Institute of Oceanology, Polish Academy of Sciences

Development of data management processes

Marcin Wichorowski

IO PAN

SeaDataNet Final Meeting, IFREMER, Brest, September 15th

Oceanographic data at IO PAN



Extensive archives of information

Institute of Oceanology (IO PAN) maintains data gathered during research activities performed for more than 50 years. These data, originating from different sources are being processed according to common practice elaborated through generations of oceanographers, however there appears to be a very strong demand for standardization of procedures, especially while facing problems regarding cooperation in data exchange





Up to late '90...

Data were managed at unit level, just for publication purposes. No policy for storage, sharing, long time preservation of data collections. High risk of data degradation and lose.

Data management performed on the basis of good practice (good will?) of data originators, in some cases only scientific papers and reports remains.

Before SDN...

National projects focusing on data management improvement. Transfer of data to digital form, development of policy and deployment of databases reducing risk of degradation. Aggregation of data collections at organisation level. Lack of standards.

2008 - letter of intent, SDN2 associated partner

Data Center Infrastructure







Development of ODC: Data sets collection





Nie zdefiniowano jednego lub więcej parametrów wyszukiwania, następujące parametry zostaną pominięte: * Projekt * Głębokość

AREX 1993 Confluence Zone-E4 CNV	AREX1993	528 5	1993-07-04 10:53:12 0	
AREX 1993 Confluence Zone-F5.CNV	AREX1993	386	1993-07-04 08:13:29.0	
AREX 1993 Confluence Zone-F6.CNV	AREX1993	361	1993-07-04 05:13:51.0	
AREX 1993 Confluence Zone-G2 CNV	AREX1993	1452.5	1993-07-04 20:54:33.0	
AREX 1993 Confluence Zone-G3 CNV	AREX1993	1052	1993-07-05 01:24:06.0	
AREX 1993 Confluence Zone-G4 CNV	AREX1993	502.5	1993-07-05 05:41:59.0	
AREX 1993 Confluence Zone-G5 CNV	AREX1993	405	1993-07-05 12:05:01.0	
AREX 1993 Confluence Zone-G6 CNV	AREX1993	420	1993-07-05 18:19:53.0	
AREX 1993 Confluence Zone-H2.CNV	AREX1993	1555.5	1993-07-06 10:56:16.0	
AREX 1993 Confluence Zone-H3.CNV	AREX1993	1201.5	1993-07-06 08:27:07.0	
AREX 1993 Confluence Zone-H4.CNV	AREX1993	653.5	1993-07-06 05:26:14.0	
AREX 1993 Confluence Zone-H5.CNV	AREX1993	455.5	1993-07-06 03:03:28.0	
AREX 1993 Confluence Zone-H6.CNV	AREX1993	402.5	1993-07-06 00:53:12.0	
AREX 1993 Confluence Zone-I2 CNV	AREX1993	1901	1993-07-06 15:37:09.0	
AREX 1993 Confluence Zone-I3 CNV	AREX1993	1653.5	1993-07-06 21:05:52.0	
AREX 1993 Confluence Zone-I4 CNV	AREX1993	1493	1993-07-07 01:16:39 0	

Development of ODC: Data set details





Development of ODC: Vocabulary browsing



ZINTEGROWANY SYSTEM RZETWARZANIA DANYCH DCEANOGRAFICZNYCH VEW DAT DATA REGISTRATION Inport Importui Słowniki Ista słowników sedatanet	PROBLEMS AND ISSUES	ADMINISTRATION PANEL	Identify as wichor to the second seco
SeaDataNet Controlled Vocabularies	Poki È		
Nazwa	P ma nazwa	Url	Definicja
NVS mappings	NERC Vocabulary Server mappings index	http://vocab.ndg.nerc.ac.uk /list/C970/401	A catalogue of the mappings between vocabularies held in the NERC Vocabulary Server.
Geo-Seas skewness descriptors	Geo-Seas sediment grain-size skewness descriptors	http://vocab.ndg.nerc.ac.uk /list/GS50/1	Terms used to describe sediment grain-size skewness coefficients in the Geo-Seas project
<u>Biological entity names</u>	BODC parameter semantic model biological entity names	http://vocab.ndg.nerc.ac.uk /list/S250/29	Terms used to describe biological entities (organisms or part thereof) in the BODC Parameter Usage Vocabulary
MEDATLAS Parameter Usage Vocabulary	MEDATLAS Parameter Usage Vocabulary	http://vocab.ndg.nerc.ac.uk /list/P09/40	MEDATLAS Parameter Usage Vocabulary
SeaDataNet data access	SeaDataNet data access mechanisms	http://vocab.ndg.nerc.ac.uk /list/L071/2	Terms for mechanisms by which data objects described by SeaDataNet Central Data Index (CDI) records may be obtained before SeaDataNet is fully implemented.
SeaDataNet APG	SeaDataNet Agreed Parameter Groups	http://vocab.ndg.nerc.ac.uk /list/P031/21	Terms agreed within the EU SeaDataNet community to describe coarse-grained groupings of related measurement phenomena.
CSR units	SeaDataNet Cruise Summary Report quantification units	http://vocab.ndg.nerc.ac.uk /list/L181/1	Terms used as units in the quantification of what was collected or measured in a Cruise Summary Report (ROSCOP) record.
SeaDataNet device categories	SeaDataNet device categories	http://vocab.ndg.nerc.ac.uk /list/L05/33	SeaDataNet device categories
<u>SeaDataNet Geospatial Feature Types</u>	SeaDataNet Geospatial Feature Types	http://vocab.ndg.nerc.ac.uk /list/L021/1	SeaDataNet profile of ISO MD_GeometricObjectTypeCodeTerms code list. Known in SEA-SEARCH as 'Library 2' or cdi_measurement_codes.
SeaDataNet Contact Roles	SeaDataNet Contact Roles	http://vocab.ndg.nerc.ac.uk /list/C865/4	Terms used to define the responsibilities for a contact (person or organisation) either within the SeaDataNet projec or for the datasets described by SeaDataNet metadata.
MD DatatypeCode	MD_DatatypeCode	http://vocab.ndg.nerc.ac.uk /list/G120/1	Datatype of element or entity
MD DimensionNameTypeCode	MD_DimensionNameTypeCode	http://vocab.ndg.nerc.ac.uk /list/G130/1	Name of the dimension
SeaVoX Platform Categories	SeaVoX Platform Categories	http://vocab.ndg.nerc.ac.uk /list/L06/9	SeaVoX Platform Categories
MD CharacterSetCode	MD_CharacterSetCode	http://vocab.ndg.nerc.ac.uk /list/G090/1	Name of the character coding standard used in the resource
SeaDataNet PDV deprecates	SeaDataNet Parameter Discovery Vocabulary deprecates	http://vocab.ndg.nerc.ac.uk /list/P022/75	Deprecated terms describing fine-grained related groups of measured phenomena designed to be used in dataset discovery interfaces.
SeaDataNet Disciplines	SeaDataNet Parameter Disciplines	http://vocab.ndg.nerc.ac.uk /list/P081/3	Terms used to classify SeaDataNet Agreed Parameter Groups to provide topic/theme level terms in a hierarchical parameter discovery interface
IHB Sea Areas	International Hydrographic Bureau (1953) sea areas	http://vocab.ndg.nerc.ac.uk /list/C161/8	Terms used for sea areas from International Hydrographic Bureau, Limits of Oceans and Seas (Special Publication No. 23), 3rd edition 1953.
	Partnership for Observation of the	http://vocab.ndg.nerc.ac.uk	Research vessels deemed to be of interest to POGO. 'Of

Data Center Infrastructure : Infrastructure monitoring

Dodaj konwerter danych Usuń konwerter danych





MONITOROWANIE INFRASTRUKTURY

This option allows you to preview the current state of IT devices of the system. You can also view the details of the work and the status of servers running services.

Definiowanie struktur bazy - danych rozszerzonych	statu	s of servers running s	ervices.						
10 - C - C - C - C - C - C - C - C - C -	Hosts								
<u>problemów</u>		Status	Status	IP Address	Last chceck	Last change	Status information		
Archiwizacia danych	⊉	s9.iopan.gda.pl	UP	10.8.2.19	2013-09-23 14:43:31	2012-09-27 23:49:12	OK - 10.8.2.19: rta 0.023ms, lost 0%	Details	
Przywrócenie danych archiwalnych	₿	s8.iopan.gda.pl	UP	10.8.2.18	2013-09-23 14:43:31	2013-08-06 18:51:25	OK - 10.8.2.18: rta 0.171ms, lost 0%	Details	
Monitorowanie infrastruktury	⊉	s7.iopan.gda.pl	UP	10.8.2.17	2013-09-23 14:43:31	2012-09-27 23:48:46	OK - 10.8.2.17: rta 0.198ms, lost 0%	Details	
Rejestrowanie typów	⊉	s6.iopan.gda.pl	UP	10.8.2.16	2013-09-23 14:43:31	2012-09-27 23:48:33	OK - 10.8.2.16: rta 0.232ms, lost 0%	Details	
Dodawanie kolumn w tabeli	8	s5.iopan.gda.pl	UP	10.8.2.15	2013-09-23 14:43:31	2013-09-01 10:32:41	OK - 10.8.2.15: rta 0.199ms, lost 0%	Details	
Rejestrowanie plików KTR	DNZ	s2.iopan.gda.pl	UP	10.8.6.2	2013-09-23 14:43:31	2013-08-31 15:51:11	OK - 10.8.6.2: rta 1.105ms, lost 0%	Details	
Wymuszenie indeksowania	₽	s11.iopan.gda.pl	UP	10.8.2.21	2013-09-23 14:43:21	2013-08-31 15:29:01	OK - 10.8.2.21: rta 0.189ms, lost 0%	Details	
<u>pinott</u>	⊉	s10.iopan.gda.pl	UP	10.8.2.20	2013-09-23 14:43:31	2013-09-01 10:41:11	OK - 10.8.2.20: rta 0.174ms, lost 0%	Details	
	DHZ	s1.iopan.gda.pl	UP	10.8.6.1	2013-09-23 14:43:31	2013-08-31 16:21:31	OK - 10.8.6.1: rta 0.757ms, lost 0%	Details	
		iDRACs4	UP	10.252.0.4	2013-09-23 14:43:31	2013-08-03 12:26:49	OK - 10.252.0.4: rta 1.054ms, lost 0%	Details	
		IDRACs3	UP	10.252.0.3	2013-09-23 14:43:31	2013-08-03 12:26:48	OK - 10.252.0.3: rta 0.896ms, lost 0%	Details	
		iDRACs11	UP	10.252.0.7	2013-09-23 14:43:31	2013-08-03 12:26:38	OK - 10.252.0.7: rta 1.113ms, lost 0%	Details	
		iDRACs12	UP	10.252.0. <mark>8</mark>	2013-09-23 14:43:31	2013-08-03 12:26:38	OK - 10.252.0.8: rta 1.080ms, lost 0%	Details	
	-	DellPowerConnect2824	UP	10.252.0.250	2013-09-23 14:43:32	2013-06-06 16:52:47	OK - 10.252.0.250: rta 3.011ms, lost 0%	Details	
	-	FCB1Brocade4424	UP	10.252.0.20	2013-09-23 14:43:21	2013-08-28 10:25:09	OK - 10.252.0.20: rta 1.025ms, lost 0%	Details	
	-	FCB2Brocade4424	UP	10.252.0.21	2013-09-23 14:43:31	2013-08-28 10:24:31	OK - 10.252.0.21: rta 1.080ms, lost 0%	Details	
	-	JuniperEX2200	UP	10.252.0.251	2013-09-23 14:43:31	2013-07-10 13:19:13	OK - 10.252.0.251: rta 2.960ms, lost 0%	Details	
		JuniperSRX3400	UP	10.252.0.254	2013-09-23 14:43:31	2013-08-28 09:03:53	OK - 10.252.0.254: rta 0.881ms, lost 0%	Details	
		LOTUS_1	UP	10.8.2.31	2013-09-23 14:43:31	2012-09-28 14:35:06	OK - 10.8.2.31: rta 0.157ms, lost 0%	Details	
		LOTUS_2	UP	10.8.2.32	2013-09-23 14:43:31	2012-09-28 14:35:46	OK - 10.8.2.32: rta 0.179ms, lost 0%	Details	
		MySql	UP	10.8.2.64	2013-09-23 14:36:31	2012-09-27 23:47:34	PING OK - Packet loss = 0%, RTA = 0.20 ms	Details	
		VideoConference	UP	10.8.2.22	2013-09-23	2013-09-14	OK - 10.8.2.22: rta	Details	

Data Center Infrastructure : Help desk management



Zintegrov Przetwarz Danych Oceanoge	WANY ZANIA RAFICZ	SYSTEM NYCH	slip					39_PROD
	DATA REC	I 2 3 4 5	6 7 8	ADMINISTRATION	PANEL 14 15	PROJECTS USER PAN	2 22 223	2
				Zgłoszenia				
<u>Identyfikator</u>	TYP	Autor	<u>Data</u> <u>stworzenia</u>	<u>Spodziewany czas</u> <u>zamknięcia</u>	Priorytet	<u>Status</u>	<u>Komentarzy</u>	Działania
ERVICE_ALERT-1	Service Alert	SYSTEM	2012-05-29 15:57	2012-05-31	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-8	Service Alert	SYSTEM	2012-06-05 16:13	2012-06-07	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-9	Service Alert	SYSTEM	2012-06-06 10:14	2012-06-08	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-10	Service Alert	SYSTEM	2012-06-06 10:58	2012-06-08	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-11	Service Alert	SYSTEM	2012-06-06 11:08	2012-06-08	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-12	Service Alert	SYSTEM	2012-06-06 11:21	2012-06-08	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-13	Service Alert	SYSTEM	2012-06-06 12:02	2012-06-08	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-14	Service Alert	SYSTEM	2012-06-06 12:15	2012-06-08	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-15	Service Alert	SYSTEM	2012-06-08 09:59	2012-06-10	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-16	Service Alert	SYSTEM	2012-06-08 12:24	2012-06-10	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-17	Service Alert	SYSTEM	2012-06-08 12:54	2012-06-10	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-18	Service Alert	SYSTEM	2012-07-14 16:10	2012-07-16	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-19	Service Alert	SYSTEM	2012-07-15 12:34	2012-07-17	Wysoki	Rozwiązane	1	Edytuj
ERVICE_ALERT-20	Service Alert	SYSTEM	2012-07-15 13:28	2012-07-17	Wysoki	Rozwiązane	1	Edytuj
	Service		2012-07-15			D		

Zaznacz wszystkie Usuń zaznaczone Daj dostęp

Login użytkownika proszącego o dostęp Uzasadnienie Szczegóły obiektu którego dotyczy zgłoszenie Data zgłoszenia Zaznacz

Data Center Infrastructure : Data transformation



1.000	2.5273	7.375
1.500	2.5464	7.376
2.000	2.5096	7.375
2 500	2 4939	7 375

Ship:

* Bottom:

.

END

Cruise:

</hop> - <hop> <from>CNV HEADER - file input</from>

<enabled>Y</enabled>

<to>Replace in string</to>

Data Center Infrastructure : Pentaho Business Analytics

•Data access and ETL

- •Data discovery and analysis (OLAP engine)
- •Data mining

















Sea Surface Temperature

Sea surface temperature determined on the basis of satellite data for a cloudless atmosphere and derivatives (map of temperature gradients, location of thermal fronts)







Ocean colour

Surface chlorophyll *a* $C_a(0)$ concentrations (a), and the coefficient of total absorption of light of wavelength 440 nm by dissolved and suspended matter in the sea water, *a*(440nm) (b), on 24th April 2011 $a(440\text{nm}) = 10^{(0.096-0.965*\log(x))}$, where *x* is the ratio of the sea's reflectance for light wavelengths 490 and 665 nm, that is $x=R_{rs}(490)/R_{rs}(665)$.







Radiation balance

Instantaneous *NET* radiation flux and its corresponding downward and upward components (short-wave $-SW_d$, SW_u ; long-wave $-LW_d$, LW_u) at 11:00 UTC on 24th April 2011 at the Baltic Sea surface







- Photosynthetically Available Radiation (400-700 nm), PAR (a)
- Photosynthetically Used Radiation (the excitation energy of marine phytoplankton pigments equal to the energy of the radiation absorbed by these pigments), *PUR* (b)
- Photosynthetically Stored Radiation (the energy incorporated into the ecosystem, that is, primary production in energy units), *PSR* (c);
- **Quantity of oxygen** O₂ released during photosynthesis in the Baltic (d)

Photosynthesis



Benefits for the community



- normalization of data and information exchange procedures deploying widely used standards (eg. those developed within SeaDataNet)
- improvement of data exchange with international organisations (OpENDAP, NetCDF, WFS)
- automatisation of metainforamation discovery and publication
- improved management of research projects with developed tools for budget and resources administration
- monitoring of the administration processes and quality management within projects
- improved efficiency of research works providing fast and easy access to data repository
- cross-functioning integration with helpdesk, groupware, eAccess and eLearning modules

Promotion of data management







Conclusions



- Cooperation with SeaDataNet has changed policy, procedures, practices of data management at IO PAN and other organisations.
- Deployment of standards elaborated within SDN frame enforced development of data centre infrastructure providing services for scientific and oceanographic community

•New challenges were trigger of development of new initiatives both inside and outside the organisation

