

The NERC Vocabulary Server and BODC Vocabulary Services

Gwenaëlle Moncoiffé
Vocabulary Management Group at BODC
National Oceanography Centre, Liverpool and Southampton

Presented at BODC 50th, 8 May 2019, Liverpool



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL



**British Oceanographic
Data Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC SCIENCE OF THE
ENVIRONMENT

Controlled Vocabularies

The semantic glue of distributed data systems

Used to capture essential elements of information about data

- Parameters
- Units
- Instruments
- Platforms
- Sea Areas
- Disciplines and many more

Enable efficient discovery and usage of the data without reference to the source

Support harmonisation and aggregation of data from multiple sources

Optimise re-usability of individual data atoms (if used correctly)



Innovation in vocab services – Success factors

Timely adoption of web technology and international standards

Use of semantic modelling for variable and parameter labelling

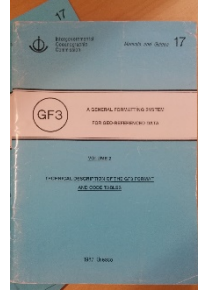
Rigorous content and technical governance and fast response time

Strong national and international collaborations



An historical perspective

1980s: IODE GETADE GF3



IODE Group of Experts on the Technical Aspects of Data Exchange - An early attempt at interoperability

1990s: The “Dark Ages”

Funding crisis led to dwindling governance and vocabulary abuse

The “Renaissance”
Building pan-European metadata catalogues underpinned by controlled vocabularies

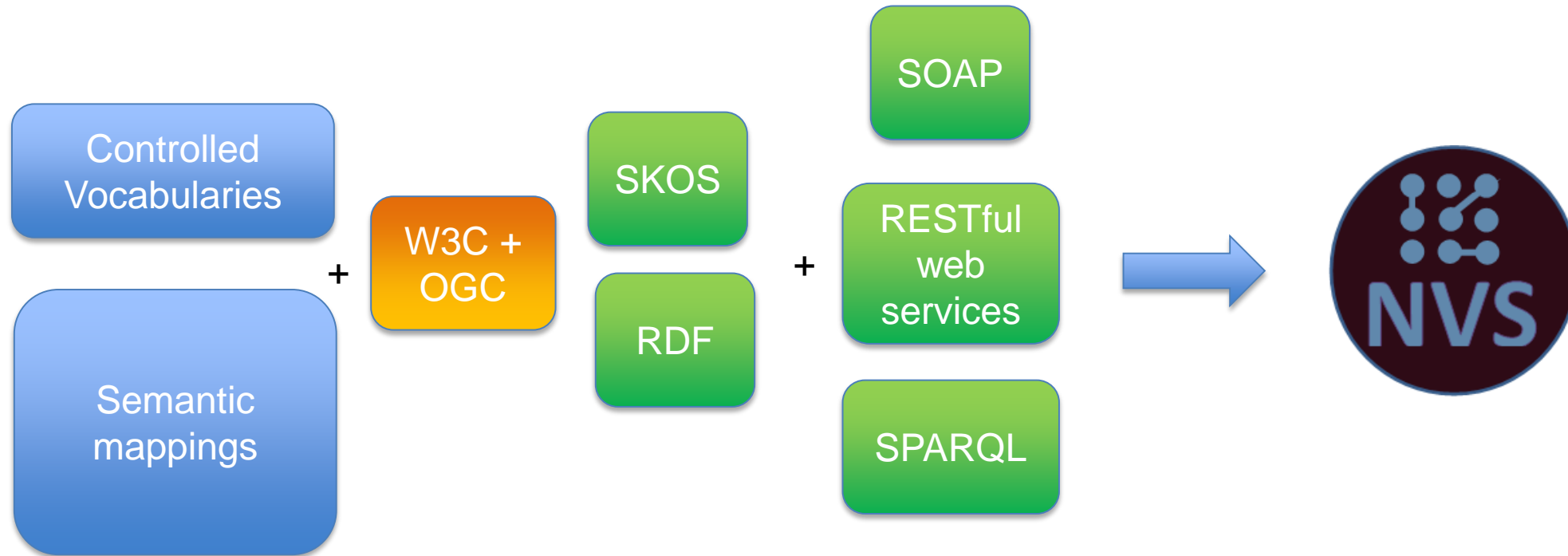
2002-2005: SEA-SEARCH



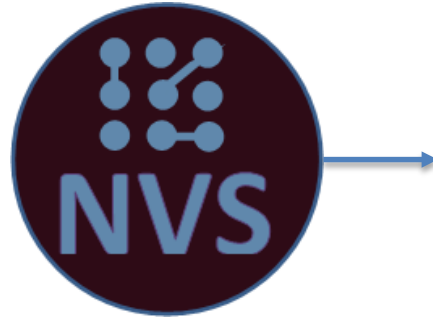
2005-2012: NERC Vocab Server



Timely adoption of web technology and standards



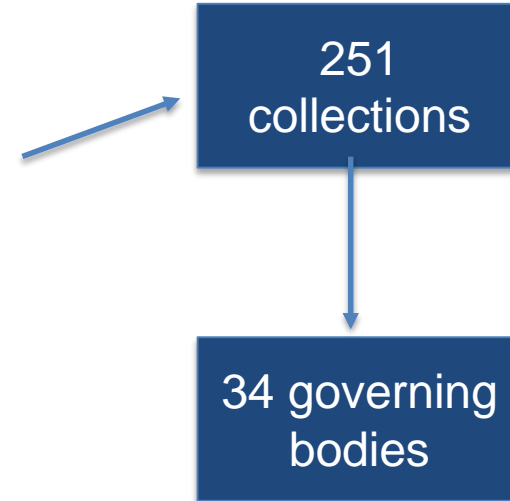
NERC Vocabulary Server (NVS)



Resources > Vocabularies > Vocabulary search

Options Collections library

A01	A02	A03	A04	A05	B02	B03	B04	B05	B06	B07	B09	B11	B12	B20
B21	B22	B39	B75	B76	C00	C10	C16	C17	C18	C19	C30	C31	C32	C33
C34	C35	C36	C37	C38	C39	C40	C41	C43	C45	C46	C47	C48	C59	C60
C61	C62	C64	C67	C71	C72	C75	C77	C86	C87	C88	C89	C96	C97	C98
D01	E01	E02	F02	G01	G02	G03	G04	G05	G06	G07	G08	G09	G10	G11
G12	G13	G14	G15	G17	G18	G20	G21	G22	G23	G25	G26	G28	G29	G30
GBX	GGB	GG3	GS1	GS2	GS3	GS4	GS5	GS6	GS8	GS9	GSA	GSB	GSC	GXM
H01	H02	H03	H04	H05	HA2	I01	I02	I03	I10	I11	I12	I13	I14	I15
L02	L03	L04	L05	L06	L07	L08	L10	L11	L12	L13	L14	L15	L18	L19
L20	L21	L22	L23	L24	L26	L27	L30	L31	L33	L34	L35	L36	L37	L38
M01	M03	M04	M05	M06	M09	M10	M11	M12	M13	M14	M15	M16	M17	M18
M20	M21	M22	M23	MVB	N01	N02	N03	N04	N05	OG1	P01	P02	P03	P04
P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P17	P18	P19	P20
P21	P22	P23	P24	P25	P26	P27	P28	P29	P30	P35	P36	P37	P38	P64
Q01	S01	S02	S03	S04	S05	S06	S07	S09	S10	S11	S12	S13	S14	S15
S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	S29	S30	V12	V22
V23	W01	W02	W03	W04	W05	W06	W07	W08	W09	W10				



BODC – 74 collections
 SDN – 40
 SWE Marine Profiles – 8
 SeaVox - 5

Vocabulary Editor Tool

VocabEditor Client (version 1.0)

List options

You have been authorised as an editor on the list(s) presented below. Please proceed with one of the following options:

- **Mappings – Bulk Upload:** Click Mappings – bulk upload button to upload mappings
- **Edit – single List:** Select a list to edit
 1. Select the list you require
 2. Click Edit single list
 3. Select one of the options

Please note — for security reasons, if you fail to interact for a period of more than 30 minutes your session will be closed. Any updates submitted prior to closing the session will be queued in the pending updates holding area and the changes will occur overnight during the scheduled vocabulary list update.

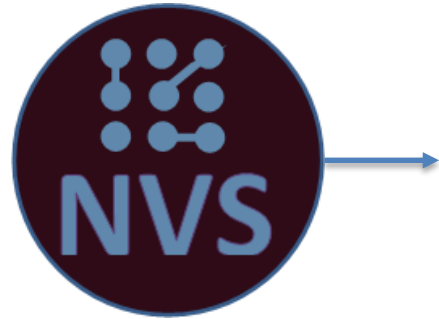
Options

Key	List ID	Short name	Definition	Version	Modified
<input type="radio"/>	M09	MEDIN EV method types	Terms used to classify techniques used to assess the socio-economic value of an ecosystem.	1	2013/11/15:02:00:10
<input type="radio"/>	M10	MEDIN EV methods	Terms used to describe techniques used to assess the socio-economic value of an ecosystem.	1	2013/11/15:02:00:10
<input type="radio"/>			Terms used to classify the conditions and processes through which natural ecosystems sustain		

https://www.bodc.ac.uk/resources/vocabularies/vocabulary_editor/



SeaDataNet Common Vocabularies



SeaDataNet PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

ABOUT US METADATA DATA ACCESS STANDARDS

Home > Standards > Common Vocabularies

COMMON VOCABULARIES

seadatanet.maris2.nl/v_bodc_vocab_v2/welcome.asp

SeaDataNet PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT BODC VOCAB LIBRARY

BODC WEBSERVICES V2 (LIBRARIES) CL12

Library	Thesaurus	Title	Alt Title	Version	Members	Modified
C16		SeaDataNet sea areas	SDN sea areas	9	127	11/7/2012 2:00:06 AM
C17		ICES Platform Codes	ICES Platforms	853	11962	5/2/2019 3:00:05 AM
C19		SeaVoX salt and fresh water body gazetteer	SeaVoX water bodies	17	263	2/21/2018 2:00:03 AM
C32		International Standards Organisation countries	ISO countries	8	281	4/17/2019 3:00:07 AM
C34		Activity purpose categories	Purpose categories	4	22	8/27/2011 3:00:05 AM
C35		European Nature Information System Level 3 Habitats	EUNIS3 Habitats	1	56	2/19/2010 2:01:37 AM
C36		Monitoring activity legislative drivers	Monitoring drivers	9	92	10/24/2018 3:00:04 AM
C37		Ten-degree Marsden Squares	Marsden-10	3	612	1/9/2009 2:00:05 AM
C38		SeaDataNet Ports Gazetteer	SeaDataNet Ports	65	4958	2/1/2019 2:00:05 AM
C39		World Meteorological Organisation sea states	WMO sea states	1	10	9/30/2009 3:01:08 AM
C45		Marine Strategy Framework Directive descriptors 2010/477/EU	MSFD descriptors 2010	3	11	2/25/2017 2:00:02 AM

A subset of 111 collections from 15 governing bodies

Use of semantic modelling for complex concepts

Application: Standardisation of parameter labelling



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL



**British Oceanographic
Data Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC SCIENCE OF THE
ENVIRONMENT

Why is this needed?

The meaning of isolated words is often context dependent

e.g. The depth of an object?

Does this relate to the physical dimension of the object or does it relate to its location in the sea or a lake?

The human brain can most of the time resolve the ambiguity if the context is known

For example if the object is a CTD and the CTD is deployed from a ship then the depth of the CTD is likely to be its location in the water column

If the object is an incubation vessel for an on-deck incubation then it is likely to be a physical dimension of the incubation container.



BODC Parameter Usage Vocabulary or P01 collection

opaque 8-byte identifier

Natural Environment Research Council [GB] | https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/P01/

NVS editor
 NVS vocabulary builder
 BODC parameter codes
 SeaVoX

Search text: Vocabulary: [advanced options](#)

Found 12533 records | Show (11 - 20) | [First](#) [Prev](#) | [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [Next](#) [Last](#) [download results](#) | [start again](#)

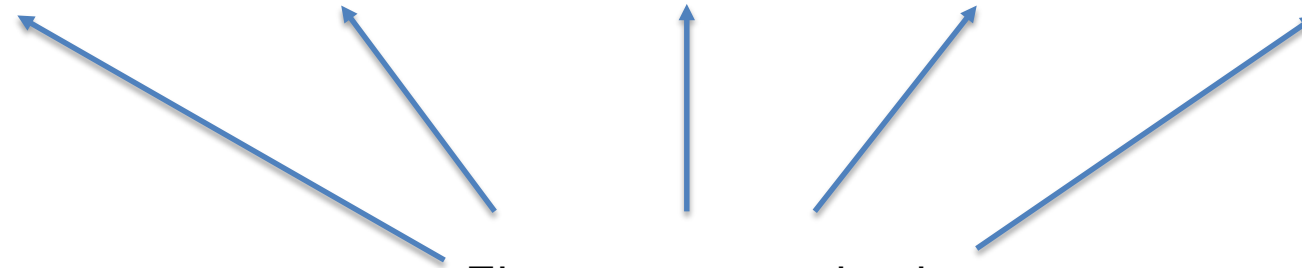
Identifier	PrefLabel	Definition	Date
FA63GCP1	Concentration of methyl hexatriaconta-7E,14E,21E-trienoate {C36:3O methyl ester} per unit volume of the water body [particulate >GF/F phase] by filtration and gas chromatography-mass spectrometry	Gas chromatography mass spectrometry (GF/F filtered)	2018-02-27
ME82GCP1	Concentration of octatriaconta-16E,23E-dien-2-one {C38:2 methyl ketone} per unit volume of the water body [particulate >GF/F phase] by filtration and gas chromatography-mass spectrometry	Gas chromatography mass spectrometry (GF/F filtered)	2018-02-27
ME83GCP1	Concentration of octatriaconta-9E,16E,23E-trien-2-one {C38:3 methyl ketone} per unit volume of the water body [particulate >GF/F phase] by filtration and gas chromatography-mass spectrometry	Gas chromatography mass spectrometry (GF/F filtered)	2018-02-27
ESTSED14	Concentration of methyl 2-hydroxy heptacosanoate per unit dry weight of sediment	specified organic compound per unit mass of dry sediment.	2018-02-27
ESTSED15	Concentration of methyl 2-hydroxy docosanoate {behenic acid methyl ester CAS 929-77-1} per unit dry weight of sediment	The amount (mass or moles) of the specified organic compound per unit mass of dry sediment.	2018-02-27
	Concentration of methyl 2-hydroxy tetracosanoate {CAS 2433-95-6} per unit dry weight	The amount (mass or moles) of the	2018-

Concentration of octatriaconta-9E,16E,23E-trien-2-one {C38:3 methyl ketone} per unit volume of the water body [particulate >GF/F phase] by filtration and gas chromatography-mass spectrometry

Structured label based on a semantic model

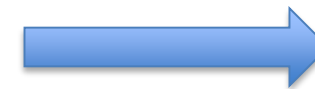
Basic conceptualisation of an observable property

the **PROPERTY** of an **OBJECT** in **RELATION** to a **MATRIX** by a **METHOD**



Elements constrained
against controlled
vocabularies

8-byte PID code



P01 Facet Search Tool at MARIS

seadatanet.maris2.nl/bandit/browse_step.php



PAN-EUROPEAN INFRASTRUCTURE FOR
OCEAN & MARINE DATA MANAGEMENT

P01 VOCABULARY - FACET SEARCH ON SEMANTIC COMPONENTS

The P01 Parameter Usage Vocabulary is based on a semantic model. This model uses a defined set of controlled vocabularies (the semantic components). The Facet Search below facilitates you to search for specific existing P01 terms using components for drilling down.

Are you missing specific P01 terms in the vocabulary, then you can compose and submit new terms for review and uptake using the [P01 Vocabulary Builder tool](#).

Filter Search

FREE SEARCH

Input string

MEASUREMENT PROPERTY (506)

[Concentration](#) (17330)

[Abundance](#) (7772)

[Absorbance](#) (2825)

[Ash-free dry weight biom...](#) (1608)

[Lipid-normalised concent...](#) (1332)

CHEMICAL SUBSTANCE (527)

[total mercury](#) (488)

[cadmium](#) (439)

[lead](#) (394)

[zinc](#) (374)

[copper](#) (347)

MATRICES (526)

[biota](#) (13487)

Found 42340 Show (1-25) < Prev [Next](#) >

[DECOMPOSED-EXPORT](#) [EXPORT](#)

Conceptid (42340)	Preflabel
D2930490	Concentration of 2,2',3,3',4,4'-hexachlorobiphenyl {PCB128 CAS 38380-07-3} per unit dry weight of biota {Zoarces viviparus (ITIS: 165324: WoRMS 127123) [Subcomponent: liver]}
D2930491	Concentration of Arochlor 1260 {CAS 11096-82-5} per unit dry weight of biota {Gadus morhua (ITIS: 164712: WoRMS 126436) [Subcomponent: liver]}
D2930492	Concentration of tributyltin cation {tributylstannyl TBT+ CAS 36643-28-4} per unit dry weight of biota {Boops boops (ITIS: 169218: WoRMS 127047) [Subcomponent: liver]}
D2930493	Concentration of 4,4'-dichlorodiphenyltrichloroethane {p,p'-DDT CAS 50-29-3} per unit dry weight of biota {Perca fluviatilis (ITIS: 168470: WoRMS 151353) [Subcomponent: liver]}
D2930494	Concentration of Arochlor 1260 {CAS 11096-82-5} per unit dry weight of biota {Zoarces viviparus (ITIS: 165324: WoRMS 127123) [Subcomponent: liver]}
D2930495	Concentration of 2,2',4,4',5,5'-hexachlorobiphenyl {PCB153 CAS 35065-27-1} per unit dry weight of biota {Patella (WoRMS 138312) [Subcomponent: flesh]}
D2930496	Concentration of 4,4'-dichlorodiphenyldichloroethylene {p,p'-DDE CAS 72-55-9} per unit dry weight of biota {Boops boops (ITIS: 169218: WoRMS 127047) [Subcomponent: muscle tissue]}
D2930497	Concentration of 2,2',3,4,4',5,5'-heptachlorobiphenyl {PCB180 CAS 35065-29-3} per unit dry weight of biota {Mya arenaria (ITIS: 81692: WoRMS 140430) [Subcomponent: flesh]}



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL



British Oceanographic
Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC SCIENCE OF THE
ENVIRONMENT

P01 Vocab builder at BODC

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Home Contact us Glossary Site map search BODC

My account Register Log in

Semantic Model

Some of BODC's vocabulary semantic model. The semantic model organises them into a structure that helps maintain a consistent look and feel.

The VOCAB BUILDER tool allows users to create new terms and users can also log in and submit terms.

The following vocabularies are available:

- Chemical entity parameter
- Biological entity parameter
- Physical entity and other parameter
- Measurement matrix concept
- Biological entity concepts (SeaVoX)

To submit new terms please use the VOCAB BUILDER tool.

Inventories

- Vocabularies
- NVS search tool
- NVS editor
- NVS vocabulary builder
- BODC parameter codes
- SeaVoX

Delivery formats

- Products
- Help and hints
- Portals and links
- Search

Found 3714 exact matches

- Select a measurement property
- Select a statistical parameter
- Select a physical entity (if applicable)
- Select a measurement-matrix relationship
- Select a matrix
- Select a sample preparation (if applicable)
- Select an analytical method (if applicable)
- Select a post-analysis processing (if applicable)

[Return to index](#)

[Login to enable new concept creation](#)

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

Home Contact us Glossary Site map search BODC

My account Register Log in

P01 Physical Entity and Other Parameter Code Builder

Preferred label

Found 6077 matches

- Select a measurement property
- Select a statistical parameter
- Select a physical entity (if applicable)
- Select a measurement-matrix relationship
- Select a matrix
- Select a sample preparation (if applicable)
- Select an analytical method (if applicable)
- Select a post-analysis processing (if applicable)

[Return to index](#)

[Login to enable new concept creation](#)

Growth and managing increasing demand



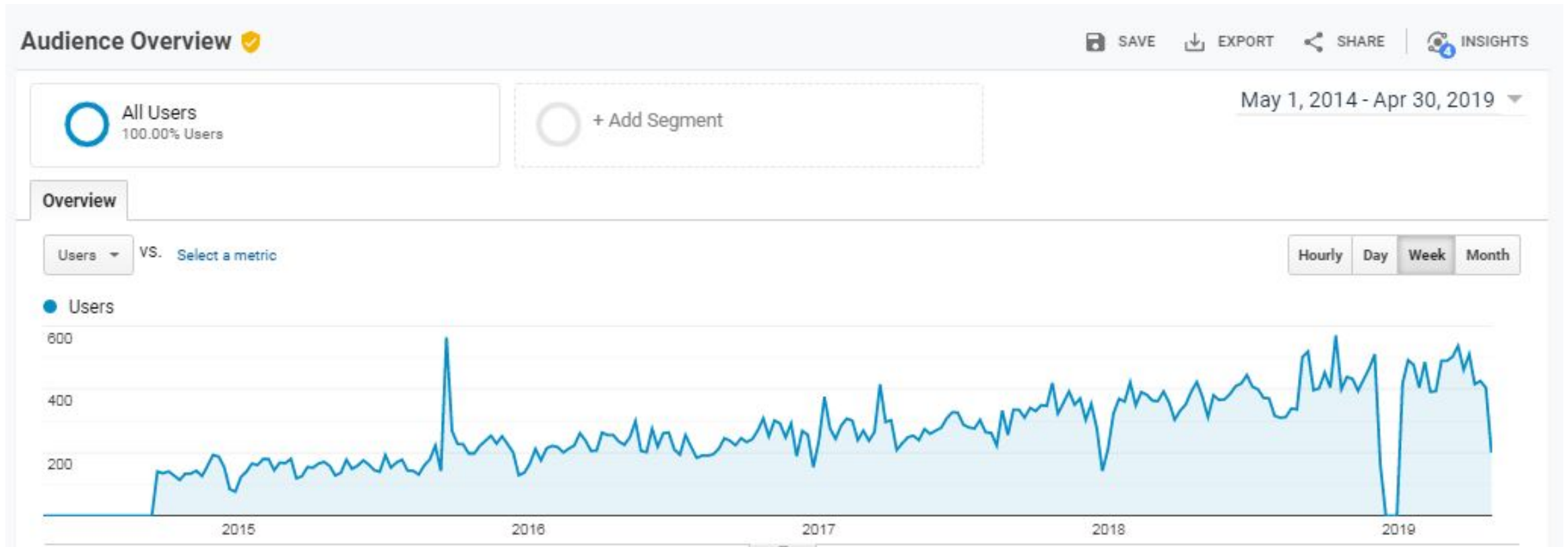
**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL



**British Oceanographic
Data Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

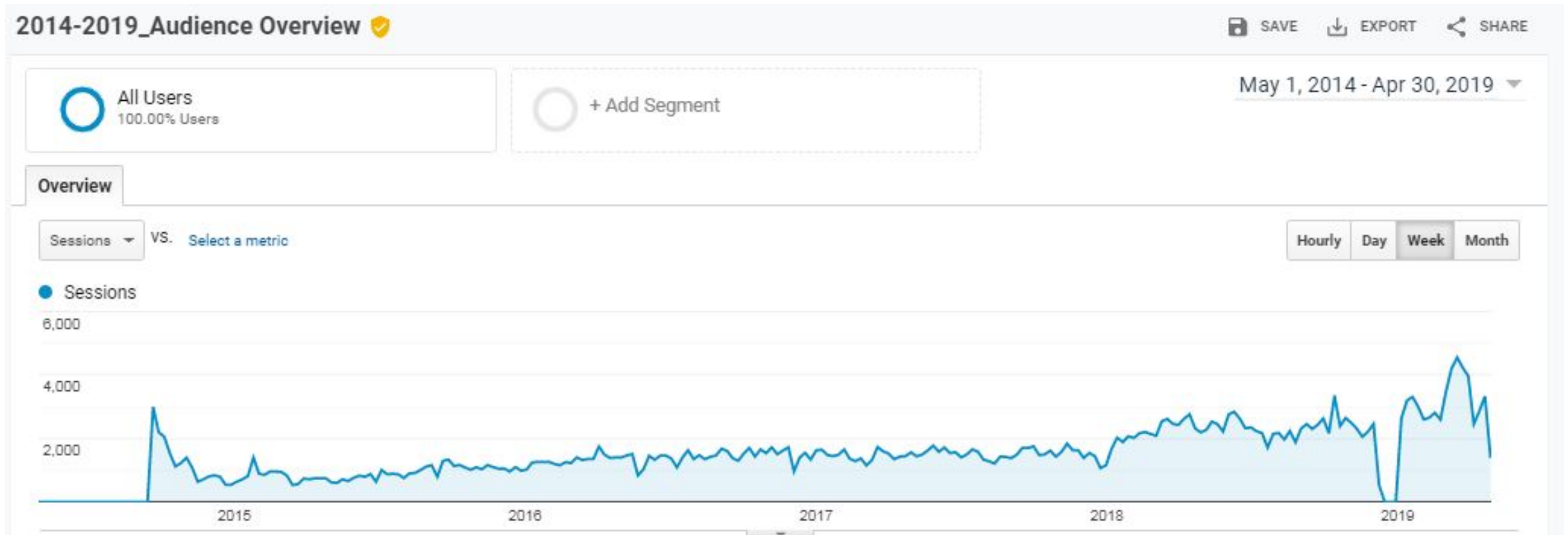
NERC SCIENCE OF THE
ENVIRONMENT

Number of NVS users per week since 2014



Number of unique NVS users per week since 01 May 2014

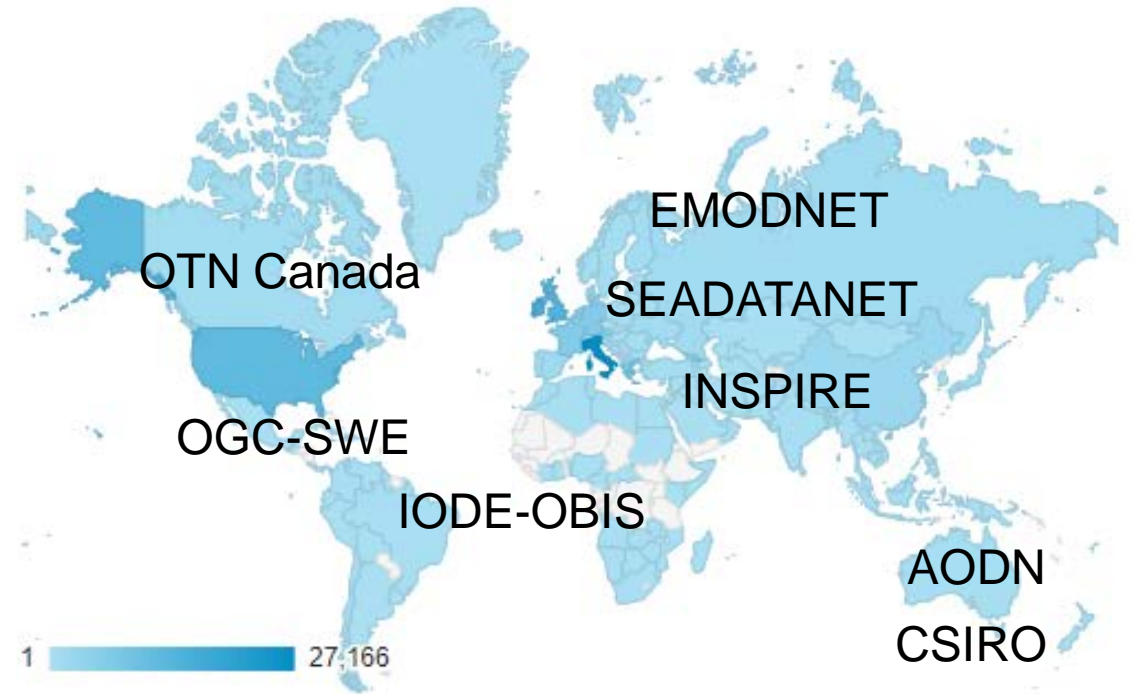
Number of NVS sessions per week since 2014



Number of NVS sessions per week since 01 May 2014

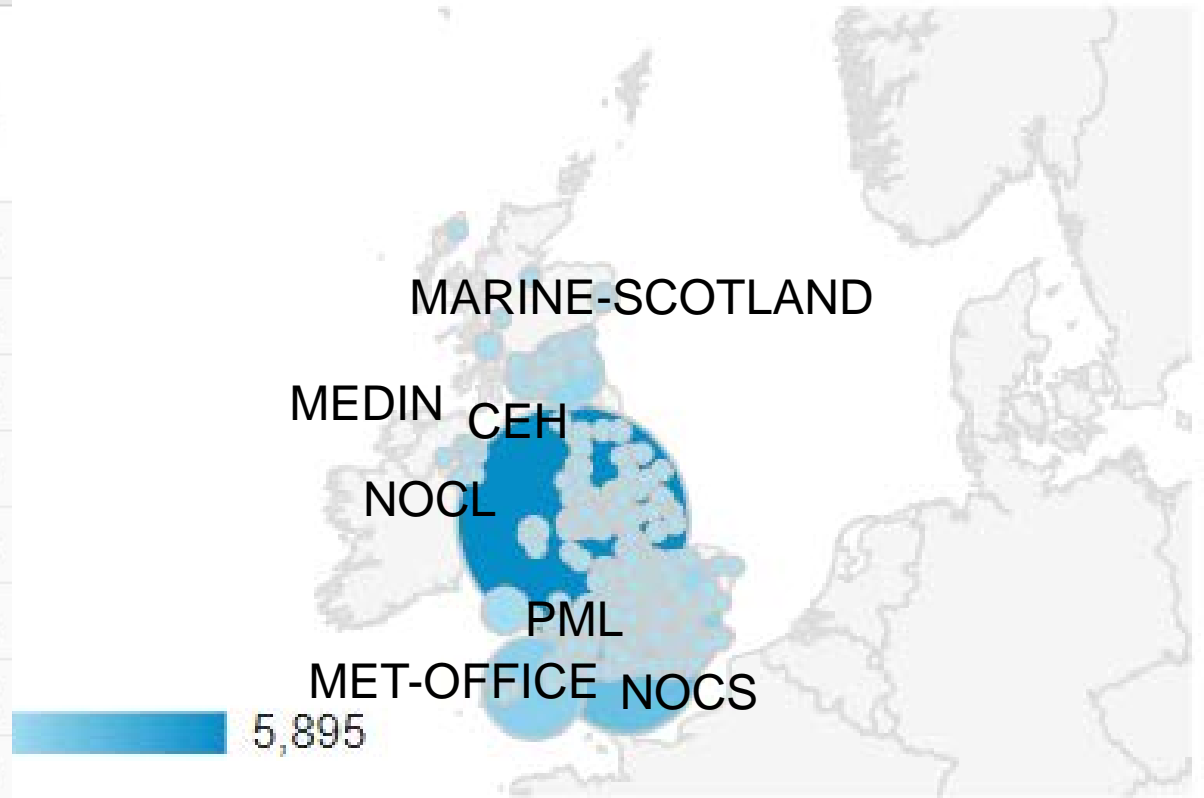
NVS International Impact

Country ?	Acquisition		
	Users ?	New Users ?	Sessions ? ↓
	15,971 % of Total: 100.00% (15,971)	15,502 % of Total: 100.01% (15,501)	138,682 % of Total: 100.00% (138,682)
1. Italy OGS-ITALY	820 (5.10%)	779 (5.03%)	27,166 (19.59%)
2. Ireland MI	300 (1.87%)	291 (1.88%)	16,082 (11.60%)
3. United Kingdom	2,900 (18.04%)	2,844 (18.35%)	14,015 (10.11%)
4. United States BCO-DMO	3,048 (18.97%)	3,002 (19.37%)	12,209 (8.80%)
5. Greece HCMR	155 (0.96%)	146 (0.94%)	8,195 (5.91%)
6. Austria EnvThes/LTER	66 (0.41%)	65 (0.42%)	7,086 (5.11%)
7. Georgia	17 (0.11%)	16 (0.10%)	6,885 (4.96%)
8. Germany MAX-PLANCK	688 (4.28%)	666 (4.30%)	6,478 (4.67%)
9. France IFREMER	744 (4.63%)	649 (4.19%)	6,207 (4.48%)
10. Netherlands MARIS	424 (2.64%)	404 (2.61%)	5,219 (3.76%)



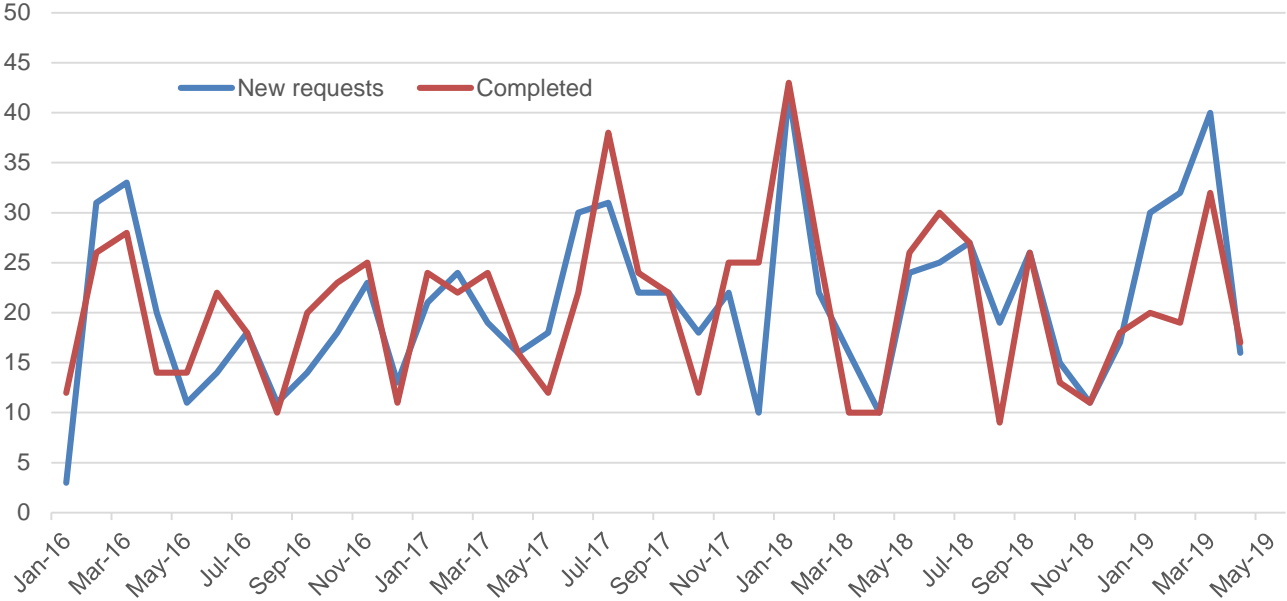
NVS usage in the UK

City ?	Users ?	New Users ?	Sessions ? ↓
	2,900 % of Total: 18.16% (15,971)	2,844 % of Total: 18.35% (15,501)	14,015 % of Total: 10.11% (138,682)
1. Liverpool	1,024 (34.74%)	1,011 (35.55%)	5,895 (42.06%)
2. Southampton	435 (14.76%)	431 (15.15%)	2,264 (16.15%)
3. London	362 (12.28%)	348 (12.24%)	1,050 (7.49%)
4. Plymouth	77 (2.61%)	74 (2.60%)	1,047 (7.47%)
5. Edinburgh	115 (3.90%)	107 (3.76%)	596 (4.25%)
6. Cambridge	42 (1.42%)	42 (1.48%)	522 (3.72%)
7. Harwell	93 (3.15%)	92 (3.23%)	259 (1.85%)
8. Corby	12 (0.41%)	9 (0.32%)	226 (1.61%)
9. Haverfordwest	17 (0.58%)	14 (0.49%)	200 (1.43%)

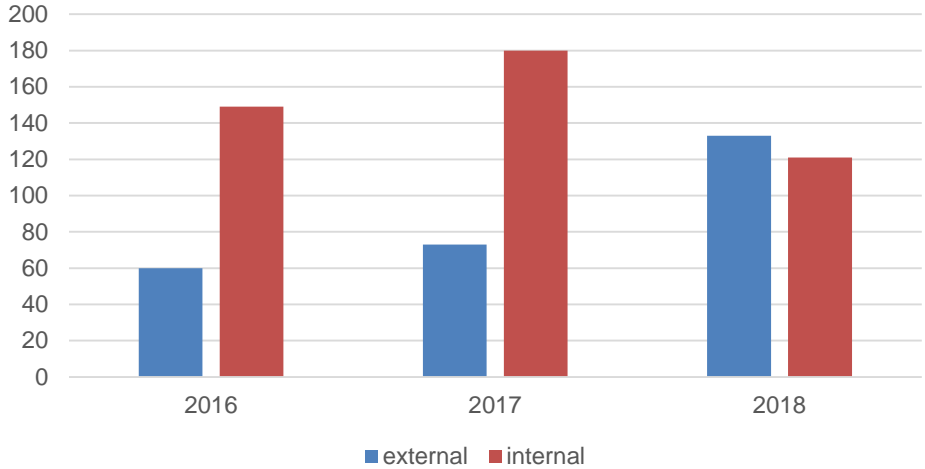


Manual Requests to the Vocab Team

BODC Vocab Helpdesk - Number of Requests Per Month

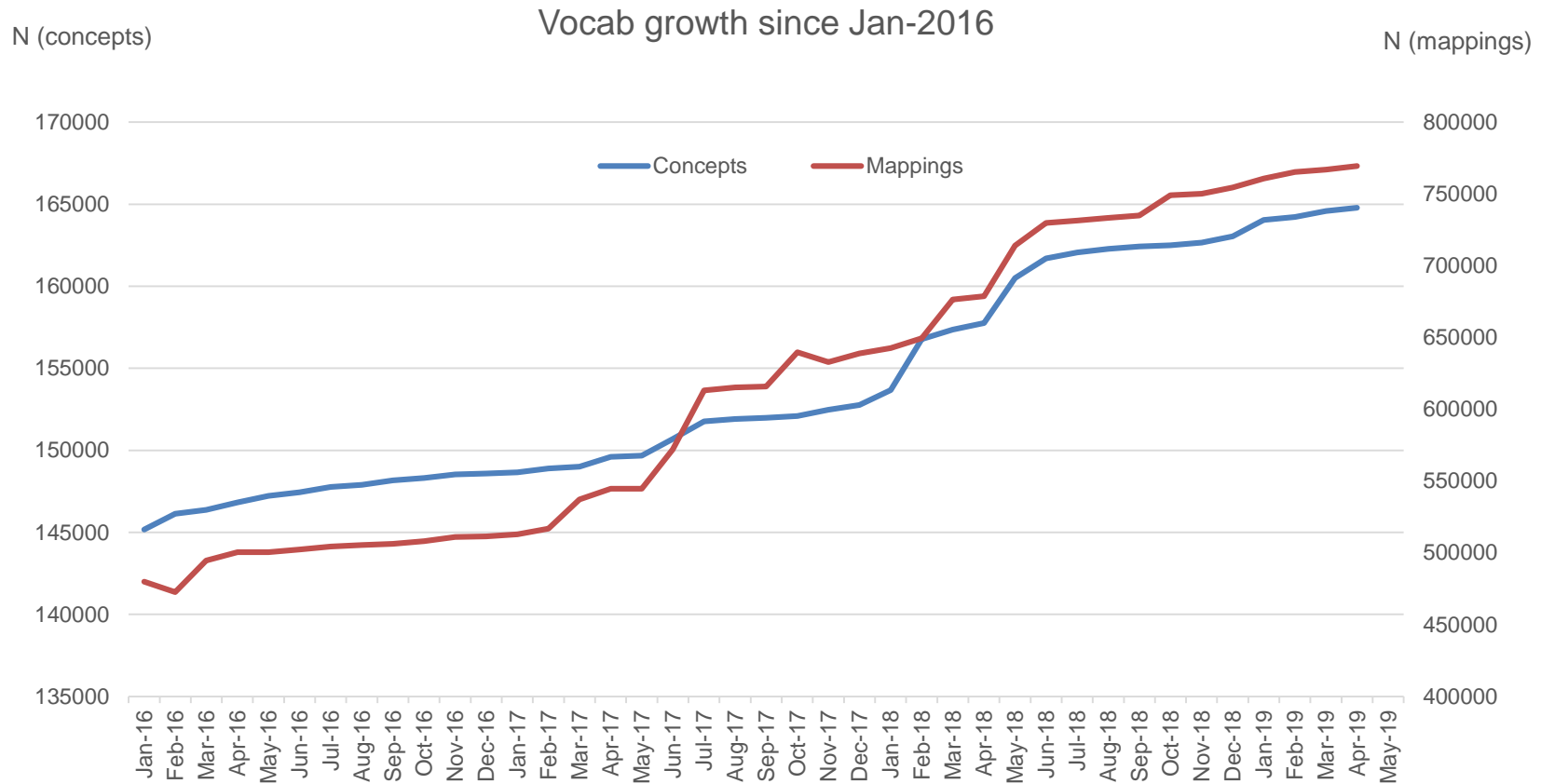


Requests for new concept creation

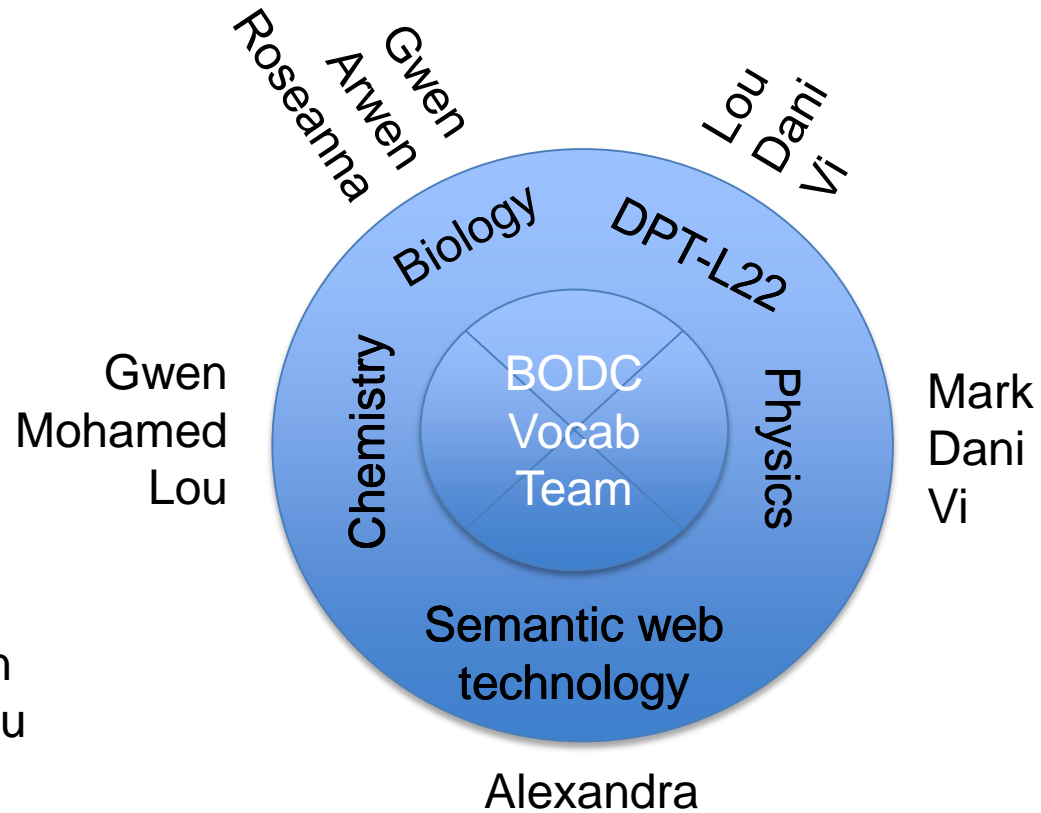


Number of requests received and completed since 01 January 2016

Growth of content - concepts and mappings



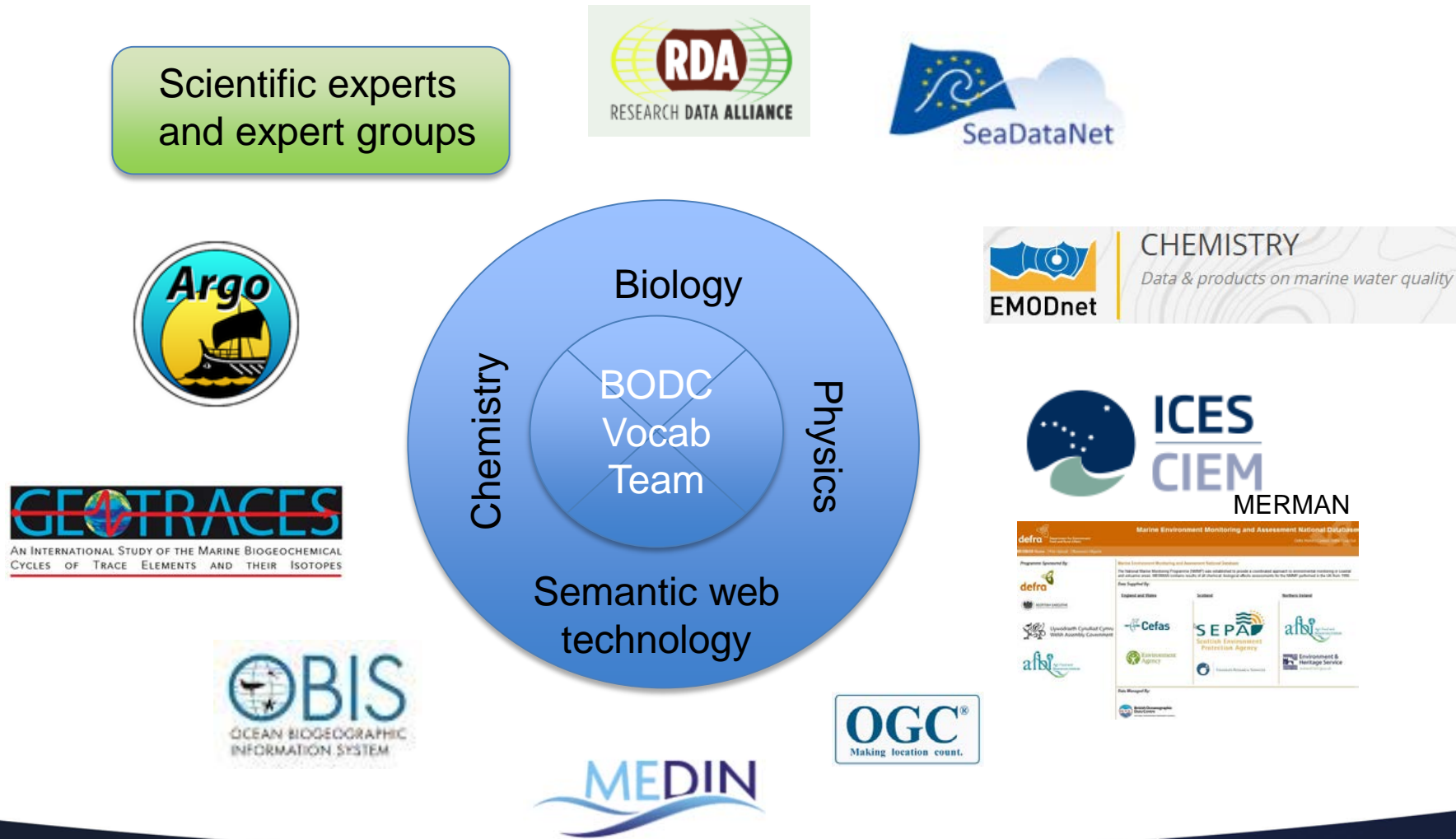
Vocab Management Team



Main gatekeeper – Gwen
Deputy Gatekeeper – Lou
Consultant - Roy

+ developing a network of external collaborators

Vocab Management Team – connections



Developing collaborations on Vocabulary Work

RDA Task Group on the harmonisation of parameter descriptions

- international and multidisciplinary working group aiming at developing a common semantic model for observable property descriptions

BCO-DMO, MI and ENVO

- work on improving workflow between vocabulary users (eg data managers), vocabulary creators, and ontology developers for application to the management of parameter and variable names, instruments identification and measurement units

MI and ICES

- To build some degree of interoperability between ICES vocabs and the NVS - focus on parameters, chemical substances, and matrices

CEH, BGS, PDC(BAS), CEDA(BADC)

- Will explore areas of collaboration between the 4 data centres for vocab services and identify requirements to support NERC Data Centre Integration



Thank you

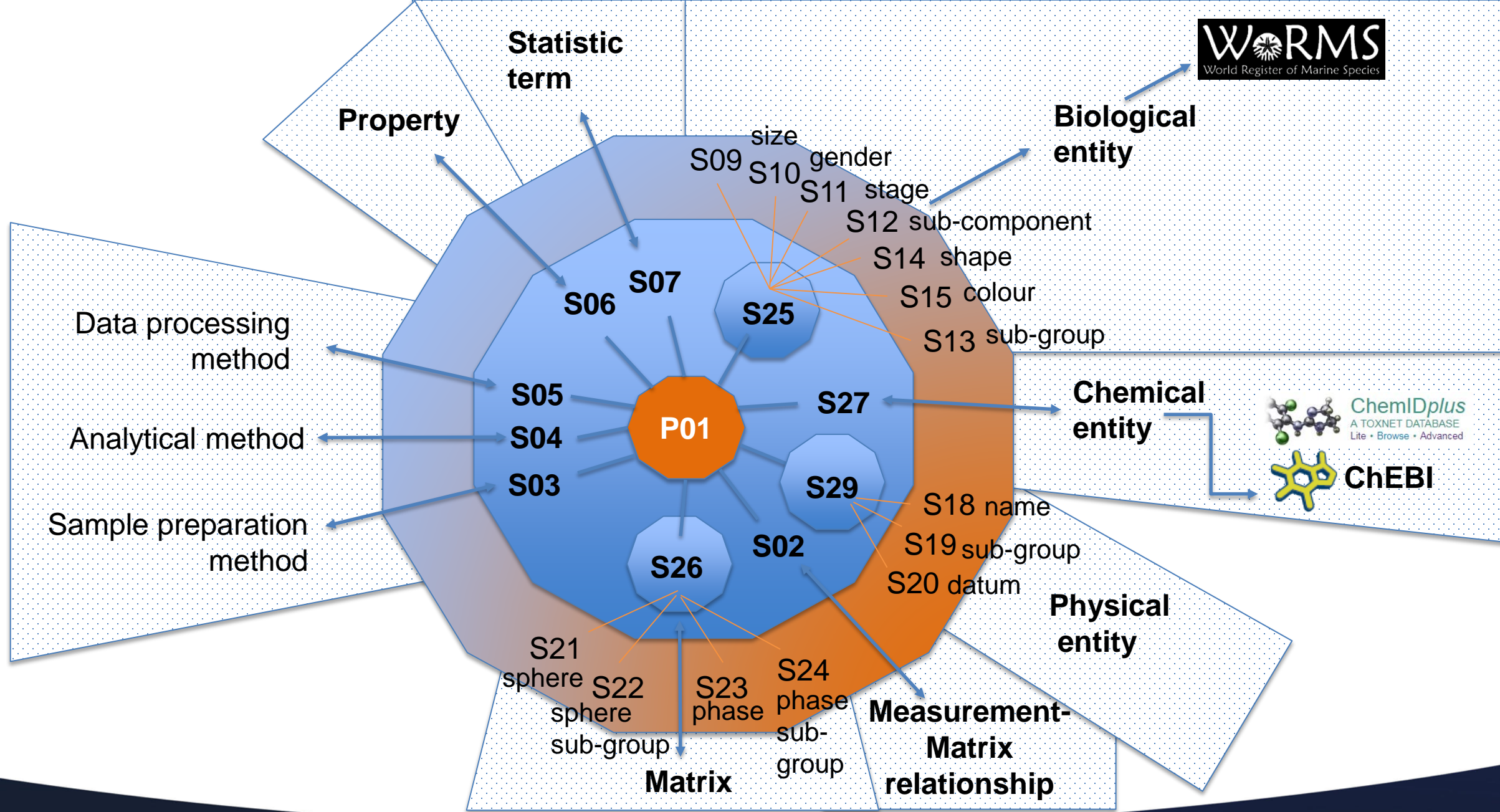


**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL



**British Oceanographic
Data Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC SCIENCE OF THE
ENVIRONMENT



The P01 PUV and its semantic model

- Deliver a practical solution for data mark-up in data files and databases
- Able to deal with the complexity with minimum loss of essential information
- New codes can be created fairly easily (by BODC staff or users who understand the model) by re-using existing concepts or creating new ones.
- It allows the capture of fine-grain information about a data value in a single header field.
- Annotating data with controlled vocabularies is time-consuming and standardized fine-grain annotation ensures optimum re-use of the data because aggregation is always an option.
- It will become particularly powerful when coupled with more formal or higher level ontologies and terminology resources like O&M/SOSA, ENVO or schema.org, and authoritative thesauri
- Data tagged with P01 codes will benefit from this
- But it does require strong gatekeeping and governance

