

I dati ecologici marini LTER e l'Open Science

Un'opportunità

workshop EcoNÀOS - 10 Aprile 2017, CNR-ISMAR Venice
Annalisa.Minelli@ve.cnr.ismar.it

In breve:

1. il concetto di life cycle (spiral) della ricerca in un'ottica Open
2. task1: armonizzazione del database;
3. task2: raccolta delle informazioni accessorie e metadatazione;
4. task3: integrazione dei dati nell'infrastruttura esistente;
5. task4: open science non solo per i dati;
6. task5: data citation.

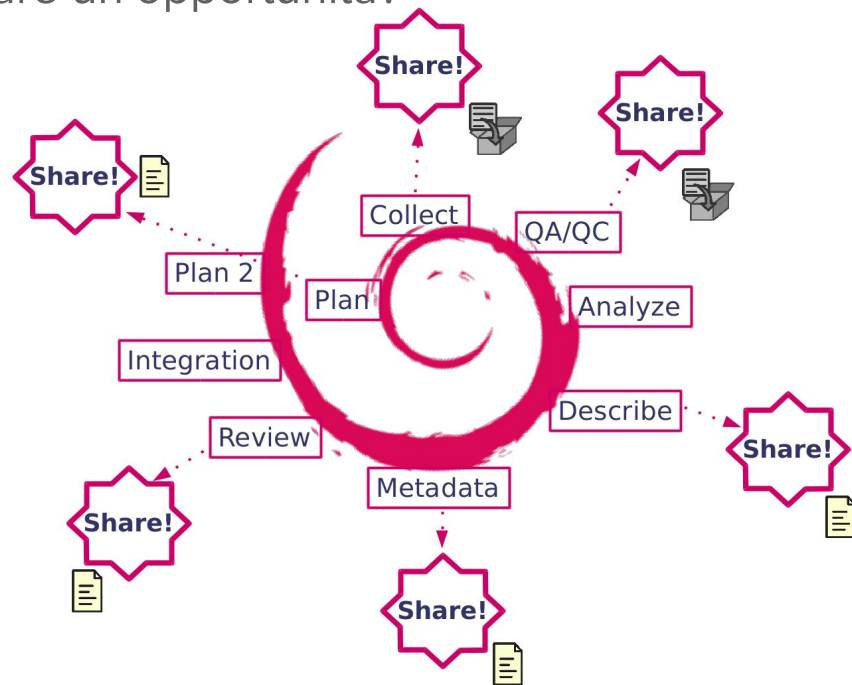
Lifecycle della della ricerca

perché aprire la ricerca potrebbe rappresentare un'opportunità?

l'integrazione della ricerca di partenza con nuove idee può portare a **nuove linee di ricerca**;

l'arricchimento del dataset iniziale (a partire da dati grezzi) può portare a studi anche di **diversa natura** sullo stesso dataset;

la condivisione non solo di risultati ma anche di idee e dati fornisce più materiale per **valutare la ricerca nel suo complesso**.



I dati ecologici LTER marini - progetto EcoNÀOS



EcoNÀOS

Ecological
North Adriatic
Open Science
Observatory System

Dati:

- 50 anni di registrazioni: dal 1965 al 2015
- Dati referenziati nello spazio e nel tempo (coordinate + data e ora)
- Più di 200 crociere oceanografiche
- Parametri abiotici e fitoplancton (presto integrati con lo zooplancton)
- ~56000 registrazioni (nucleo iniziale)
- Diversi strumenti e metodi di campionamento e di analisi (diverse le tecnologie)

Task1: armonizzazione del database

Armonizzazione:

1. sintattica: standardizzazione formale dei dati;
2. strutturale: adozione di modelli di dati condivisi;
3. semantica: propria alla semantica del dato.

È fondamentale per:

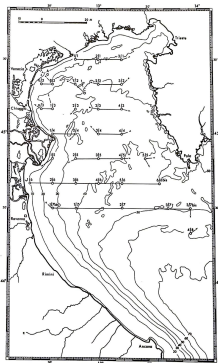
- renderli riutilizzabili e ricercabili (Findable and Reusable);
- renderli il più possibile utilizzabili da chiunque (Interoperable);
- renderli persistenti: esempio del cambio nella tassonomia di una specie;
l'armonizzazione permette ai ricercatori di sfruttare al meglio il dato - in diversi periodi storici.

	A	B	C	D	E	F
1	N	CRUISE	STATION	SHIP	dd/mm/yy	hh:mm
2						
3	1	PP/1	B	VERCELLI	12/04/1965	9.33
4	2	PP/1	B	VERCELLI	12/04/1965	9.33
5	3	PP/1	B	VERCELLI	12/04/1965	9.33
6	4	PP/1	B	VERCELLI	12/04/1965	9.33
7	5	PP/1	C	VERCELLI	12/04/1965	12.20
8	6	PP/1	C	VERCELLI	12/04/1965	12.20
9	7	PP/1	C	VERCELLI	12/04/1965	12.20
10	8	PP/1	C	VERCELLI	12/04/1965	12.20
11	9	PP/2	A	VERCELLI	28/04/1965	6.42
12	10	PP/2	A	VERCELLI	28/04/1965	6.42
13	11	PP/2	A	VERCELLI	28/04/1965	6.42
14	12	PP/2	A	VERCELLI	28/04/1965	6.42
15	13	PP/2	B	VERCELLI	28/04/1965	9.10
16	14	PP/2	B	VERCELLI	28/04/1965	9.10
17	15	PP/2	B	VERCELLI	28/04/1965	9.10
18	16	PP/2	B	VERCELLI	28/04/1965	9.10
19	17	PP/2	C	VERCELLI	28/04/1965	11.20
20	18	PP/2	C	VERCELLI	28/04/1965	11.20

Task1: armonizzazione del database

Prime osservazioni:

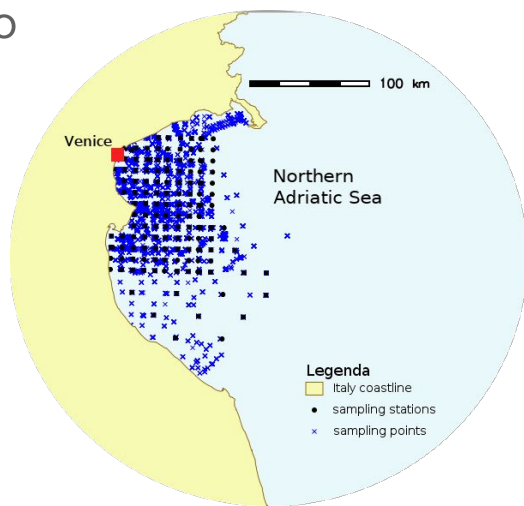
- punti a terra
- stazioni ricorrenti
- eterogeneità nel formato descrittivo



fonte: P. Franco (1965)

Azioni adottate:

- correzione della posizione geografica dei punti a terra
- individuazione delle stazioni “standard” di campionamento e studio della loro variabilità nel tempo
- rimozione di eventuali eterogeneità nel descrittore delle stazioni



Task1: armonizzazione del database

TODO:

- standardizzazione strutturale degli altri parametri
- individuazione di formati standard condivisi per i singoli parametri (armonizzazione sintattica)
- gestione efficiente della tassonomia (armonizzazione semantica)

Task2: raccolta di dati ancillari storici

Prime osservazioni:

- dati raccolti in un periodo storico ampio
- diversi metodi e strumenti
- influenza sull'attendibilità del dato/misura

Azioni adottate:

- ricostruzione storica delle crociere e degli strumenti utilizzati

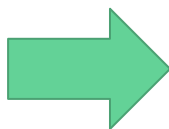
Crociere '65, '66, '78 e '79:

- bollettini oceanografici (P. Franco)
- ricerca bibliografica

Ship P. Vercelli	Latitude 45°00'.0 N	Marsden no. 10° 179/1° 42	Date 23.7.1965	Max. depth sample m 26.5	Barometer mbar 1,007	Air temp. dry wet 24.3	Wind speed 4.0					
Station 12740-0.0	Longitude 12°40'-0.0 E	Sonic depth m 30.3	Time (GMT) 1909 + 0145	Secchi disc m 25.0	Visibility	Cloud type cover	directionSE force 1					
Depth	t	s	q	A	zAD	O ₂	O ₂ %	pH	Tot. alk.	Spec. alk.	Tot. CO ₂	
St. 0,5	24.38	36.40	24.70	325.2	0,0000	4.89	4.60	106.44	8.18	2.91	0.139	2.58
St. 1	24.18	36.51	24.75	320.5	0,0016	5.01	4.63	106.13	8.31	2.82	0.136	2.37
Obs. 5	23.57	37.49	25.67	233.1	0,9127	4.97	4.64	107.13	8.31	2.75	0.129	2.30
Obs. 5,5	23.49	37.61	25.79			4.97	4.65	107.00	8.31	2.74	0.128	2.29
Obs. 10	20.64	37.77	26.71	134.6	0,9219	5.42	4.87	111.21	8.29	2.85	0.132	2.46
Obs. 11	20.01	37.81	26.92			5.52	4.92	112.14	8.28	2.87	0.133	2.49
Obs. 20	16.32	37.83	27.73	38.2	0,9305	5.47	5.24	106.67	8.29	2.79	0.129	2.41
Obs. 25,5	14.35	37.84	28.32			5.43	5.47	99.28	8.30	2.73	0.127	2.35

Ship P. Vercelli	Latitude 45°00'.0 N	Marsden no. 10° 179/1° 42	Date 24.7.1965	Max. depth sample m 11.5	Barometer mbar 1,007	Air temp. dry wet 21.5	Wind speed 0.5					
Station 12720-0.0	Longitude 12°30'-0.0 E	Sonic depth m 37.5	Time (GMT) 0614 + 0118	Secchi disc m 8.0	Visibility	Cloud type cover	directionSE force 1					
St. 0,5	24.32	22.41	1413	1350.7	0,0000	4.09	5.05	96.91	8.00	3.44	0.270	
Obs. 1	24.85	22.79	18.11	956.9	0,0057	5.16	4.88	106.13	8.11	3.20	0.219	
Obs. 1,5	24.58	32.97	21.96			5.43	4.71	115.34	8.21	3.13	0.167	2.77
Obs. 5	23.99	34.68	23.42	447.5	0,9338	5.20	4.70	110.71	8.21	3.14	0.159	2.73
Obs. 6	23.82	35.17	23.84			5.14	4.70	109.39	8.21	3.14	0.157	2.72
Obs. 10		36.31				5.08			8.16	3.09	0.150	2.73
Obs. 11,5		36.74				5.06			8.14	3.07	0.147	2.73

fonte: P. Franco (1965)



Preparazione dei Metadati

Task3: condivisione dei dati e interoperabilità

Prime osservazioni:

- il progetto non influirà sul metodo di lavoro dei ricercatori
- distinguo: uso interno ed esterno dei dati

Azioni adottate:

- inserimento di alcuni dati in influxDB
- inserimento di un campione di dati in GET-IT

TODO:

- studio delle migliori tecniche di sfruttamento del dato (in funzione delle sue caratteristiche): 3D e timestamped
- integrazione del dato nelle piattaforme di gestione esistenti

The screenshot shows the GET-IT web application interface. The header includes the Ritmare logo, a search bar, and a 'Sign in' button. The navigation menu contains links for HOME, LAYERS, SENSORS, DOCUMENTS, VIEWS, PEOPLE, SEARCH, and SERVICES. The main content area displays two sensor entries:

- Autosol laboratory salinometer**
URL: http://p7.irea.cnr.it/sensors/vesk.ve.ismar.cnr.it/procedure/Autosol/noModelDeclared/noSerialNumberDeclared/20150903114559504_50768
A 'Sensor details' button is visible below the URL.
- Spettrofluorimeter Beckman Gold System**
URL: http://p7.irea.cnr.it/sensors/vesk.ve.ismar.cnr.it/procedure/Beckman/Gold_System/noSerialNumberDeclared/20150903122743511_56166
A 'Sensor details' button is visible below the URL.

On the left side, there is a 'Provider' section for ASD, listing details for CNR ISMAR, including the name Mauro Bastianini, organization, position (researcher), and address (Castello 2737/II, Arsenale Tesa 104, Postcode 30122, City).

da: <http://vesk.ve.ismar.cnr.it/sensors/>

Task4: condivisione del ciclo di vita della ricerca

Prime osservazioni:

- progetti di ricerca
- dati (grezzi e rifiniti)
- metadata
- tools e codice eventualmente sviluppato
- risultati della ricerca
- revisioni

Azioni adottate:

- scrittura di un research idea (RIO?)
- pubblicazione codice di pretrattamento geografico dei dati (Github)



Task5: citazione di un dataset dinamico

Prime osservazioni:

- citazione di progetti di ricerca e prodotti della ricerca (papers)
- dataset integrabile:
 - nuove osservazioni - stessi sensori
 - vecchie e nuove osservazioni - sensori diversi
- citazione del dataset e di porzioni del dataset tramite URI (DOI)

TODO:

- messa a punto di uno workflow per la citazione dinamica del dataset (e porzioni)

Azioni adottate:

- studio dei migliori metodi in corso
- collaborazione con il Dr. Silvello (UniPD)
- partecipazione alla 9° Plenary di RDA (Research Data Alliance) - prossima affiliazione al WG sul data citation



Grazie!

..ed ora:

- Questionario breve
- Discussione libera

Lavori scientifici citati nella presentazione:

- Franco, Paolo. "OCEANOGRAPHY OF NORTHERN ADRIATIC SEA: 2. HYDROLOGIC FEATURES: CRUISES JANUARY-FEBRUARY AND APRIL-MAY 1966." ARCH OCEANOGR LIMNOL. 17 (SUPPL), P 1-97. 1972. ILLUS. (1972).

