology without rigorous analysis of existing ontologies and a more accurate integration of the suggested models within the existing ontology landscape. While we believe this should be our eventual goal, we agree with the reviewer that the current state is premature and inaccurate. We defer the creation of a comprehensive knowledge graph, based on the ontologies suggested by the reviewer and, if necessary, a minimal extension of these, to future work. We thank the reviewer for the suggestion to use a more loosely defined relation model and the Kumu tool for visualizing the model. We have done so and provided a description, and a reference to a public Kumu model instead of the Ontology in the previous revision. Accordingly, we have revised the manuscript as follows (lines 187-195):

"*For the purpose of providing an interactive map, we created an online ocean data ecosystem relational model available at* https://kumu.io/odini/ocean-data-ecosystem *the map is a long-term reference, open, and may be updated and extended. To facilitate contributions and comments from the public, we make a public issue system available* https://gitlab.com/odini_dev/data-ecosystem-ontology*) and invite readers to suggest additions and corrections .*

The methodology we used is based on a thorough literature and website review of over 90 scientific articles and over 100 websites. Articles and examples selected to illustrate the elements of the model have been selected based on using search terms such as 'ocean data,' 'ocean data interoperability,' and 'marine ontologies.' The examples are not exhaustive and are intended as a starting point for ocean data professionals, who are encouraged to explore additional resources specific to their research areas."

In addition, the concluding paragraph of the paper was modified as follows (lines 1137-1149):

"In summary, by mapping the market landscape in the field of ocean data, this review paper is meant to enable the reader, especially the new entrant to the ocean data field, to establish an understanding of the ocean data sector. To maintain long-term relevance, the ecosystem model presented is aimed to be used as a tool in further characterization of the ocean data ecosystem. The examples given are not exhaustive, and the reader may further identify relevant examples within their domain. The model has been placed as an open online resource, describing the elements of the data ecosystem as concepts, and examples as instances with relationships. The model is open to be further validated and refined by the ocean data community. The results bridge gaps between different disciplines and levels of familiarity with ocean data. We provide an up-to-date analysis of ocean data sources and emerging solutions and a summary of relevant data standardization efforts such as marine standards, vocabularies, and ontologies. By characterizing the ocean data ecosystem, we intend to assist the scientific community in identifying the gaps, current needs and future vision of the ocean data ecosystem. This work aims to contribute to the development of needs-based solutions, components, products, services, and technologies, thus contributing to the evolution of the ocean data ecosystem and promoting data-based ocean research."