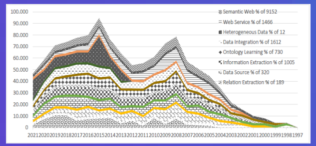
ICBO 2023
14th International Conference on Biomedical Ontology

USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES

The potential of ontologies for the empirical assessment of machine learning techniques in operational oceanography

Table 1
The main features provided by ontologies in support of PAMI

Ontology feature	Utility in PAMI
Classes and relations	When ontology reasoning is applied to sensor data, rdf:type will be connected to a class name of an ontology
Domain vocabulary	Ontologies provide a domain vocabulary that can be exploited to create a dense network of relationships among the entities, and serve software applications, and GIS
Metadata and descriptions	Biodiversity data, especially in marine domain, have database entities represented as ontologies where these last are primarily used for metadata that describe raw data providing contextual information
Axioms and formal declarations	Ontology axioms and applied reasoning on them are related to the recognition of object presence in a time interval



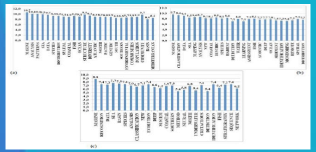
Levels of data visualization in machine learning methods for pattern analysis (marine data) (metadata generated from a base histogram

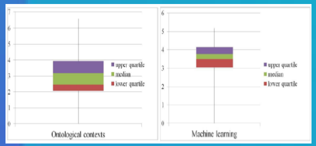
	analytics	association	case study	classification	comparison	complex	correlation	development	information	mapping	model	monitor	observation	pattern	prediction	recognition	regression	sensor	simulation	time series	other	SUM
Papers	1	1	4	70	7	3	1	4	2	16	20	12	2	13	13	7	1	12	3	2	16	210
disseminative				1					2			4		1						2	3	13
observational			1	11								2	2	1				12			4	33
analytical	1	1	3	34	7	3	1					6		10			1				3	70
model-developmental				24				4		16	20			1	13	7			3		4	92

***Enrique Wulff**, Marine Science Institute of Andalusia, CSIC*

Machine learning for pattern analysis time statistics (marine data)

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	SUM
Papers	3	3	7	5	4	6	9	12	16	15	27	55	48	210
disseminative							1		1	1	1	5	4	13
observational	1		1			1		1			6	12	11	33
Analytical	1	1	3	2	1	4	4	6	6	6	7	14	17	70
model-developmental	1	2	3	3	3	1	4	5	9	8	13	24	16	92





ICBO2023 – 30/08-01/09/2023

Goal of this Contribution

USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES

- To identify relevant pattern analysis research in marine data classification and recognition, and to review its intersection with the state-of-the art in marine ontologies
- Focus on the 3D modeling and analysis domain, computer vision and interactions are described for machine learning (ML) and marine ontologies
- Show how the use of ontologies for representing database entities has been advantageous in the field of Operational Oceanography (Riga et al., 2021)*

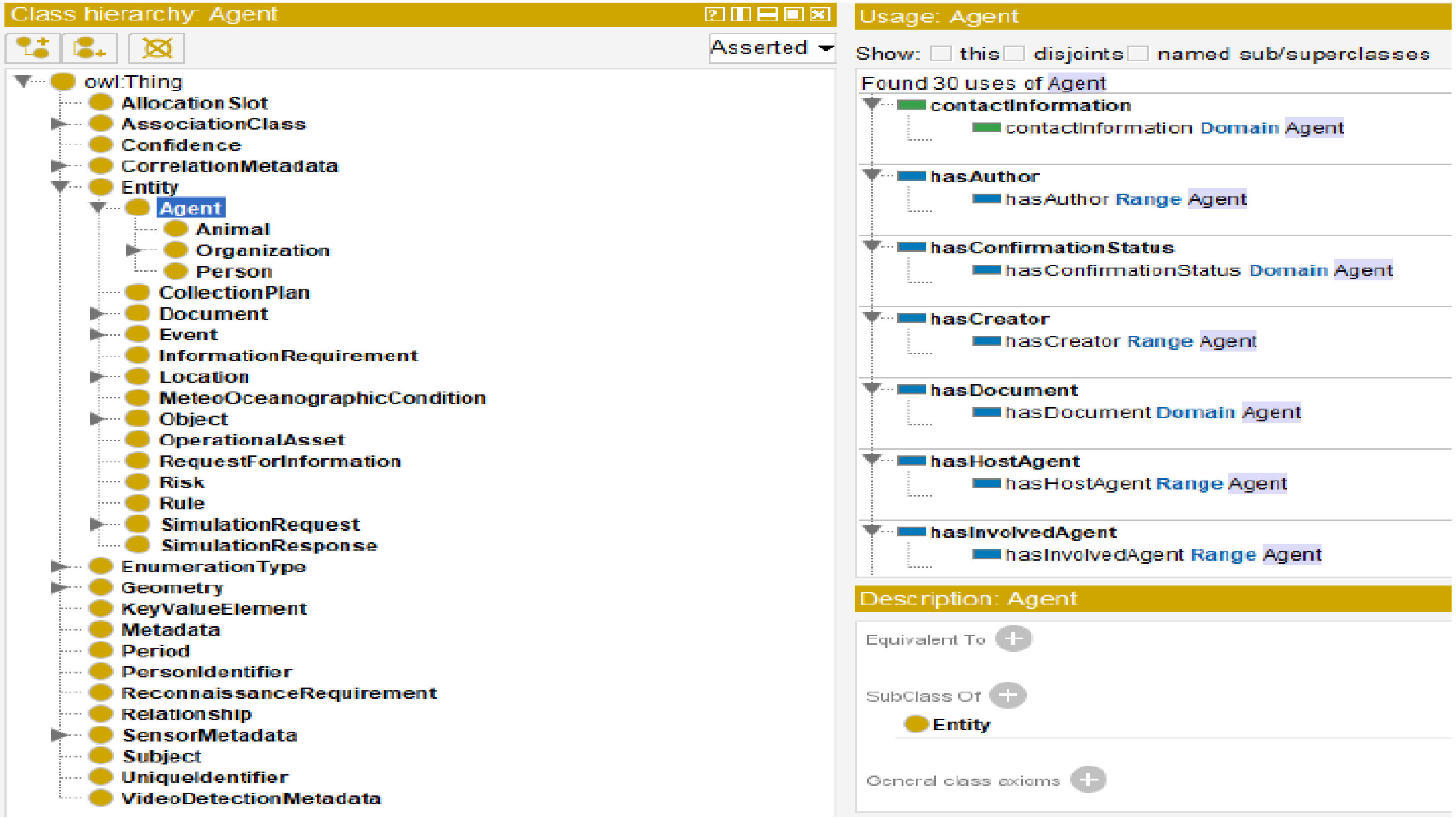
*M. Riga, E. Kontopoulos, K. Ioannidis, S.; Kintzios, S. Vrochidis, I. Kompatsiaris, EUCISE-OWL: An ontology-based representation of the Common Information Sharing Environment (CISE) for the maritime domain, *Semantic Web* 12 (2021) 603-615. doi: 10.3233/SW-200403

A key phrase from John Delaney underpins several ideas in these goals

- “There are emergent technologies throughout the fields around oceanography which we will incorporate into oceanography, and through that convergence we will make oceanography into something even more magical”

OCEANOGRAPHY

Visualization of the CISE ontology on Protégé

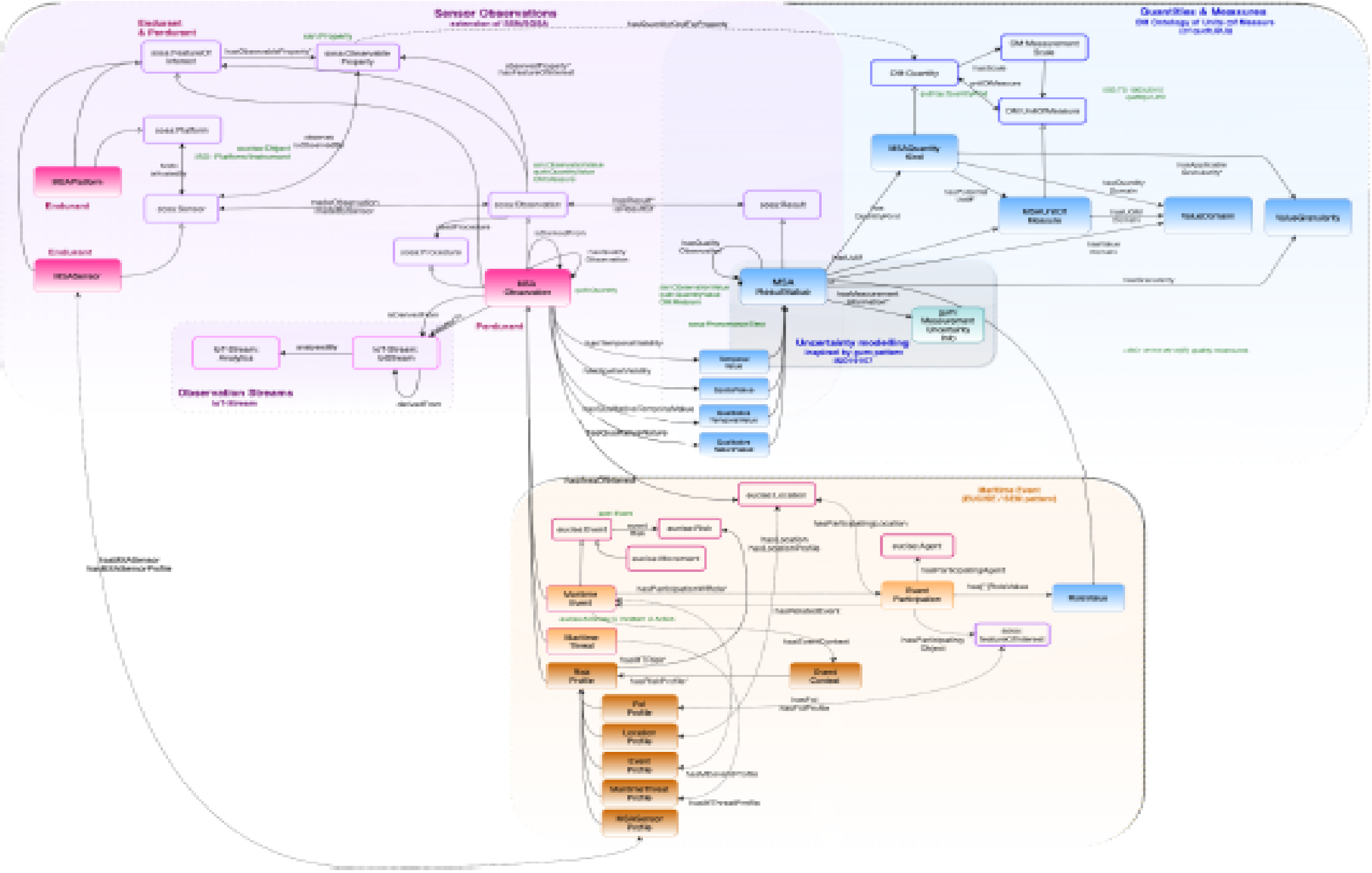


central classes of the CISE model concerning the Location entity

ICBO 2023
14th International Conference on Biomedical Ontology

Maritime Situational Awareness Heterogeneous Sensor Network Ontology (MSA-HSN)

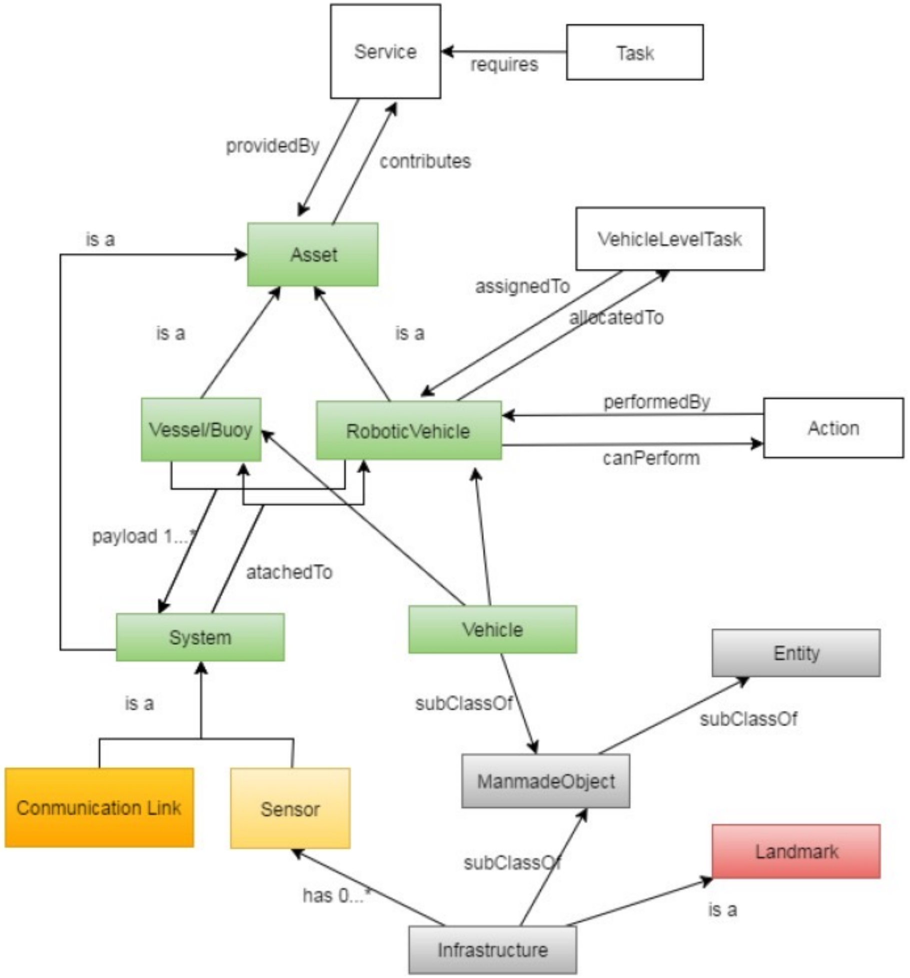
USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES



MSA-HSN, extended to support entity profiles and event context. Boxes represent top-level ontology concepts, including sensors and observations (violet background), observation values (light blue), and maritime events (orange). Arrows represent object properties linking concepts. Empty boxes are concepts from existing ontologies.

SWARMs Ontology

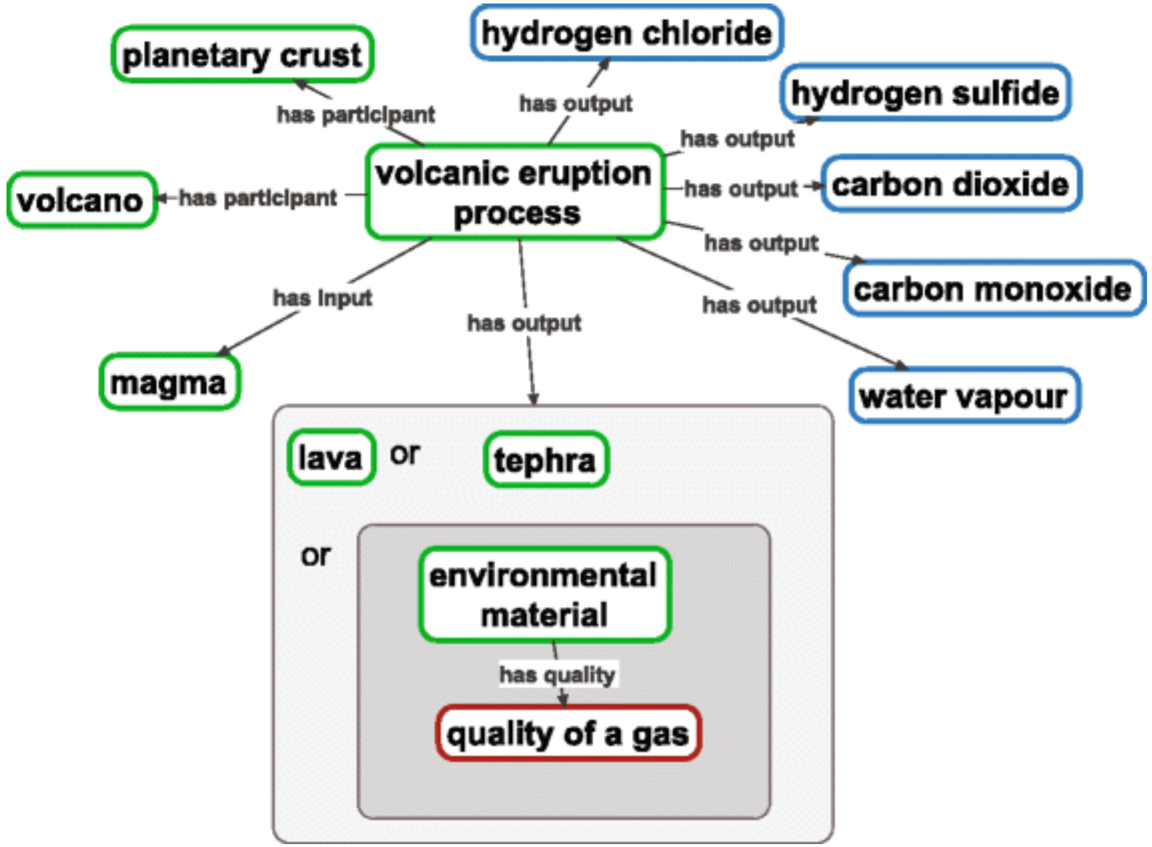
USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES



A representation of the overall structure of the core ontology for the cooperation of underwater robots

Environment Ontology (ENVO)

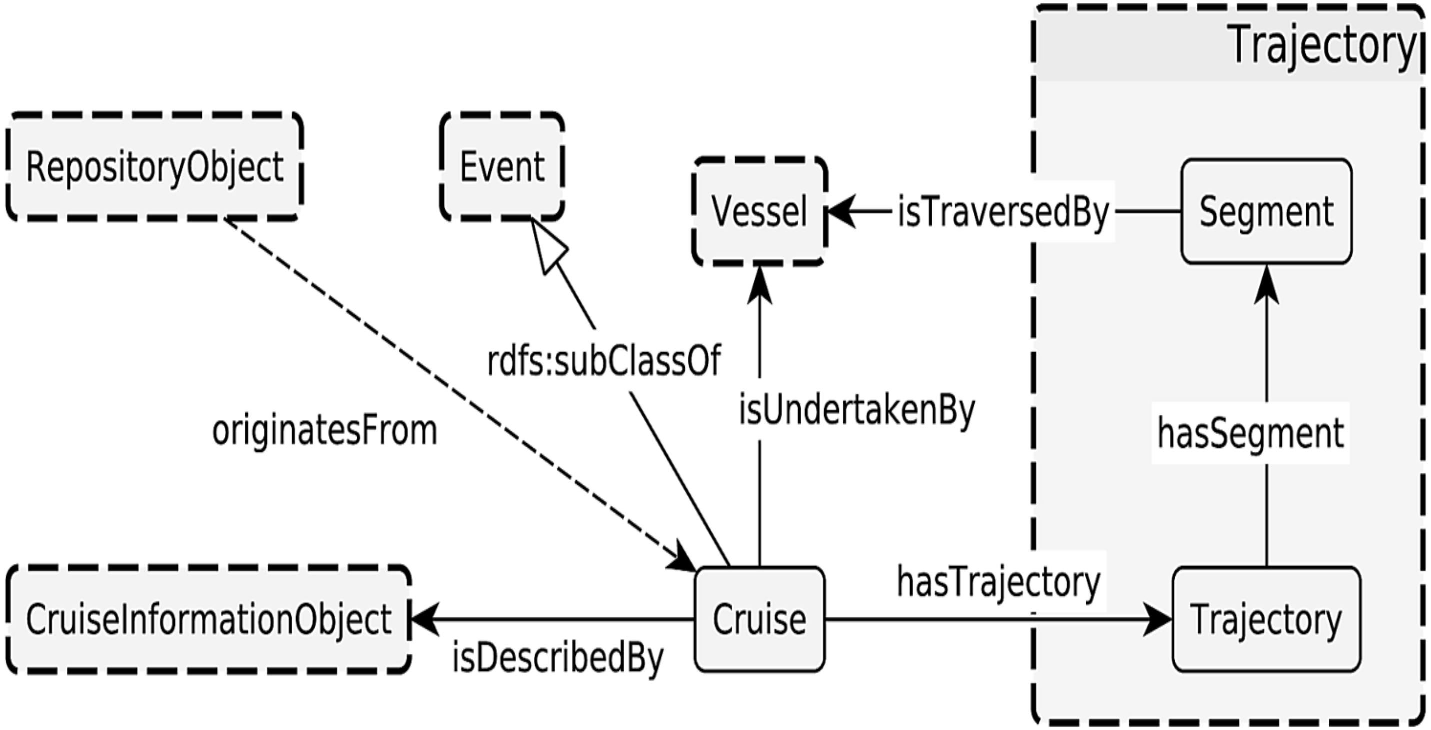
USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES



ENVO bridges domains with increased scope, semantic density, and interoperation.

An oceanographic cruise ODP

USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES



An ODP (ontology design pattern) is a reusable solution to a data modelling problem.

Agenda, including:

- Pattern Analysis and Machine Intelligence (PAMI) and Marine Ontologies
- Ontologies and Marine Robotics
- Ontology research trends review
- State-of-the-art of Ontologies in marine data classification and recognition
  - A framework for interactive visual analysis of heterogeneous marine data
  - PA, DC & DR identified through ontologies for marine data
  - Machine learning levels of visualization and their temporal perspective
- Conclusions

Data-driven Future in Marine Sciences

No interoperability

Relevance: semantic data, metadata standards

Not enough data

Relevance: particularly AI, NN, ML, frontier technologies

Few annotations in images

Relevance: monitoring of coastal seas, random forest applications

Use cases unpredictable

Relevance: cross-domain innovation applications

Addressing the interoperability

Marine Ontologies offer:

- A tool and method to assess the added value robotic technology brings into the marine environment (autonomous underwater vehicles (AUVs) or (ocean floor observation systems) OFOSs)
- The mechanism for describing sensors and sensor networks work in the context of Sensor Web applications
- A sustainable approach to harmonized data documentation
- Enables data re-use, data valorization, collaborative innovation, and to unambiguously set definitions and interconnect concepts in various field (more agile value chain interactions)

The main features provided by ontologies in support of PAMI

USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES

Ontology feature	Utility in PAMI
Classes and relations	When ontology reasoning is applied to sensor data, rdf:type will be connected to a class name of an ontology
Domain vocabulary	Ontologies provide a domain vocabulary that can be exploited to create a dense network of relationships among the entities, and serve software applications, and GIS
Metadata and descriptions	Biodiversity data, especially in marine domain, have database entities represented as ontologies where these last are primarily used for metadata that describe raw data providing contextual information
Axioms and formal declarations	Ontology axioms and applied reasoning on them are related to the recognition of object presence in a time interval

Contribution to Marine Data Potential

USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES

Marine ontologies interoperability Standardized data documentation

Increase in data exchange and integration More data use and re-use

Increase in data value

Data visualization potential, using semantic web axioms

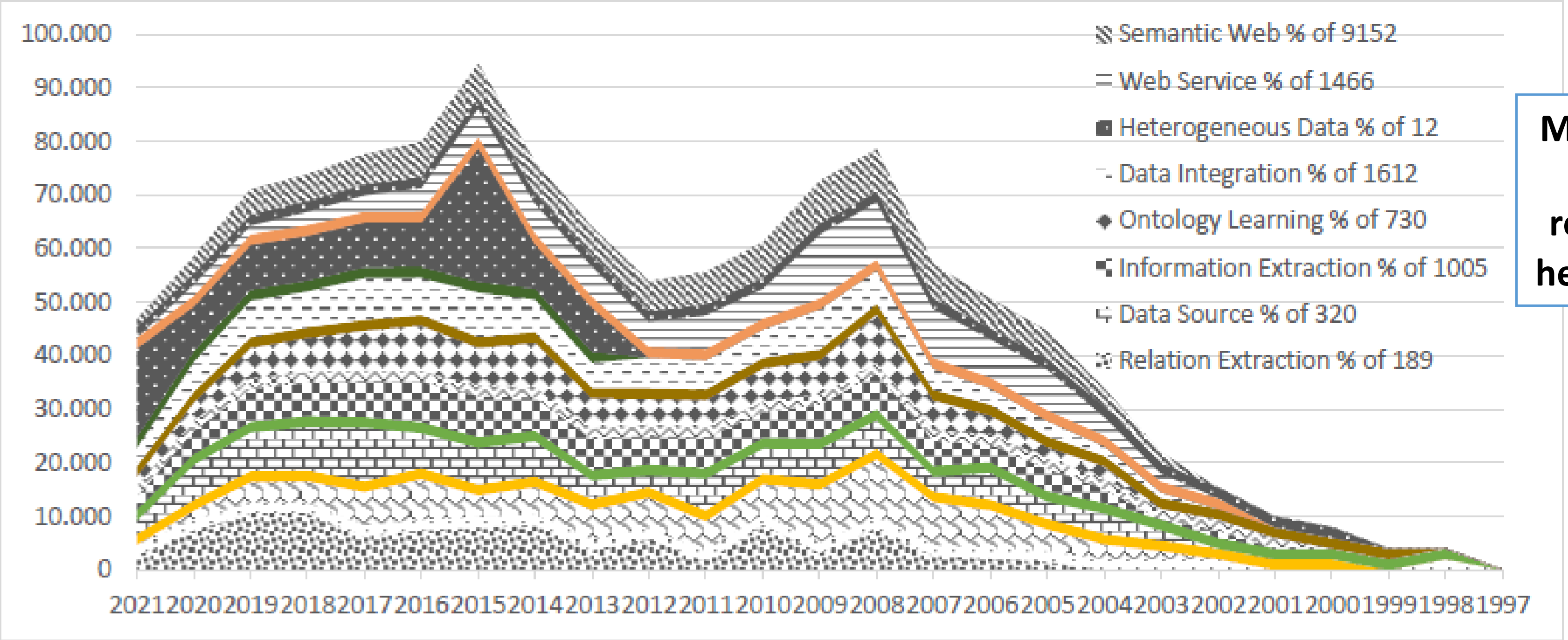
Increase in use cases

Marine robotics is a successful application / lesson learned

Ontology Research Trends

USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES

ONTOLOGY TOPICS (semantic web, web services)



MARINE ONTOLOGY (data source, relation extraction, heterogeneous data)

Contribution to semantic annotation

- Breakthrough annotation potential by making heterogeneous marine data accessible to ML
- Unlocking the potential of compositional definitions of a sequence of images
- Re-use of data across domains (compositional definitions, ontological certainty)
- Marine ontologies facilitate annotation of patterns using a multiple expert approach
- Highlighting ontological feedback to the domain of visualization (85.7% context detected)

Ontological contexts detected in the data sources analyzed

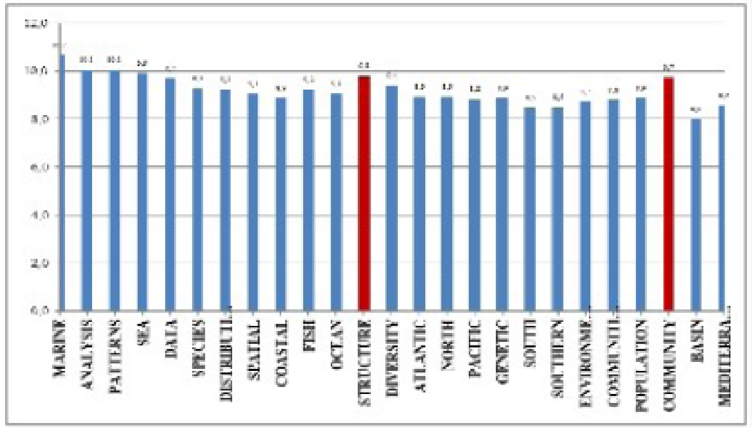
USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES

Num	Experimental Organism	A/B	DV	Ontology
1	Dynamena pumila	B	Y	Gene Ontology (GO) and KEGG pathway
2	Takifugu rubripes	A	Y	Gene Ontology (GO)
3	phytoplankton	A	Y	Gene Ontology (GO)
4	nd	A	Y	nd
5	Dreissena polymorpha	B	Y	Gene Ontology (GO)
6	Atlantic salmon	A	Y	Gene Ontology (GO) and UniProt Knowledgebase
7	Micromonas polaris; Pyramimonas tychotreta	B	Y	Gene Ontology (GO)
8	Crassostrea gigas	A	Y	Gene Ontology (GO)
9	Nd	B	Y	Genomic Standards Consortium's MIxS and Environment Ontology (ENVO); EMP Ontology (EMPO) of microbial environments
10	Chlamys farreri	A	Y	Gene Ontology (GO) and Eukaryotic Orthologous Groups (KOG) and Kyoto Encyclopedia of Genes and Genomes (KEGG)
11	Nd	B	Y	Protégé environment (ontology)
12	Nd	B	Y	Protégé environment (ontology)
13	Eucheuma denticulatum	A	Y	Gene ontology (GO)
14	48 species of freshwater and marine fish	A	Y	Gene ontology (GO)
15	Larimichthys crocea	A	Y	Gene ontology (GO)
16	Seriola lalandi	B	Y	Gene ontology (GO)
17	marine and FW sticklebacks	B	Y	Gene ontology (GO)
18	Genypterus chilensis	A	Y	Gene ontology (GO)
19	Ceriops	A	Y	Gene ontology (GO)
20	Human	A	Y	Gene ontology (GO)
21	Zostera muelleri	B	N	Gene ontology (GO)

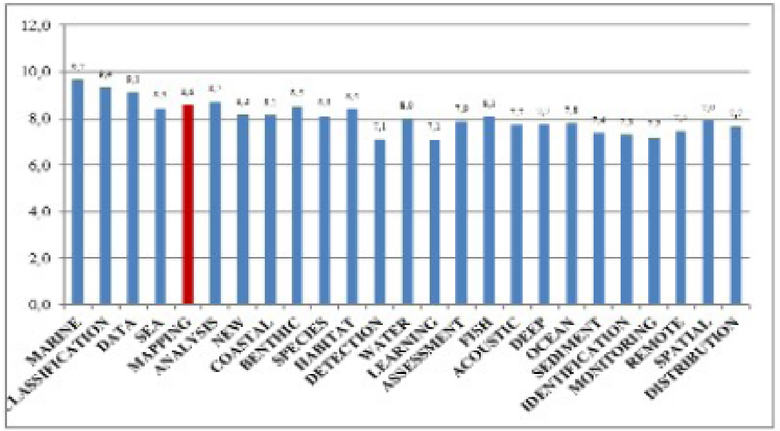
(A/B: applied/theoretical; DV: data visualization feedback)

A proof of principle shows needs for sensor data fusion

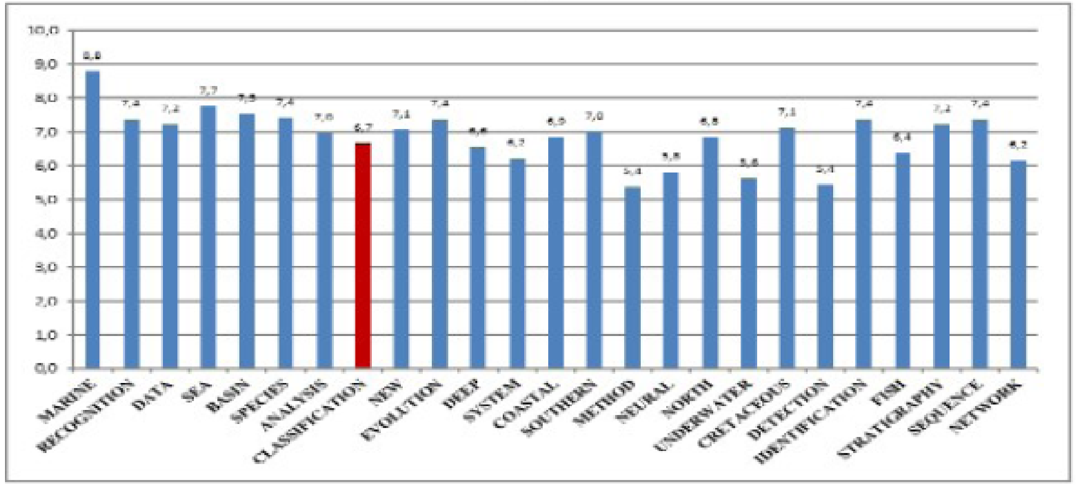
USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES



(a)



(b)



(c)

(a) (structure and communities cause a positive effect on modelling required to discriminate relevant from non-relevant images)

(b) (mapping is the main technique to classify marine data)

(c) (a visual recognition task mapping is ensured by a visual language with data classification)

Support to image visualization and exploitation on 2D, 3D, and motion imagery

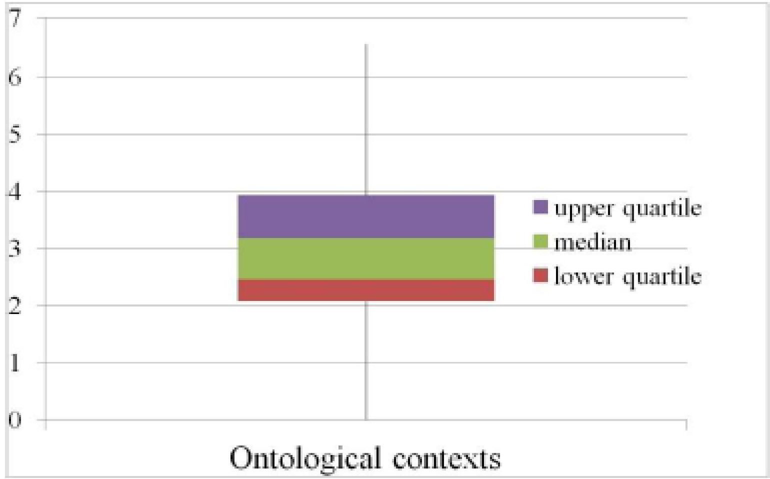
- ML attests its power on heterogeneous marine data:

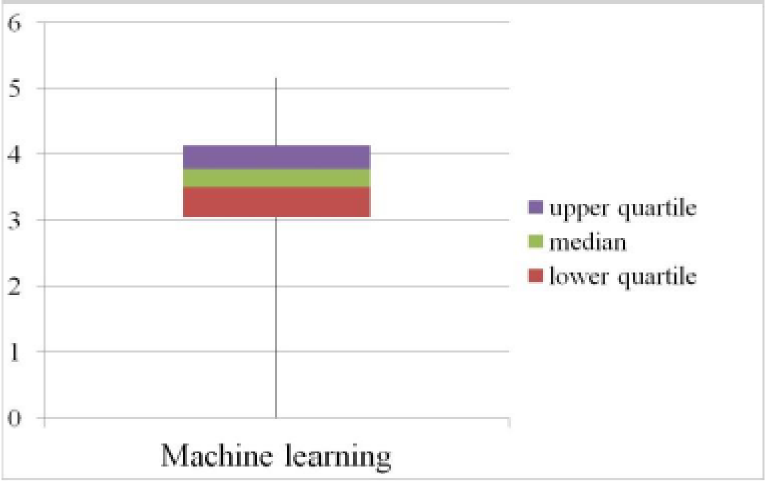
	analytics	association	case study	classification	comparaison	complex	correlation	development	information	mapping	model	monitor	Observation	pattern	prediction	recognition	regression	sensor	simulation	time series	other	SUM
Papers	1	1	4	70	7	3	1	4	2	16	2	12	2	13	13	7	1	12	3	2	16	210
disseminative				1					2			4		1						2	3	13
observational			1	11							2	2	2	1				12			4	33
analytical	1	1	3	34	7	3	1					6		10			1				3	70
model-developmental				3				4		16				1	13	7			3		4	92

- Best analytical and modelling for innovation from demonstrators

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	SUM
Papers	3	3	7	5	4	6	9	12	16	15	27	55	48	210
disseminative							1		1	1	1	5	4	13
observational	1		1			1		1			6	12	11	33
analytical	1	1	3	2	1	4	4	6	6	6	7	14	17	70
model-developmental	1	2	3	3	3	1	4	5	9	8	13	24	16	92

Data Visualization Scenarios





- the resulting lower quartile of 0.45 reached a best score for ML than for ontologies (0.38)
- ML decisions based outperformed ontology-driven coding for image classification
- in spite of ontology mapping for underwater IoT (IoUT) supports better interoperability protocols in the context of computer vision

Conclusions

This approach has led to accurate predictions of the level of visualization importance for the example of how should marine databases be represented :

- The aim is to characterize marine ontologies to select data visualization techniques
- Promote results in interoperability as supported by ontologies
- Align with further statistics coverage developments across pattern analysis applications
- Continue to support the machine learning techniques, as it is clear that deep learning is a core components of the wider computer vision task with marine data.

Towards a theoretical model for visualization with marine data

USE OF ONTOLOGIES FOR REPRESENTING DATABASE ENTITIES

- Beyond the GIS community, Common Marine Ontologies are set to contribute to identify predicting and moderating variables for information perception of visual data.
- This will lead to a multiplier effect of potential applications, related to the top current cognition research, and unlock considerable potential for underwater and robotic vehicles and FAIR services in the context of a digital space for oceanography.

The potential of ontologies for the empirical assessment of machine learning techniques in operational oceanography

Thank you very much for your attention!

Enrique Wulff, *Marine Science Institute of Andalusia, CSIC*