



SeaDataCloud

Technical implementation plan for next period

Dick MA Schaap – SeaDataCloud Technical Coordinator

SDC Review, Brussels – Belgium, 6 December 2018
sdn-userdesk@seadatanet.org – www.seadatanet.org

Technical development lines

- Upgrading of CDI service
- Developing the Virtual Research Environment (VRE)
- Implementing Linked Data concept for directories
- Maintaining and upgrading vocabularies
- Handling new data types
- SWE ingestion service for NRT and RT data sets
- Brokerage service for third party portals
- Publishing service for WP11 data products

Upgrading of CDI service

- **Local software tools** at data centres to prepare ingestions
- **Replication Manager** (RM) at data centres for exchanging to Import Manager and EUDAT cloud
- **Import Manager dashboard** to steer import and validation process
- **EUDAT cloud** with HTTP API and central data cloud with all unrestricted files and versions in multiple formats, where relevant
- Upgraded **CDI User Interface**, ordering and downloading facility, with integrated MySeaDataNet

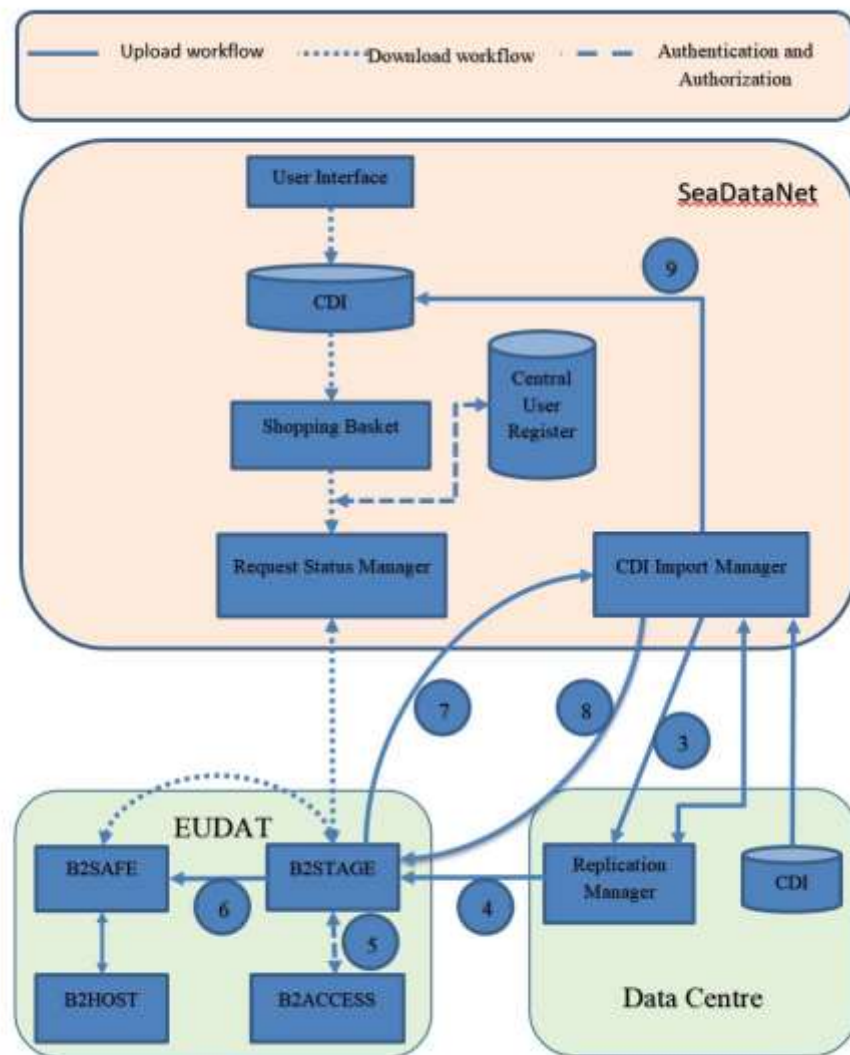


Figure ...: High level architecture

Which EUDAT centres?

- EUDAT is participating with 5 processing centres: CSC (FI), STFC (UK), DKRZ (DE), CINECA (IT) and GRNET (GR)
- The Import process will include only one EUDAT centre – **CSC** in order to lower risk as the import is complex and EUDAT centres have different infrastructures. This means that all RMs will exchange data with CSC, controlled by the Import Manager service which runs at MARIS
- Initially the CDI shopping system will communicate only with CSC as central cloud. Later in time EUDAT will establishing horizontal replication and load balancing between the 5 EUDAT centres. Users should not notice.

Installed base of CDI nodes



Installed base of CDI nodes

- 89 nodes with Download Manager
- 23 nodes with Interim solution
- 1 node with own software

Important condition is that the CDI service must stay operational while upgrading; this also applies for the related services for EMODnet portals and buffers!

Installed base of CDI nodes

No.	Partner organization (abbreviation)	EDMO code	Country	Download Manager version	Participating in the following projects:						
					SDC	EMODnet Bathymetry - HRSM	SDN	Geo-Seas	Upgrade Black Sea SCENE	EMODnet Chemistry	WaTeR
1	IFREMER	486	France	1.4.7	X	X	X	X		X	
2	NERI (DCE - DMU)	729	Denmark	1.4.6	X		X			X	
3	Shom	540	France	1.4.4	X	X	X	X			
4	RIHMI-WDC (aka RNODC)	681	Russia	1.4.4	X		X		X	X	
5	IMR	612	Norway	1.4.4	X		X			X	
6	NERC-BODC	43	United Kingdom	1.4.6	X	X	X	X		X	
7	OGS	120	Italy	1.4.6	X	X	X			X	
8	SMHI	545	Sweden	1.4.4	X		X			X	
9	MHI-DMIST	727	Ukraine	1.4.4			X		X	X	
10	ICES	730	Denmark	1.4.6	X		X			X	
11	IEO	353	Spain	1.4.6	X	X	X			X	
12	BSH-DOD	96	Germany	1.4.6	X	X	X	X			
13	FMI	1725	Finland	1.4.4	X		X			X	
14	NERC-BGS	42	United Kingdom	interim		X		X			
15	UHI-MB	841	Ukraine	1.4.4					X	X	
16	IMS-METU	696	Turkey	1.4.6	X		X			X	
17	MSI	713	Estonia	1.4.6	X		X			X	
18	SOI	723	Russia	1.4.4						X	
19	MI	396	Ireland	1.4.51	X	X	X			X	
20	HCMR-HNODC	269	Greece	1.4.6	X	X	X			X	

Installed base of CDI nodes

No.	Partner organization (abbreviation)	EDMO code	Country	Download Manager version	Participating in the following projects:						
					SDC	EMODnet Bathymetry - HRSM	SDN	Geo-Seas	Upgrade Black Sea SCENE	EMODnet Chemistry	WaTeR
21	RWS *	1526	Netherlands	1.4.7			X			X	
22	TNO	635	Netherlands	own software	X		X	X			
23	BRGM	1022	France	1.4.3				X			
24	RBINS-MUMM (RBINS-BMDC)	1578	Belgium	1.4.6	X		X	X		X	
25	IMWM (aka IMGW)	193	Poland	1.4.6	X		X			X	
26	NIOO	633	Netherlands	1.2.0	X		X				
27	SYKE	1104	Finland	1.4.4	X					X	
28	PANGAEA	3234	Germany	1.4.4			X			X	
29	IORAS (aka SIO-RAS)	685	Russia	1.4.3	X		X		X	X	
30	ENEA	136	Italy	1.4.7	X		X				
31	NIMRD	697	Romania	1.4.4	X	X	X		X	X	
32	MBSS-NIB	1229	Slovenia	1.4.4	X		X			X	
33	GEUS	1543	Denmark	1.4.4				X			
34	MRI	583	Iceland	1.4.6	X		X				
35	KNMI	632	Netherlands	1.1c	X		X				
36	IHPT	590	Portugal	1.4.4	X	X	X			X	
37	NIOZ	630	Netherlands	1.4.6	X	X	X			X	X
38	<u>UkrSCES</u>	1167	Ukraine	1.4.6	X				X	X	
39	PGI	609	Poland	1.4.4				X			
40	SMA	1215	Sweden	interim		X					
41	IOLR	963	Israel	1.4.6	X	X	X			X	
42	NGU **	1415	Norway	1.4.4				X			
43	IBSS	840	Ukraine	1.4.4			X		X	X	
44	ISMAR Ancona	144	Italy	1.4.51						X	
45	ISPRA	3009	Italy	1.4.4						X	

Installed base of CDI nodes

No.	Partner organization (abbreviation)	EDMO code	Country	Download Manager version	Participating in the following projects:						
					SDC	EMODnet Bathymetry - HRSM	SDN	Geo-Seas	Upgrade Black Sea SCENE	EMODnet Chemistry	WaTeR
46	LIAE	698	Latvia	1.4.6	X		X			X	
47	IO-BAS (aka BGODC)	692	Bulgaria	1.4.6	X	X	X	X	X	X	
48	VLIZ	422	Belgium	1.4.6	X		X			X	
49	BGR	1496	Germany	1.4.4				X			
50	OC-UCY	711	Cyprus	1.4.6	X		X			X	
51	NCSD	3092	Greece	1.4.4				X			
52	RWS - ICT	1527	Netherlands	interim		X					
53	ONU	1169	Ukraine	1.2.0					X		
54	IOF (aka IZOR)	700	Croatia	1.4.4	X		X				
55	<u>OceanWise</u>	2607	United Kingdom	interim		X					
56	BSCS	1546	Turkey	1.2.0					X		
57	DEU-IMST	801	Turkey	1.4.3					X		
58	NIMH	961	Bulgaria	1.4.3					X		
59	<u>IOI-UoM</u>	708	Malta	1.4.4	X	X	X				
60	SOVON	3393	Netherlands	1.4.4							X
61	GAMMA	1265	Georgia	1.4.3					X	X	
62	IPMA	3288	Portugal	1.4.3		X					
63	IGG - NRC	1545	Lithuania	1.4.4				X			
64	NHS	543	Norway	interim		X					
65	SHODB	731	Turkey	1.4.6			X				
66	IIM	1338	Italy	interim		X					
67	INHR	691	<u>Morocco</u>	1.4.3			X				
68	TSU-DNA	693	Georgia	1.4.3	X		X		X	X	
69	LU	2239	Latvia	interim				X			
70	EPA	2108	Lithuania	1.4.4			X				
71	IGME-ES	310	Spain	1.2.0		X		X			
72	INSTM	1232	Tunisia	1.4.6	X		X				
73	IMBK	2432	Montenegro	1.4.51	X		X			X	

Installed base of CDI nodes

No.	Partner organization (abbreviation)	EDMO code	Country	Download Manager version	Participating in the following projects:						
					SDC	EMODnet Bathymetry - HRSM	SDN	Geo-Seas	Upgrade Black Sea SCENE	EMODnet Chemistry	WaLTeR
74	EGK	1609	Estonia	1.4.3				X			
75	IMARES	1340	Netherlands	1.4.3							X
76	IFR	191	Bulgaria	1.4.6					X		
77	MAL	4665	Latvia	interim		X					
78	CNR - ISAC	149	Italy	1.4.6	X		X			X	
79	CNR-Spezia (ISMAR Lerici)	134	Italy	1.4.6	X		X				
80	IKI-RAS	761	Russia	1.4.4					X		
81	SNU-FF	733	Turkey	1.2.0					X		
82	UCC (MaRel)	1848	Ireland	1.4.4				X			
83	IU-FF-IMS	802	Turkey	1.4.3					X	X	
84	MOW	2243	Belgium	interim		X					
85	OGS - IRI (aka OGS RIMA)	2276	Italy	1.4.4	X	X	X	X		X	
86	NLHS (aka RNLHS)	574	Netherlands	interim		X					
87	RSHU	924	Russia	1.4.4			X				
88	GSI	366	Ireland	1.4.3		X		X			
89	BSTU	732	Turkey	1.4.4					X		
90	UB	2172	Spain	1.2.0				X			
91	CMR (NEA)	2267	Georgia	1.4.4					X	X	
92	ISMAR	145	Italy	interim		X					
93	NOA	1703	Greece	1.4.4				X			
94	IGS	2228	Ukraine	1.2.0					X		
95	CDG-CNRS	565	France	1.4.4				X			
96	DHMO	2231	Ukraine	1.4.6					X		
97	Geo SU (IGV)	2210	Sweden	interim		X					
98	IHM	1461	Spain	interim		X					
99	UTM-CSIC	2489	Spain	1.4.6	X	X					
100	MARUM	1568	Germany	interim		X					

Installed base of CDI nodes

No.	Partner organization (abbreviation)	EDMO code	Country	Download Manager version	Participating in the following projects:						
					SDC	EMODnet Bathymetry - HRSM	SDN	Geo-Seas	Upgrade Black Sea SCENE	EMODnet Chemistry	WaTeR
101	CONISMA	239	Italy	interim		X					
102	CNR-IAMC-Napoli	252	Italy	interim		X					
103	AU-FAF	2176	Turkey	1.4.4					X		
104	CNR-IGAG	2918	Italy	interim		X					
105	IOPAN	195	Poland	1.4.4	X	X	X				
106	GRID - Arendal	2543	Norway	interim		X					
107	GeoECOMAR	850	Romania	interim		X					
108	Jardfeingi	3289	Faroe Islands	interim		X					
109	MR (Marine Research)	3339	Romania	interim		X					
110	ICM	2157	Spain	interim		X					
111	GIS	4666	Slovenia	interim		X					
112	CNRS-MIO	3078	France	1.4.6	X						
113	SBR	521	France	1.4.7	X						
114	MBA	45	United Kingdom	1.4.7	X						

CDI upgrading activities

- Finalising technical developments for a robust and operational system with versions for development, testing, import, and production (like current CDI system)
- Connecting all SeaDataNet nodes to the upgraded CDI service
- Validating and working up the central CDI database and associated unrestricted data collection, including completing different formats.
- Integrating all components for public launch.

Timing

- Overall aim is publicly launching upgraded CDI service in **March 2019**.
- **Till end November 2018:** making the import cycle robust and lots of testing to ensure that the new **import and retrieval processes** work, both for unrestricted and for restricted data: **SDC technical team**.
- **Begin December 2018 – end February 2019:** rolling out the new technical configuration in test and in parallel to current operational system; installing and configuring **Replication Managers**. Each month 30 nodes in groups: **Data Centres with support by MARIS and IFREMER**
- **Begin January 2019 – end February 2019:** configuring the new system for the Data Centres with **Indirect Solution**. In operation MARIS will intermediate for new and updated entries towards the Import Manager. At present access default is 'restricted'; discussion on releasing (subsets of) data: **Data Centres with MARIS and IFREMER**

Timing

- **Begin December 2018 – end February 2019:** thorough testing of the new connections, both for unrestricted and for restricted data. Feedback loop to developers for possible bug fixing and refining: **Data Centres with MARIS and IFREMER**
- **Till end January 2019:** further improving the front side (user interface – shopping – RSM): **MARIS** in dialogue with Beta testers from the **SeaDataNet data centre community**.
- **Till end December 2018:** further expanding the central QA-QC services as part of import, following requirements: **MARIS, IFREMER and EUDAT**

Timing

- In parallel to technical configuring of the network nodes, the central data buffer of unrestricted data has to be fully checked and worked up as V0 database. This requires actions focused on data handling:
 - **September till end December 2018:** MARIS will **harvest** all **unrestricted** data sets in all formats. Checking in batch with OCTOPUS is ongoing for ODV files. OCTOPUS has a number of auto-correction functions. Feedback to data centres for corrections and updates of ODV and CDI files. (Hopefully also inclusion of unrestricted data from Interim Solution data centres).
 - **January – February 2019:** Freezing of the CDI contents (> 2.25 Million entries). **NO** mutations or additions in that short period.

Timing and versioning

- **January – February 2019:** Completing the V0 data collection with other file types (MedAtlas => ODV + NetCDF; ODV => NetCDF; NetCDF => ODV) by central conversion and updating of file types in CDIs. Other filetypes can not be converted. This action should result in a clean and consistent V0 CDI database and V0 central database of unrestricted data files.
- **Mid March 2019** integrating all components and switching to operation
- The CDI service will then be open again for updates and new data sets. From this point on we will work with **versioning**. This implicates that each updated CDI submission will be imported as a new version of the CDI and an associated data set (max possibly MedAtlas + ODV + NetCDF; or other file types, seismic as ZIP of SEG-Y + UKOAA nav file).

Versioning

- A cross table will safeguard the relations at CDI level between the different versions of a CDI and its links to the corresponding data files (by their PIDs).
- Queries in the CDI catalogue will focus on the current = latest version. However the CDI detail page will display links to the older versions, if any.
- Data products might have been built with older versions of data sets. The products include links to CDIs and their used version. This will be maintained allowing for reproduceability.

Full cooperation required

- The migration from the current CDI service to the upgraded CDI service is a challenging operation.

Full cooperation and engagement is required from the staff of the CDI nodes for:

- **Configuring and testing the new connections (RM and IS)**
- **Validating and correcting their data holdings**
- **Trying out and giving feedback on the new interfaces**

Virtual Research Environment (VRE)

- **VRE Online: 30 April 2019**

- D10.2: 30 April 2019: Virtual Research Environment (VRE) in the cloud operational; GRNET
- D10.4: 30 April 2019: Sub-setting application operational in VRE; IFREMER
- D10.6: 30 April 2019: Ocean Data View online operational in VRE: AWI
- D10.8: 30 April 2019: DIVA online operational in VRE; ULg
- D10.14: 30 April 2019: Visualisation services operational in VRE; Deltares
- D10.20: 30 April 2019: MySeaDataCloud operational; STFC

- **Expansion with Biology QA-QC in 3 phases:**

- D10.10: 21 December 2018: Phase 1 of Biology Data QC online operational; VLIZ
- D10.11: 31 October 2019: Phase 2 of Biology Data QC online operational; VLIZ
- D10.12: 30 April 2020: Phase 3 of Biology Data QC online operational; VLIZ

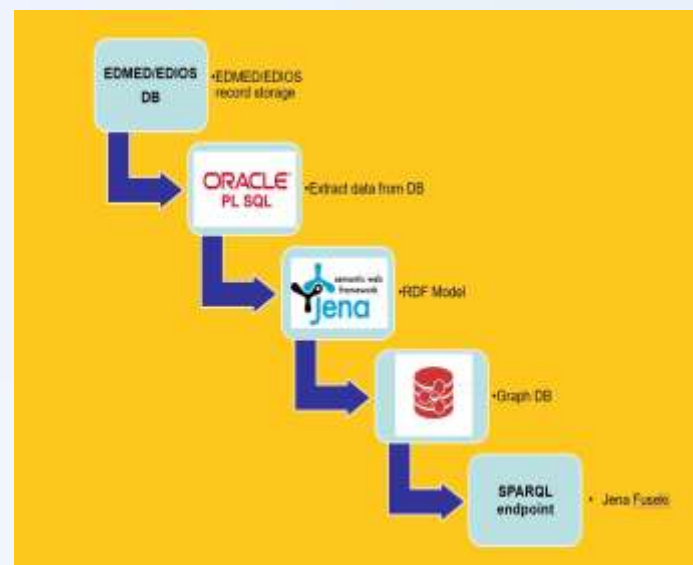
Linked Data principle implemented

- **RDF triple store and SPARQL services implemented for:**

- Vocabularies
- EDIOS
- EDMED
- EDMO

- **Well underway for:**

- EDMERP
- CSR
- CDI



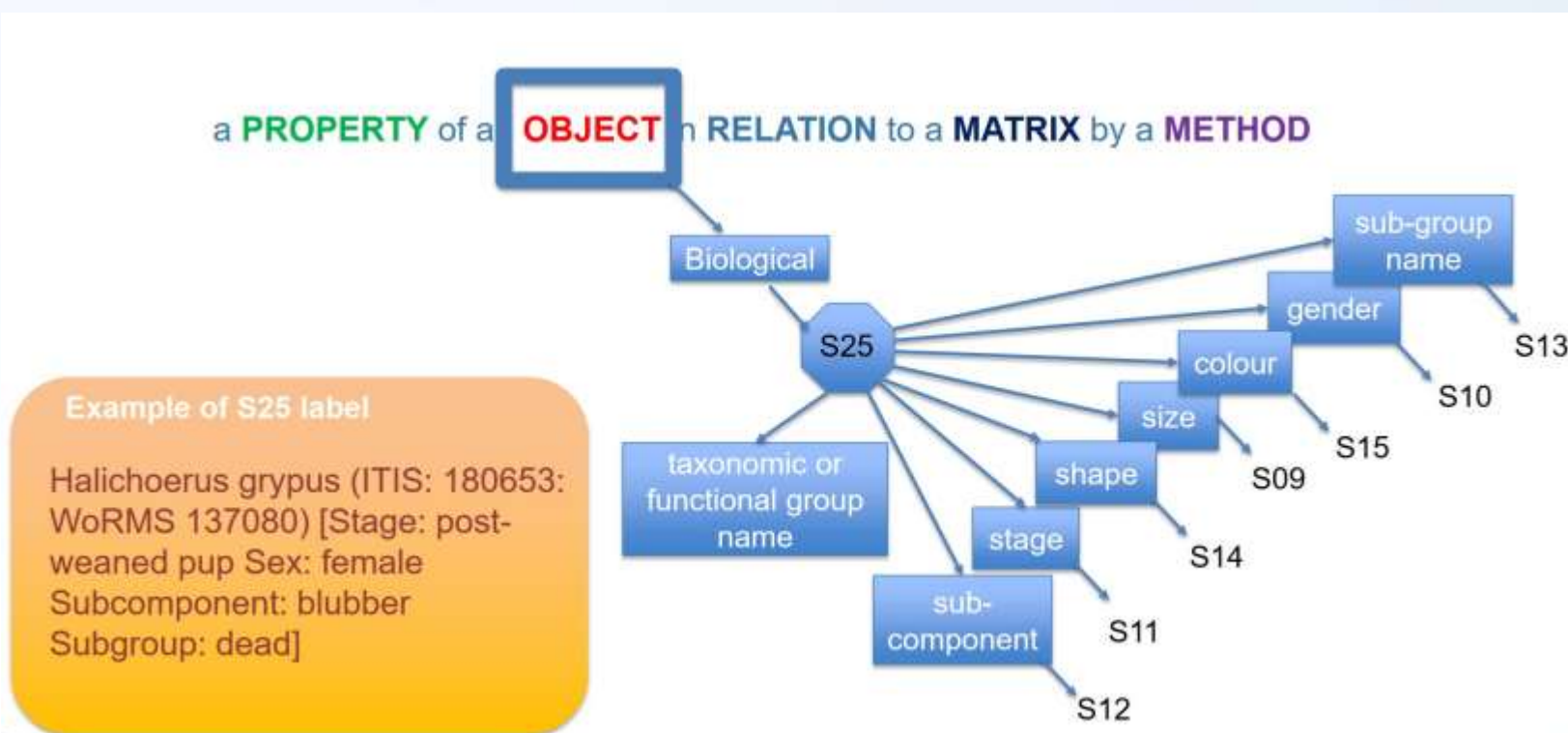
- **New URLs structure:**

- HTML search page: [http://\[metadir\].seadatanet.org/search/](http://[metadir].seadatanet.org/search/)
- SPARQL endpoint web interface: [http://\[metadir\].seadatanet.org/sparql/](http://[metadir].seadatanet.org/sparql/)
- SPARQL endpoint machine interface: [http://\[metadir\].seadatanet.org/sparql/sparql](http://[metadir].seadatanet.org/sparql/sparql)
- Details of a record: [http://\[metadir\].seadatanet.org/report/{ID}/](http://[metadir].seadatanet.org/report/{ID}/)

Maintaining and upgrading Vocab

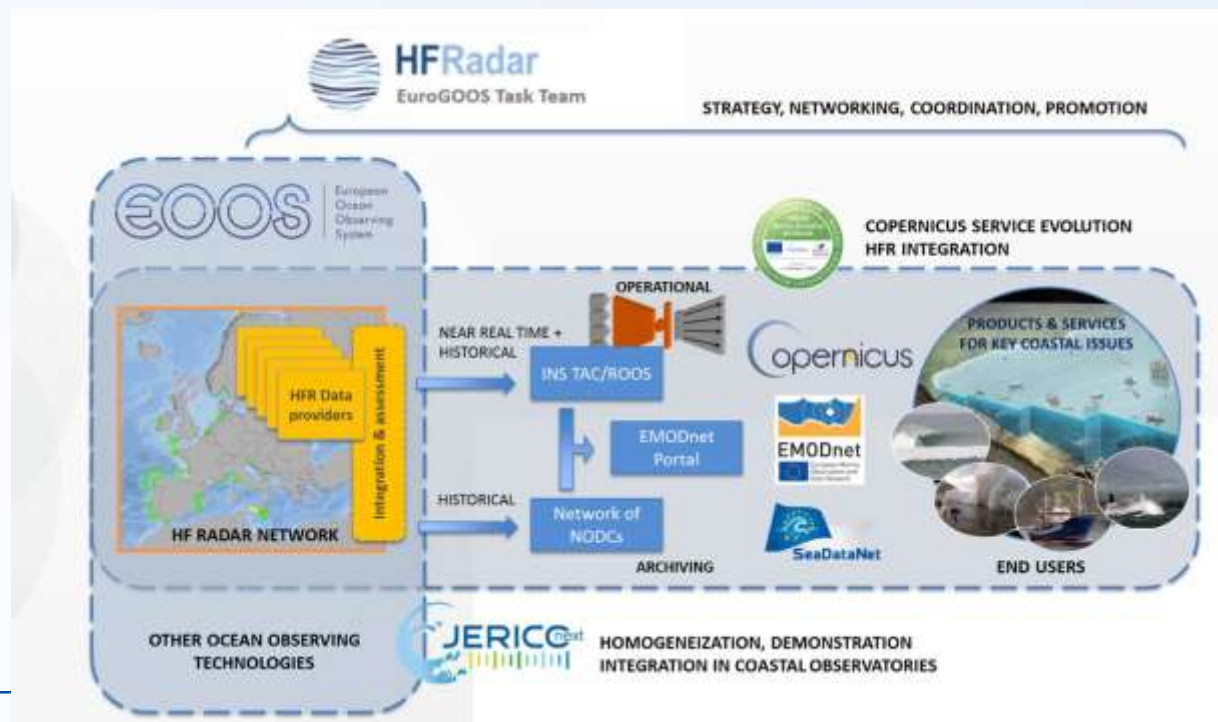
- Finalising 31 October 2020

- D8.2: 31 October 2020: Report outlining the status of the common vocabularies and vocabulary server, including vocabulary builder, and new vocabularies added; BODC



Handling new data types

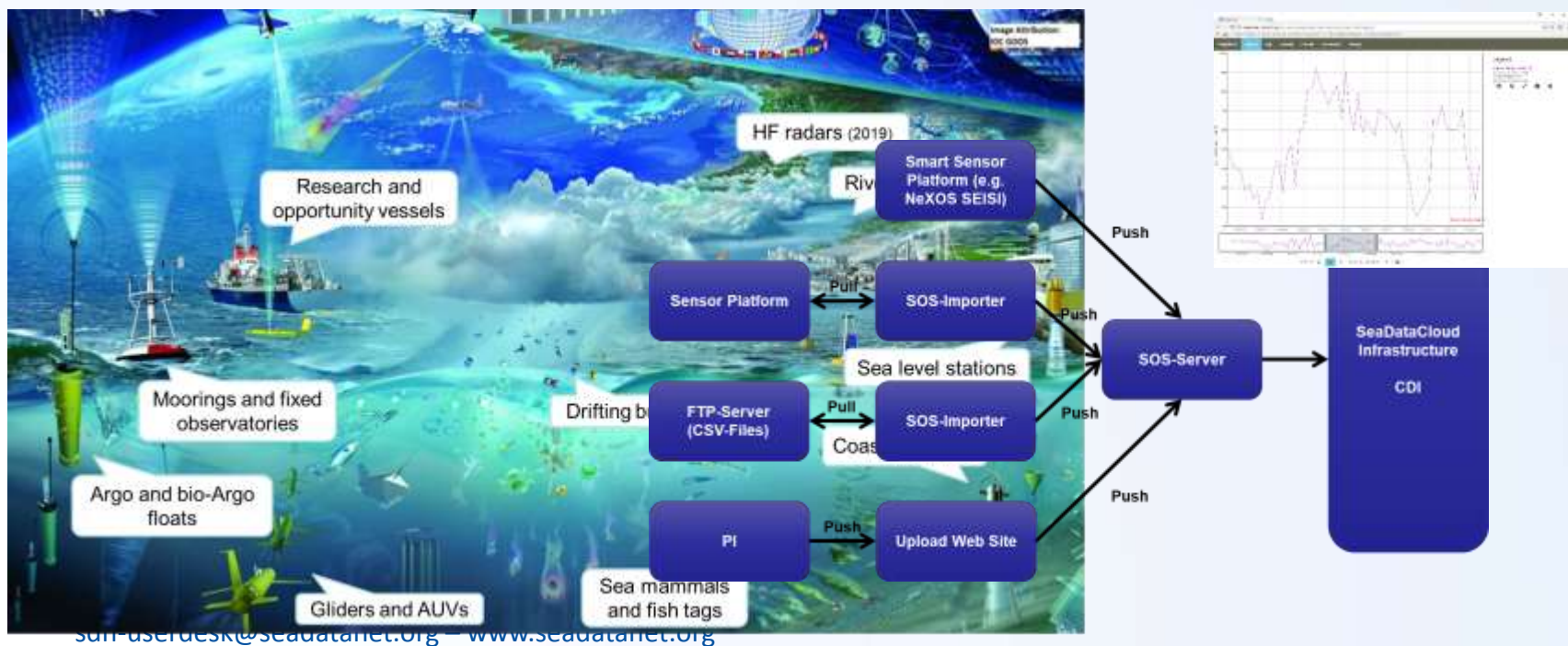
- **Data management protocols ready and published:**
 - D9.12: 30 June 2019: SeaDataNet data management protocols for HF Radar data; ETT
 - D9.13: 30 June 2019: SeaDataNet data management protocols for Flow Cytometer data; VLIZ
 - D9.14: 30 June 2019: SeaDataNet data management protocols for glider data; BODC



SWE Ingestion service for operational data streaming

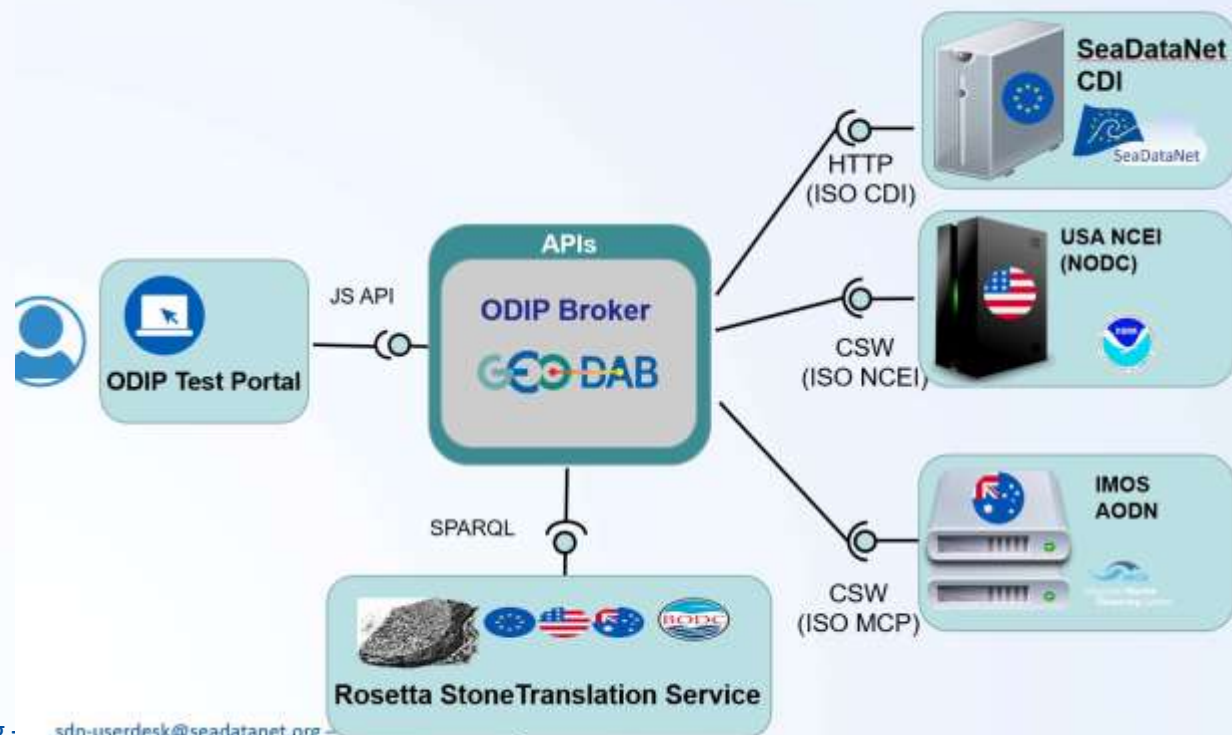
- Implementation in steps:

- D9.11: 28 February 2019: **Pilot** directory of monitoring platforms and sensors operational and feeding into EDIOS; IFREMER
- D10.18: 31 May 2019: SOS viewing services for data streams operational; 52North



Brokerage service for third party portals

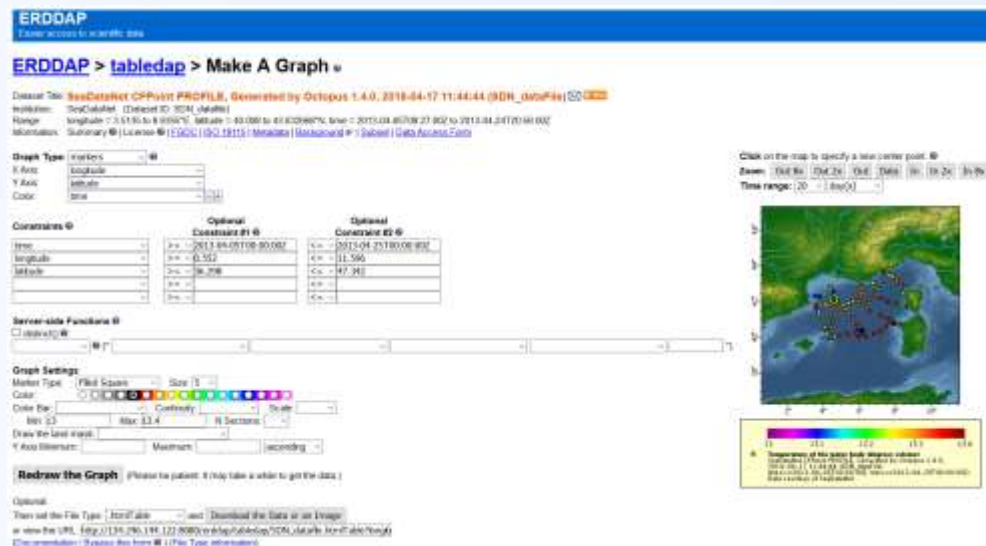
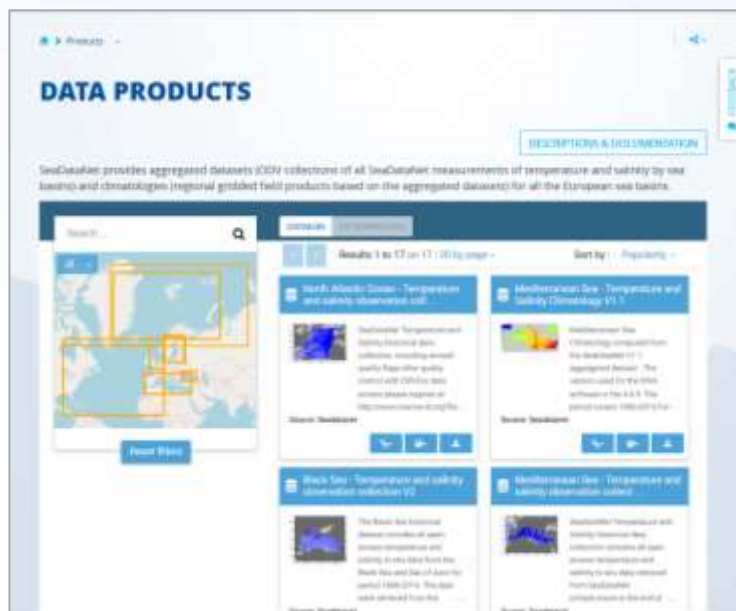
- **Operational: 30 April 2020:**
 - D9.16: 30 April 2020: External data sets ingested in the EMODnet Ingestion service; IFREMER

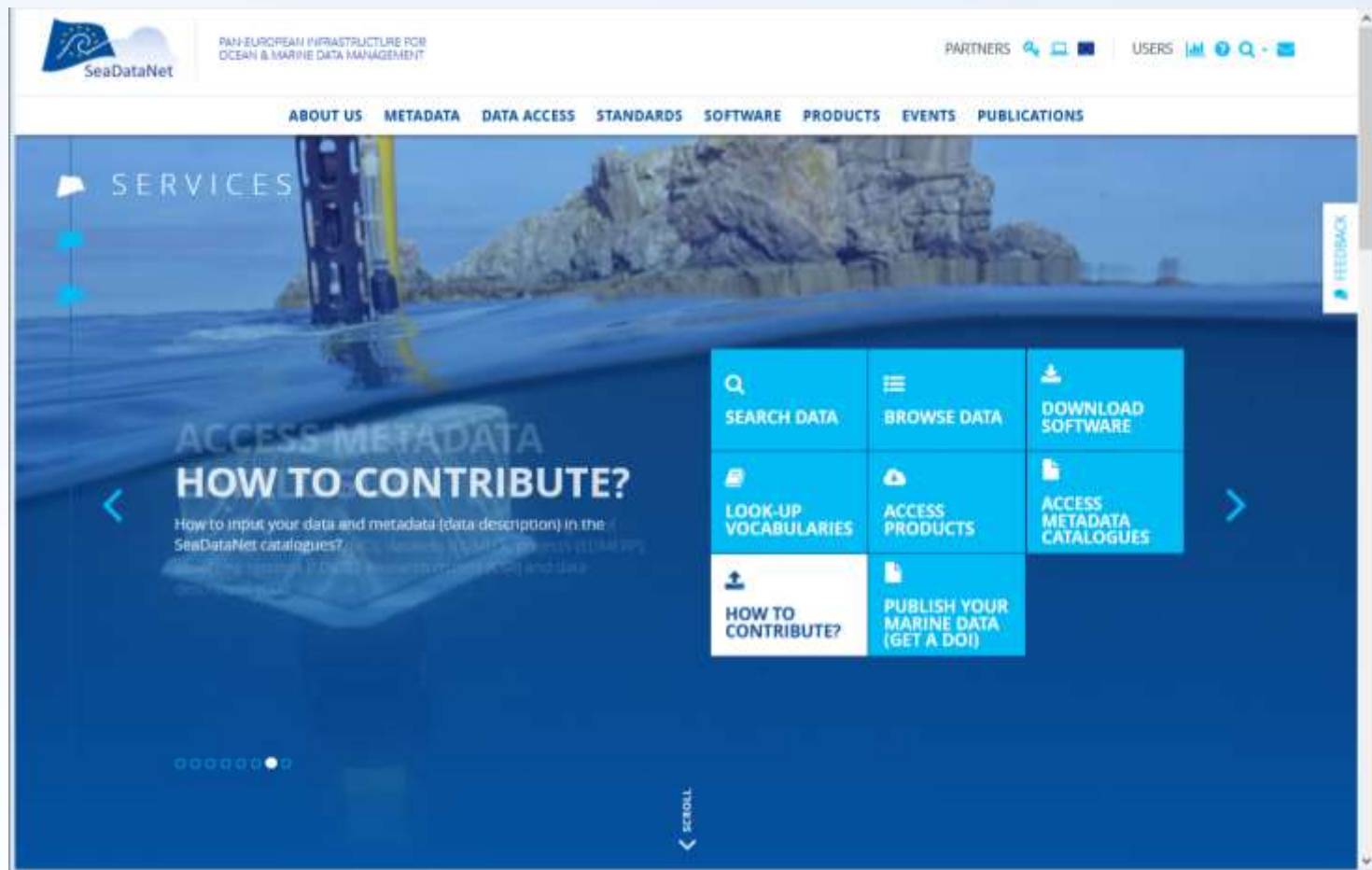


Publishing service for WP11 data products

- **Visualisation and subsetting services:**

- D10.15: 31 December 2018: Specification of Oceanotron services and development plan; IFREMER; Oceanotron changed to ERDDAP
- D10.16: 29 February 2020: Upgraded Oceanotron services operational for SeaDataNet WP11 data products; IFREMER; Oceanotron changed to ERDDAP





www.seadatanet.org