Interconnecting Arctic observatory data through machine-actionable knowledge representation: are ontologies fit for purpose?

Kai Blumberg

Submission Date 15.03.18

Contents

Table of Contents **2**

Introduction **2**
Rapid effects of climate change on Polar systems . . . 2
Polar ocean observatories and marine monitoring programs . . . 2
Policy and SDGIOs . . . 3
UN sustainability development goals in response to climate change . . . 3
Need for semantics in Environmental data . . . 3
FAIR . . . 3
Linking earth science data initiatives such ESIP Open knowledge network to the UN SDGIO's 4
Ontologies and the OBO Foundry . . . 4
competency questions: . . . 5

Materials and Methods **6**
Datasets used in Datastore . . . 6
programs used: . . . 6
Semantic Data Annotation . . . 6
sparql query scripting . . . 6

Results **6**
Competency Questions . . . 7
Lookup author of ontology term . . . 7
Retrieve any data which is about a subclass of sea ice . . . 7
What compounds play a role as algae metabolites? . . . 7
PCO contributions & Plankton Ecology . . . 8
cryoMIxS . . . 8
post compositional data annotation model . . . 9
Example of Post Compositional Data Annotation with Ontology Terms . . . 9
Vocamp Virtual Glacial Hackathon . . . 9
AWI DBPEDIA contributions . . . 10
UNEP SDGIO . . . 10

Discussion **10**
Querying Semantically Annotated Data . . . 10
Creating Classes vs post compositional annotation for data annotation . . . 10
Semantics as AWI Public Outreach . . . 10

Conclusion **10**

Outlook **10**

Appendices **11**

Python Scripts and Documentation 11

script 1 11

script 2 11

References **11**

Table of Contents

Introduction

Rapid effects of climate change on Polar systems

Anthropogenic green house gas emissions are leading to increased climate change and weather extremes.

purl	label	purl	la
http://purl.obolibrary.org/obo/CHEBI_17992	Sucrose	http://purl.obolibrary.org/obo/CHEBI_17992	S
http://purl.obolibrary.org/obo/CHEBI_80716	aplysiatoxin	http://purl.obolibrary.org/obo/CHEBI_80716	a
http://purl.obolibrary.org/obo/CHEBI_90820	11(R)-HEPE(1-)	http://purl.obolibrary.org/obo/CHEBI_90820	1
http://purl.obolibrary.org/obo/CHEBI_86386	3-mercaptopropionate	http://purl.obolibrary.org/obo/CHEBI_86386	3
http://purl.obolibrary.org/obo/CHEBI_17754	Glycerin	http://purl.obolibrary.org/obo/CHEBI_17754	G
http://purl.obolibrary.org/obo/CHEBI_17754	glycerol	http://purl.obolibrary.org/obo/CHEBI_17754	g
http://purl.obolibrary.org/obo/CHEBI_16810	2-oxoglutarate(2-)	http://purl.obolibrary.org/obo/CHEBI_16810	2
http://purl.obolibrary.org/obo/CHEBI_16914	salicylic acid	http://purl.obolibrary.org/obo/CHEBI_16914	s
http://purl.obolibrary.org/obo/CHEBI_16914	Salicylic Acid	http://purl.obolibrary.org/obo/CHEBI_16914	S
http://purl.obolibrary.org/obo/CHEBI_16811	Methionine	http://purl.obolibrary.org/obo/CHEBI_16811	M

Microbes and Biogeochemical cycles

> The prokaryotic and eukaryotic microorganisms that drive the pelagic ocean's biogeochemical cycles are currently facing an unprecedented set of comprehensive anthropogenic changes [1]

Polar ocean observatories and marine monitoring programs

Polar marine monitoring initiatives such as FRAM … are working to gauge the effects of climate change on such rapidly changing environments.

FRAM & HAUSGARTEN

At the forefront of climate change affected environments are polar habitats.

HAUSGARTEN intro: [2]

FRAM intro: [3]

AtlantOS

//maybe mention this?

the Atlantic Ocean Observation Systems (AtlantOS) 1st AtlantOS Briefing Paper

Policy and SDGIOs

Making Marine Life Count: A New Baseline for Policy [4] Just use a little bit from this as policy intro.

DOOS Consultative Draft (no DOI) for insight into functions that can be understood as ecosystem services of the deep, and thus linked to natural capital.

UN sustainability development goals in response to climate change

The effects of increased climate change and extreme weather events are hardest felt by indigenous people and the global precariat subsiding via land and ocean subsistence farming and fishing.

UN publication: TRANSFORMING OUR WORLD: THE 2030 AGENDA FOR SUSTAINABLE DEVELOPMENT no DOI reference for the sustainable development goals and targets.

The UN framework for SDG's have setup targets for improvements to many global issues such as UN SDG 14 for ocean health.

14.1

> By 2025, prevent and significantly reduce marine pollution of all kinds, in particular from land-based activities, including marine debris and nutrient pollution

link the nitrogen phosphorus data to the concept of those cycle being out of balance as doccumented in the Planetary Boundaries: Exploring the Safe Operating Space for Humanity paper. [5]

Need for semantics in Environmental data

Observatories generate considerable volumes and varieties of data. The management and integration of such data remains a major obstacle, as the data are often not semantically interoperable. I.e. the data cannot be used in combination, because they are not annotated with a controlled vocabulary of interconnected terms which would allow for a computer to perform logical reasoning upon them.

FAIR

the FAIR data guiding principles (machine-focused findability, accessibility, interoperability reusability) [6]

PANGAEA

observational networks often upload their data to open access repositories such as the PANGAEA

Although vast quantities of environmental data are freely available to the scientific community, integrated analysis of such data is hindered by a lack of logical connections between different types of data.

Linking earth science data initiatives such ESIP Open knowledge network to the UN SDGIO's

There exist a variety of earth and life science initiatives attempting to capture and represent the knowledge associated with environmental data. ...

The knowledge required to interface the concepts needed for the Sustainable development goals are represented in a machine operable form via the SDGIO sustainable development goals interface ontology.

role of data in 2015 - 2020 ESIP Strategic Plan

link to my log

OPeNDAP

> OPeNDAP will be a fundamental component of systems which provide machine-to-machine interoperability with semantic meaning in a highly distributed environment of heterogeneous datasets.

Open-source Project for a Network Data Access Protocol There is a need for semantic interoperability ...

knowledge outreach

Knowledge graphs are becoming more popular and useful, need to bridge the gap between patchy but growing resources such as Wikipedia, and expert knowledge (locked away in text books), using an ontology helps to bridge this, it can be applied to querying Wikipedia data and for improved semantic representation make data FAIR. Ontology for an agreed upon term structure

Ontologies and the OBO Foundry

Ontology, a human and machine readable semantic representation of domain knowledge ...

An ontology is a hierarchically structured, machine and human readable representation of the knowledge used by experts to describe entities, and capture the relationships between them [7]. In informatics, ontologies exist in the form of a knowledge graph, where nodes represent entities, and edges represent logical relations linking entities together (i.e. axioms). Ontologies provide a digital semantic infrastructure upon which advanced querying, discovery and analysis of data can occur.

Ontologies are a methodology to systematically structure and connect data, allowing users to ask more complicated questions involving the synthesis of disparate data types which currently can not be combined.

//revise a bit from lab rotation: Because, no single knowledge graph can encompass the needs of interdisciplinary projects, work must be done in a coordinated fashion with other ontology researchers and developers. In order to interconnect ontologies representing scientific knowledge from different domains, the Open Biological and Biomedical Ontology (OBO) Foundry and Library was created [7]. The OBO Foundry and Library established a set of principles by which to develop and coordinate ontologies such that the scientific knowledge they represent and hence the data they link can interoperate. These ontologies share a common upper level in the hierarchy and use of the same types of logical connective operations to interlink their knowledge. Following these principles are a family of ontologies representing scientific knowledge from non-overlapping domains, which can be used in combination to describe natural phenomena in greater depth. OBO compliant ontologies make use of the Basic Formal Ontology (BFO), to ensure they have a compatible hierarchical structure, and use logical relations from the Relations Ontology (RO), to standardize the connections between their knowledge.

OBO compliant ontologies can be benefit observatory networks such as Hausgarten FRAM, by providing connections between data collected by researchers of different disciplines studying overlapping entities.

//example from my rotation add something like this. > For example sea ice physicists studying the reflectivity of various ice mass features, may have light intensity data that would help microbial ecologists studying photosynthetic bacteria in brine channels, to calculate the light dependent growth rates of such bacteria

SDGIO

United Nations Environment Programme

SDGIO is an OBO compliant ontology

uses the same interoperable semantic standards to ENVO. Although UNEP PURLS cannot currently be queried.

ENVO for representing environmental semantics.

ENVO papers: [8] [9]

The Environment Ontology (ENVO) represents expert knowledge about different types of environments[8][9]. ENVO is an OBO aligned ontology.

Environmental knowledge represented by ENVO is used to annotate data from a variety of life science disciplines including oceanography and polar research. [8][9]

Gene Ontology

go paper: [10]

GO frequently used to interpret omic data [10]. It has been used to do genomewide RNA expression profile data to compare samples based on shared biological pathways. [11]

The combination of GO and ENVO is less frequently attempted. [12]

Paring GO with ENVO is a potential avenue for future study allowing researchers to ask questions such as > "What is the omic potential of microbes associated with particular environments?".

Example Ontology uses

A communal catalogue reveals Earth's multiscale microbial diversity. //Uses EMPO a light-weight application ontology built on ENVO the Earth Microbiome Project Ontology [13] //good to have an example which demonstrates the utility of ENVO for an application ontology to provide utility.

//from my rotation rewrite example > Thesen et al.13. show how such a federated semantic approach can enhance handling of environmental and phenotype data, in order to ask increasingly complex questions such as "Which crop varieties are expected to do well in a particular location over the next century?". Thesen et al Emerging semantics to link phenotype and environment [14]

competency questions:

In order to leverage growing data and knowledge representation semantic infrastructure we test if a semantic knowledge web represented by an ontologies can be used in combination with AWI data to address competency questions such as:

Materials and Methods

Datasets used in Datastore

1. Inorganic nutrients measured on water bottle samples at AWI HAUSGARTEN during POLARSTERN cruise MSM29. [15]
2. Physical oceanography and current meter data from mooring TD-2014-LT. [16]
3. Chlorophyll a measured on water bottle samples during POLARSTERN cruise ARK-XXIV/2. [17][18]
4. Global chlorophyll "a" concentrations for diatoms, haptophytes and prokaryotes obtained with the Diagnostic Pigment Analysis of HPLC data compiled from several databases and individual cruises. [19][20]
5. Biogenic particle flux at AWI HAUSGARTEN from mooring FEVI7. [21][22]
6. Snow height on sea ice and sea ice drift from autonomous measurements from buoy 2015S22, deployed during the Norwegian Young sea ICE cruise N-ICE 2015. [23][24]
7. Sea ice thickness at Ice Camp 1 on 2013-09-01 (GEM2IceTh_DiveHole_IceStation1). [25][26]
8. Ice-algal chlorophyll a and physical properties of multi-year and first-year sea ice of core CASIMBO-CORE-1_10. [27][28]
9. //TODO add genomic data preferably some FRAM data from eddie like an otu table and or functional genomic table.

programs used:

sparql, python, N3, turtle, any23, owl, Protégé

Semantic Data Annotation

Semantic annotation of example data was conducted in the RDF serialization turtle, drawing upon its blank node feature to facilitate scripting owl code in RDF. Annotations make use ontology terms from the OBO Foundry [7]. Ontology terms can be search for using Ontobee A linked data server hosting ontologies and their terms. [29]

sparql query scripting

scripts to perform queries were written in python verion?

using the rdf-lib module

Queryies preformed against the ontobee endpoint http://sparql.hegroup.org/sparql/ a serive provied by the He Group [29]

The script makes use of a conjunctive graph object from the rdf-lib module, to emulate an RDF triple store.

Results

In my masters thesis work I have devised a semantic data annotation and querying schema. It allows for the phenomena inhering in data, to be represented and searched in the same way as ontology classes. Annotating data to be semantically inter-operable with existing ontologies, allows us to ask questions of interdisciplinary data, making use of the connections between phenomena encoded within ontologies.

In my masters thesis work I have been writing scripts to assemble and query a demonstration datastore comprised of semantically annotated AWI data. As a part of my proposed work, I would create a human and machine-readable web accessible endpoint to host a variety of AWI data, as well as a the semantic search tools to facilitate querying it.

Competency Questions

experiments to test knowledge model against competency questions.

Lookup author of ontology term

see my thesis here

Retrieve any data which is about a subclass of sea ice

//easy to bang out see my thesis here

What compounds play a role as algae metabolites?

easy enough to answer Make use of the CHEBI class: algal metabolite

purl

querying the ontobee sparql endpoint

```
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
SELECT DISTINCT ?purl (STR(?label) as ?label)
WHERE
{
  ?purl rdfs:subClassOf/owl:someValuesFrom obo:CHEBI_84735.
  ?purl rdfs:subClassOf/owl:onProperty obo:RO_0000087.
  ?purl rdfs:label ?label.
}
GROUP BY ?purl
LIMIT 10
```

This query gives us the purls and the labels of the first 20 classes which are subclasses of 'has role' some algal metabolite

using the restriction has role.

The group by ?purl is to ensure we don't get duplicates of purls which have duplicated labels such as http://purl.obolibrary.org/obo/CHEBI_15756 which has labels: `hexadecanoic acid` and `Hexadecanoic acid`

Returning the following results:

Compounds serving as algal metabolites.

Table 2: Compounds serving as algal metabolites.

purl	label
http://purl.obolibrary.org/obo/CHEBI_17992	Sucrose
http://purl.obolibrary.org/obo/CHEBI_80716	aplysiatoxin
http://purl.obolibrary.org/obo/CHEBI_90820	11(R)-HEPE(1-)
http://purl.obolibrary.org/obo/CHEBI_86386	3-mercaptopropionate
http://purl.obolibrary.org/obo/CHEBI_17754	Glycerin
http://purl.obolibrary.org/obo/CHEBI_17754	glycerol
http://purl.obolibrary.org/obo/CHEBI_16810	2-oxoglutarate(2-)
http://purl.obolibrary.org/obo/CHEBI_16914	salicylic acid
http://purl.obolibrary.org/obo/CHEBI_16914	Salicylic Acid
http://purl.obolibrary.org/obo/CHEBI_16811	Methionine

PCO contributions & Plankton Ecology

//Assuming I get any of this stuff pushed to PCO and or ENVO.

I may still be able to write about the proposed design patterns for PCO, even if I don't get to submit a pull request.

Tilman Satelite Data

Paper: Diatom Phenology in the Southern Ocean: Mean Patterns, Trends and the Role of Climate Oscillations. [30] //Associated with the plankton ecology project using Tillman satellite chlorophyll data and the plankton bloom ontology classes.

cryoMIxS

original MIxS paper: [31]

//talk about my contributions to the cryoMIxS project. Including work from my lab rotation.

ENVO releases of interest:

Ecotone, Polar express, Hot tub time machine.

Sample grid table.

Table 3: Sample grid table.

Fruit	Price	Advantages
Bananas	$1.34	• built-in wrapper • bright color
Oranges	$2.10	• cures scurvy • tasty

post compositional data annotation model

//Maybe this could go in material and methods but I'll argue this is a result in the sense of semantic research, a model for data annotation.

In this work we present a novel semantic data annotation model. Semantics have been used to represent data ... //TODO FIND REFS. In this model data annotations are composed of terms from the OBO Foundry. Data annotations are written in The RDF turtle specification, and structured as nested owl classes. Annotating the data as owl classes ensures parity to the OBO ontologies. This enables us to perform sparql queries on the annotated data in the same manor as would be done to query OBO Foundry ontologies.

In order to emulated owl code written in RDF, we chose the turtle RDF format for its ability to nest blank nodes within strings of triples.

//ADD THE is about property in the data model, it could also be cool to have a vue figure which explains the workflow.

The creation of ontology classes involves the composition of axioms, the links between classes, which are assembled from other preexisting ontology classes and relational properties. In ontology development this is refereed to as precomposition, which has the effect of taking a set of ontology classes and properties and joining them together in a specific way and assigning this assemblage to be a novel class.

The proposed semantic data annotation model allows for this process to be done in reverse. This is not necessary when an appropriate term for annotation already exists, however, in cases where the appropriate annotation term is lacking, it can be created from a combination of other terms. This practice, referred to as "post composition", enables a user to annotate their data with axioms that comprise a non existent ontology term. By writing the data annotations as owl classes, they are functionally equivalent to existing ontology classes, in terms of their ability to be searched for using a sparql query.

This allows for the phenomena inhering in data, to be represented in a machine readable semantic layer prior to their incorporation as ontology terms.

The model makes use of owl equivalence classes, to structure the annotation as the intersection (and) and or union (or) of post compositionally annotated classes.

Thus the proposed data annotation model will allow for users, who are not ontologists, to post compositionally annotate their data. //ADD section about how I'll write a tool to automate this in the outlook.

Example of Post Compositional Data Annotation with Ontology Terms

change this competency question example to be about how to annotate data which is about a **marine environment determined by a diatom community** or a **marine environment determined by a diatom community bloom** instead of being about I intened to create these classes.

Vocamp Virtual Glacial Hackathon

vocamp:

> VoCamp is a series of informal events where people can spend some dedicated time creating lightweight vocabularies/ontologies for the Semantic Web/Web of Data.

Virtual-Hackahon-on-Glacier-topic

//to be held on Feb. 2nd. I should have an example of moving snow and ice related data ready to demonstrate by then.

AWI DBPEDIA contributions

Contributing semantic knowledge to the website Wikipedia in the form of an improved heirarchaly structure but aligning with ENVO.

Implement and talk about dpbedia contributions, hopefully they'll let me edit. My intention is to align dbpedia glacial semantics to those in ENVO, should be relatively quick and easy once I can edit.

UNEP SDGIO

Despite operating within a semantically which is interoperable with the OBO Foundry the UNEP ontology is currently non queryable. Future work needs to be done to improve the way SDGIO purls are hosted via UNEP so that they can be querable. This would allow for the the incorporation of data mobililzed via semantics to the UN SDGs to help achieve their objectives.

Discussion

Querying Semantically Annotated Data

using polar semantics to annotate AWI Polar data in a machine-readable way. This allows for knowledge to be captured in a data querying

Creating Classes vs post compositional annotation for data annotation

Semantics as AWI Public Outreach

AWI Education & Communication

Contributions to semantic models such as those discussed in this work serve to improve AWI public outreach efforts to educate and communicate polar research outputs to the public. Dissemination of AWI knowledge has been demonstrated in this work via the contributions made to the open source encyclopedia Wikipedia. This was achieved by aligning the dpbedia ontology glacial semantics to those of ENVO, which were contributed during this work.

Conclusion

This work has demonstrated that semantics can be used to mobilize polar data.

Outlook

I believe the use of ontologies and semantics data annotation could serve as a valuable tool to address broad biological questions, such as those in the Raes et al. 2017 paper, about which mechanism, temperature or productivity is responsible for marine microbial diversity.

An outlook for the goals presented in this work would be to semantically annotate a wide variety of interdisciplinary AWI datasets in order render such data machine-readable and query-able. This creates the possibility to ask deeper questions of large data sets to address fundamental biological questions such as: "Does microbial diversity coincide with temperature or with primary productivity sourced from nitrogen fixation?"

Such questions could be asked of semantically annotated and machine-readable genomic datasets, which contain basic metadata. Such data could be sourced from anywhere, in house AWI data or already published data, from a variety environmental locations. Working with a data publication service such as PANGAEA to host such data in an open machine-readable web accessible format would allow for complex queries and questions to be asked.

For example to address the aforementioned question, we would perform a query to gather all datasets which include temperature, functional genomic and taxonomic information. From this ecological analysis could be conducted such as testing if temperature tends to correlate with microbial diversity, or with samples enriched in nitrogen fixation genes. The intentional interoperability between the Environment Ontology and the Gene Ontology would facilitate a query for the latter.

Appendices

Python Scripts and Documentation

script 1 . . .

script 2 . . .

References

1. **Hutchins DA, Fu F**. Microorganisms and ocean global change. *Nature Microbiology* 2017;2:17058.

2. **Soltwedel T, Bauerfeind E, Bergmann M, Budaeva N, Hoste E *et al.*** HAUSGARTEN: Multidisciplinary investigations at a deep-sea, long-term observatory in the arctic ocean. *Oceanography* 2005;18:46–61.

3. **Soltwedel T, Schauer U, Boebel O, Nothig E-M, Bracher A *et al.*** FRAM - FRontiers in arctic marine monitoring visions for permanent observations in a gateway to the arctic ocean. In: *2013 MTS/IEEE OCEANS - bergen*. IEEE. Epub ahead of print June 2013. DOI: 10.1109/oceans-bergen.2013.6608008.

4. **Williams MJ, Ausubel J, Poiner I, Garcia SM, Baker DJ *et al.*** Making marine life count: A new baseline for policy. *PLoS Biology* 2010;8:e1000531.

5. **Rockström J, Steffen W, Noone K, Persson, Chapin FSI *et al.*** Planetary boundaries: Exploring the safe operating space for humanity. *Ecology and Society*;14. Epub ahead of print 2009. DOI: 10.5751/es-03180-140232.

6. **Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M *et al.*** The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 2016;3:160018.

7. **Smith B, Michael Ashburner, Rosse C, Bard J, Bug W *et al.*** The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 2007;25:1251–1255.

8. **Buttigieg P, Morrison N, Smith B, Mungall CJ, and SEL**. The environment ontology: Contextualising biological and biomedical entities. *Journal of Biomedical Semantics* 2013;4:43.

9. **Buttigieg PL, Pafilis E, Lewis SE, Schildhauer MP, Walls RL *et al.*** The environment ontology in 2016: Bridging domains with increased scope, semantic density, and interoperation. *Journal of Biomedical Semantics*;7. Epub ahead of print September 2016. DOI: 10.1186/s13326-016-0097-6.

10. **Ashburner M, Ball CA, Blake JA, Botstein D, Butler H *et al.*** Gene ontology: Tool for the unification of biology. *Nature Genetics* 2000;25:25–29.

11. **Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL *et al.*** Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 2005;102:15545–15550.

12. **Henschel A, Anwar MZ, Manohar V**. Comprehensive meta-analysis of ontology annotated 16S rRNA profiles identifies beta diversity clusters of environmental bacterial communities. *PLOS Computational Biology* 2015;11:e1004468.

13. A communal catalogue reveals earth's multiscale microbial diversity. *Nature*. Epub ahead of print November 2017. DOI: 10.1038/nature24621.

14. **Thessen AE, Bunker DE, Buttigieg PL, Cooper LD, Dahdul WM *et al.*** Emerging semantics to link phenotype and environment. *PeerJ* 2015;3:e1470.

15. **Bauerfeind E, Kattner G, Ludwichowski K-U, Nöthig E-M, Sandhop N**. Inorganic nutrients measured on water bottle samples at AWI HAUSGARTEN during POLARSTERN cruise MSM29. Epub ahead of print 2014. DOI: 10.1594/PANGAEA.834685.

16. **Bauerfeind E, von Appen W-J, Soltwedel T, Lochthofen N**. Physical oceanography and current meter data from mooring TD-2014-LT. Epub ahead of print 2016. DOI: 10.1594/PANGAEA.861860.

17. **Nöthig E-M, Bauerfeind E, Metfies K, Simon S, Lorenzen C**. Chlorophyll a measured on water bottle samples during POLARSTERN cruise ARK-XXIV/2. Data Set; PANGAEA. Epub ahead of print 2015. DOI: 10.1594/PANGAEA.855799.

18. **Nöthig E-M, Bracher A, Engel A, Metfies K, Niehoff B *et al.*** Summertime plankton ecology in fram straita compilation of long- and short-term observations. *Polar Research* 2015;34:23349.

19. **Soppa MA, Peeken I, Bracher A**. Global chlorophyll "a" concentrations for diatoms, haptophytes and prokaryotes obtained with the Diagnostic Pigment Analysis of HPLC data compiled from several databases and individual cruises. Data Set; PANGAEA. Epub ahead of print 2017. DOI: 10.1594/PANGAEA.875879.

20. **Losa SN, Soppa MA, Dinter T, Wolanin A, Brewin RJW *et al.*** Synergistic exploitation of hyper- and multi-spectral precursor sentinel measurements to determine phytoplankton functional types (SynSenPFT). *Frontiers in Marine Science*;4. Epub ahead of print July 2017. DOI: 10.3389/fmars.2017.00203.

21. **Bauerfeind E, Nöthig E-M, Beszczynska A, Fahl K, Kaleschke L *et al.*** Biogenic particle flux at AWI HAUSGARTEN from mooring FEVI7. Data Set; PANGAEA. Epub ahead of print 2009. DOI: 10.1594/PANGAEA.714844.

22. **Bauerfeind E, Nöthig E-M, Beszczynska A, Fahl K, Kaleschke L *et al.*** Particle sedimentation patterns in the eastern fram strait during 20002005: Results from the arctic long-term observatory HAUSGARTEN. *Deep Sea Research Part I: Oceanographic Research Papers* 2009;56:1471–1487.

23. **Nicolaus M, Itkin P, Spreen G**. Snow height on sea ice and sea ice drift from autonomous measurements from buoy 2015S22, deployed during the Norwegian Young sea ICE cruise N-ICE 2015. Data Set; Alfred Wegener Institute, Helmholtz Center for Polar; Marine Research, Bremerhaven; PANGAEA. Epub ahead of print 2015. DOI: 10.1594/PANGAEA.846861.

24. **Nicolaus M, Hoppmann M, Arndt S, Hendricks S, Katlein C *et al.*** Snow height and air temperature on sea ice from Snow Buoy measurements. Epub ahead of print 2017. DOI: 10.1594/PANGAEA.875638.

25. **Ricker R, Krumpen T, Schiller M**. Sea ice thickness at Ice Camp 1 on 2013-09-01 (GEM2IceTh_DiveHole_IceStation1). Data Set; PANGAEA. Epub ahead of print 2017. DOI: 10.1594/PANGAEA.870689.

26. **Arndt S, Meiners KM, Ricker R, Krumpen T, Katlein C *et al.*** Influence of snow depth and surface flooding on light transmission through antarctic pack ice. *Journal of Geophysical Research: Oceans* 2017;122:2108–2119.

27. **Lange BA, Michel C, Beckers J, Casey JA, Flores H *et al.*** Ice-algal chlorophyll a and physical properties of multi-year and first-year sea ice of core CASIMBO-CORE-1_10. Data Set; PANGAEA. Epub ahead of print 2015. DOI: 10.1594/PANGAEA.842359.

28. **Lange BA, Michel C, Beckers JF, Casey JA, Flores H *et al.*** Comparing springtime ice-algal chlorophyll a and physical properties of multi-year and first-year sea ice from the lincoln sea. *PLOS ONE*

2015;10:e0122418.

29. **Ong E, Xiang Z, Zhao B, Liu Y, Lin Y *et al.*** Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. *Nucleic Acids Res* 2017;45:D347–D352.

30. **Soppa M, Völker C, Bracher A**. Diatom phenology in the southern ocean: Mean patterns, trends and the role of climate oscillations. *Remote Sensing* 2016;8:420.

31. **Yilmaz P, Kottmann R, Field D, Knight R, Cole JR *et al.*** Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature Biotechnology* 2011;29:415–420.