

# Raccolta di dati ausiliari Armonizzazione strutturale Invito all'interoperabilità

---

Cosa è stato fatto

workshop EcoNÀOS - 10 Aprile 2017, CNR-ISMAR Venice  
[Annalisa.Minelli@ve.cnr.ismar.it](mailto:Annalisa.Minelli@ve.cnr.ismar.it)

# In breve:

1. fare ordine in (e arricchire) un database: dati ausiliari
2. armonizzazione sintattica del database:
  - a. correzione della posizione geografica
  - b. correzione dei nomi delle stazioni
  - c. prime osservazioni
3. invito all'interoperabilità: a che cosa (e chi) servono i dati?
  - a. uso interno
  - b. uso esterno

# I dati ausiliari

database di partenza: ~56000 osservazioni, dal 1965 a oggi

## CONSISTENZA?

- evoluzione negli strumenti di campionamento e di laboratorio (tecnologica)
- evoluzione nelle tecniche di misurazione (metodologica)
- evoluzione nelle unità di misura (scientifica)

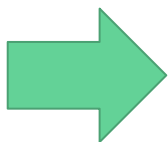


come si può, se non **analizzare**, almeno **rappresentare** il dataset nel suo insieme?

# I dati ausiliari

## Operazioni effettuate:

1. recupero di informazioni storiche: crociere del 1965, 1966, 1978, 1979 (reports di Paolo Franco)
2. individuazione nel database dei dati provenienti da quelle crociere
3. analisi dei metodi e strumenti utilizzati per la misurazione di ogni parametro
4. approfondimento e documentazione teorica e fotografica per ognuno
5. confronto con i dati in nostro possesso



- stima della precisione delle misure e della possibilità di confronto dei dati con il resto delle registrazioni
- preparazione dei metadati dei “sensori”

# I dati ausiliari

Output: tre report che andranno ad integrare i **metadata** ed aumentano la consapevolezza dell'**attendibilità** del dato

Paolo Franco, 1970. Oceanography of Northern Adriatic sea. 1 hydrologic features: cruises july-august and october-november 1965. Archivio di oceanografia e limnologia, Vol. 16, supplemento, pp.93.

CRUISES 1965	Summer 16 July - 17 August 1965
	Autumn 12 October - 3 November 1965

Materials and methods

VESSEL



NOTE

On the whole samplings and analytical methods followed the lines indicated by Strickland and Parsons in their handbook of sea-water analysis<sup>1</sup> (1960-1955).

<sup>1</sup> Strickland, J.D.H. Parsons, T.A. 1960. A practical handbook of seawater analysis. B.U. Fish. Res. Bd Can. 167, p. 311.

Paolo Franco, 1972. Oceanography of Northern Adriatic sea. 2 hydrologic features: cruises January-february and april-may 1966. Archivio di oceanografia e limnologia, Vol. 17, supplemento, pp.97.

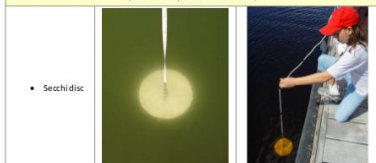
CRUISES 1966	12 January - 5 February 1966
	19 April - 14 May 1966

Materials and methods

VESSEL

- Chartered M/Y Sea Queen [missing photos and/or other data about the vessel](#)

TRANSPARENCY AT NOON<sup>1</sup> STATIONS (unit measure = meter)  
(stations sampled in diurnal hours)



- Secchi disc
- Submarine photometer [missing photos, manufacturer brand or other data about the instrument](#)

<sup>1</sup> Noon stations: diurnal stations.

Paolo Franco, 1982. Oceanography of Northern Adriatic sea. Data from the cruises of the years 1978 and 1979. Archivio di oceanografia e limnologia, Vol. 20, supplement 2, pp.207.

CRUISES 1978-79	1978	1	11-14 Mar
		2	3-22 Jul
		3	25-28 Sep
		4	28 Nov - 11 Dec
	1979	5	11-17 Mar
		6	1-16 Aug
		7	8-29 Oct

Materials and methods

VESSEL

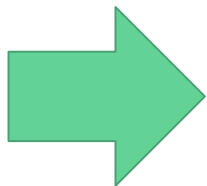


- RV Umberto D'Ancona

# Armonizzazione sintattica del database

## Criticità:

1. non punti in stazioni precise ma nubi di punti attorno a stazioni predefinite
2. punti a terra
3. punti senza coordinate
4. punti senza coordinate né nome della stazione
5. eterogeneità formale nei nomi delle stazioni predefinite



correzione delle anomalie e tentativo di armonizzazione

# Armonizzazione sintattica del database

## Nubi di punti e stazioni predefinite:

- assenza dei dispositivi GPS a bordo fino agli anni '90
- decisione di mantenere le nubi di punti
- vettorializzazione delle stazioni predefinite:
  - tre versioni dal 1965 al 1978
  - 161 diversi nomi di stazioni predefinite individuate

long	lat	name	periods
12:29E	45:24N	A	summer_1965,autumn_1965
12:52E	45:16N	C	summer_1965,autumn_1965
12:46E	45:30N	1/1	summer_1965,autumn_1965
13:14E	45:20N	3/2	summer_1965,autumn_1965
12:46E	45:20N	Y/2	autumn_1965
12:46E	45:25N	04/2A	1978,1979
13:21E	44:45N	09/2E	1978,1979
13:13E	44:30N	08/0G	1978,1979

# Armonizzazione sintattica del database

Carte storiche delle tre versioni di stazioni predefinite:

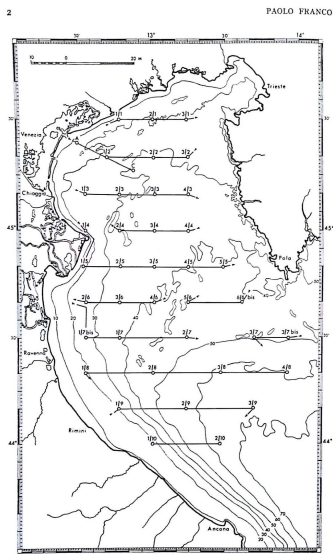


FIG. 1. Track and stations pattern of the summer cruise (Bathymetry in meters from Istituto Idrografico della Marina, Chart 1253).

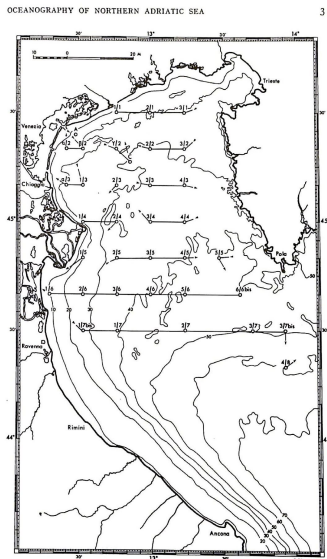
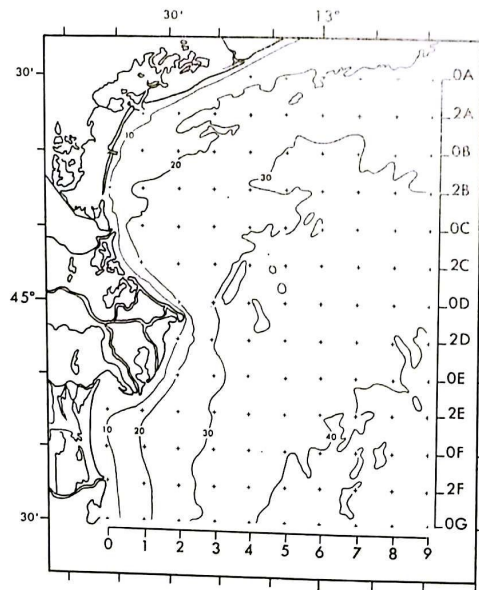


FIG. 2. Track and stations pattern of the autumn cruise (Bathymetry in meters from Istituto Idrografico della Marina, Chart 1253).



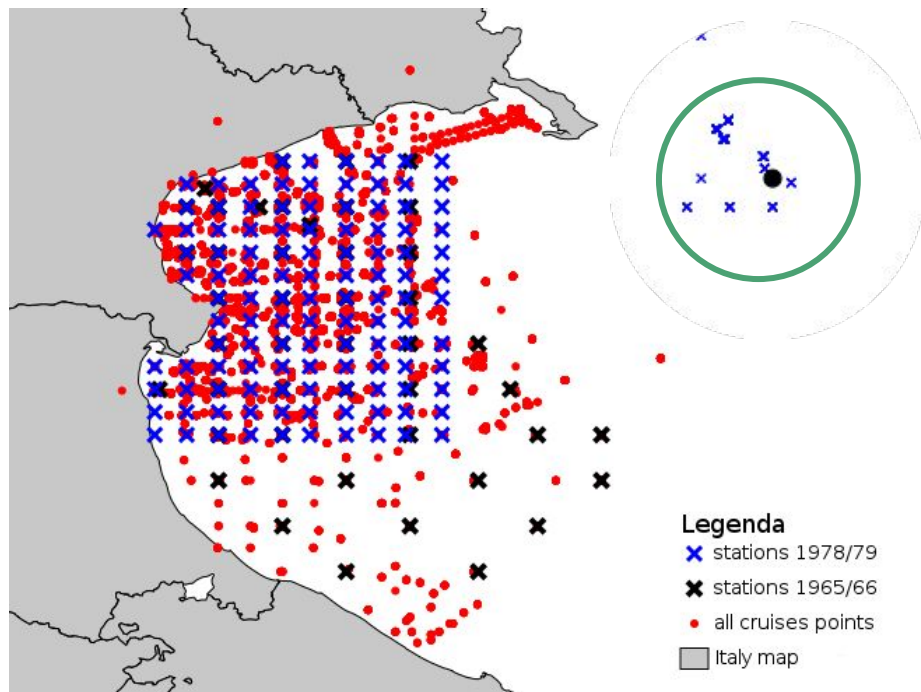
- nomenclatura numerica (es. 1/4, 5/4 etc.) negli anni '60
- nomenclatura alfanumerica (es. 02/0B, 05/2A etc.) negli anni '70

fonte: P. Franco (1965, 1966, 1980)



# Armonizzazione sintattica del database

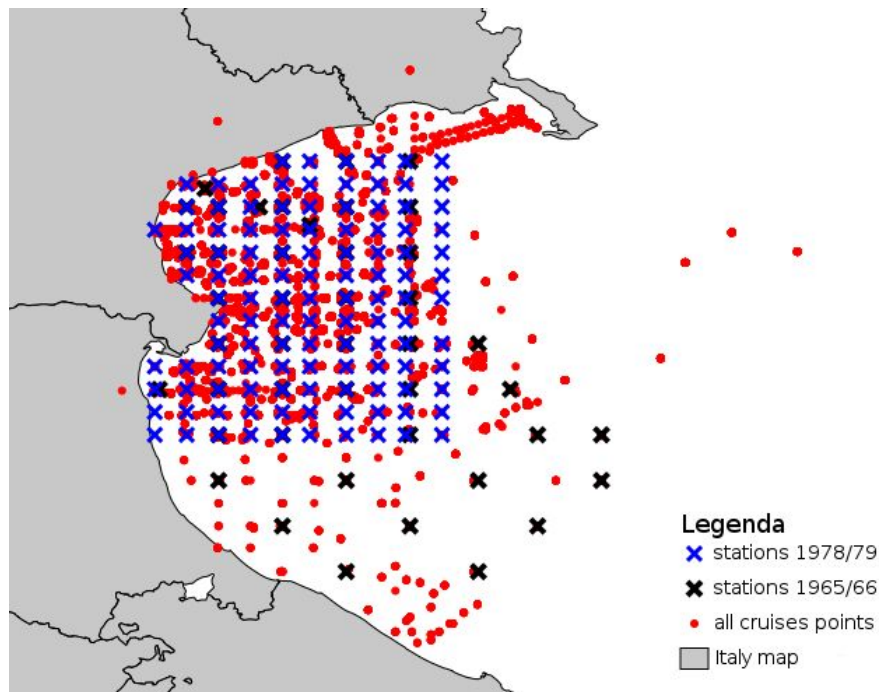
## Vettorializzazione delle stazioni e importazione dei dati in GRASS GIS



- nubi di punti attorno alle stazioni predefinite
- punti che cadono a terra (~25)
- punti non identificati da coordinate (~10)
- punti non identificati né da coordinate né da nomi di stazione (1)

# Armonizzazione sintattica del database

## Vettorializzazione delle stazioni e importazione dei dati in GRASS GIS



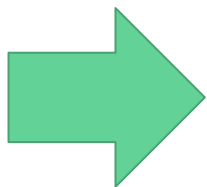
- nubi di punti attorno alle stazioni predefinite
- punti che cadono a terra (~25)
- punti non identificati da coordinate (~10)
- punti non identificati né da coordinate né da nomi di stazione (1)

# Armonizzazione sintattica del database

## Correzione punti a terra e armonizzazione del formato delle stazioni

Perché automatizzare?

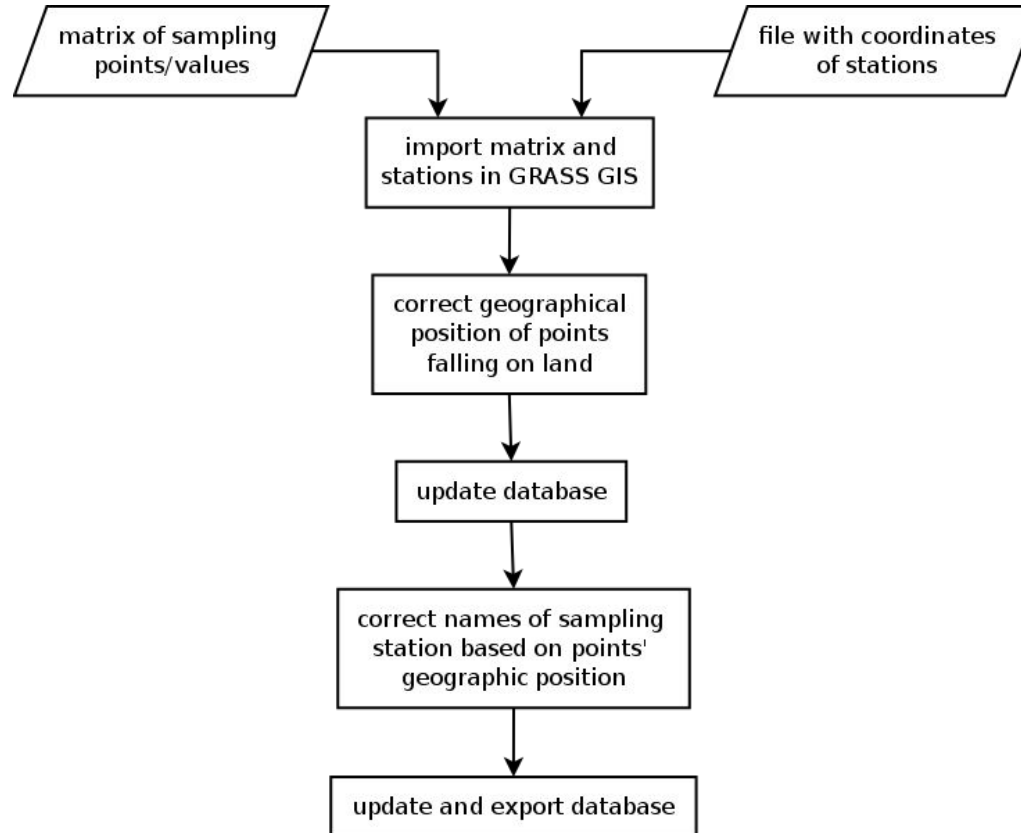
- per i punti a terra: possibilità di errori di posizionamento (anche con i dispositivi più avanzati)
- per i nomi delle stazioni: caso troppo ricorrente e troppe varianti per un medesimo punto



correzione geografica e semantica, con l'implementazione di una routine (script) che coinvolge GRASS GIS e Python

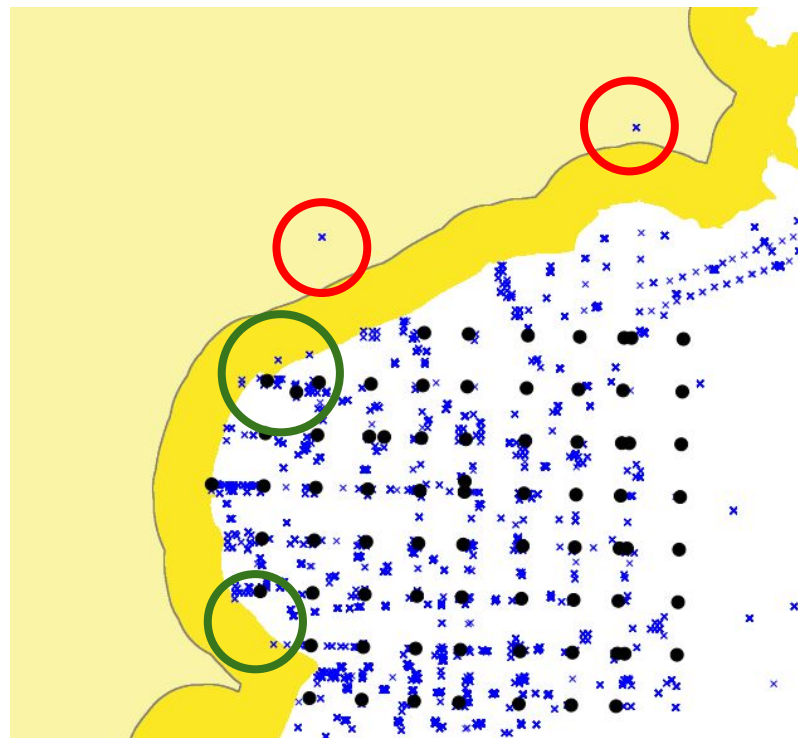
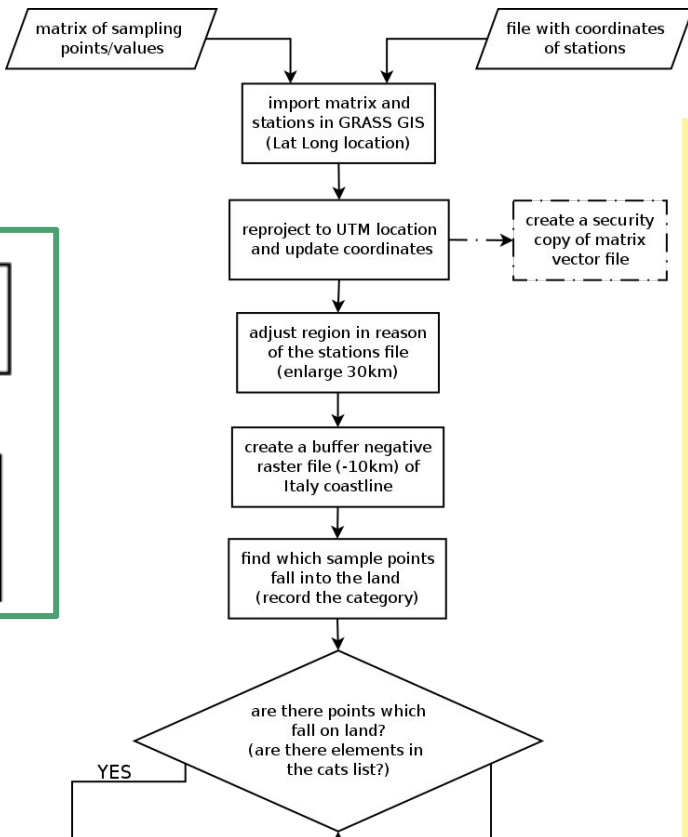
# Armonizzazione sintattica del database

Lo script Python



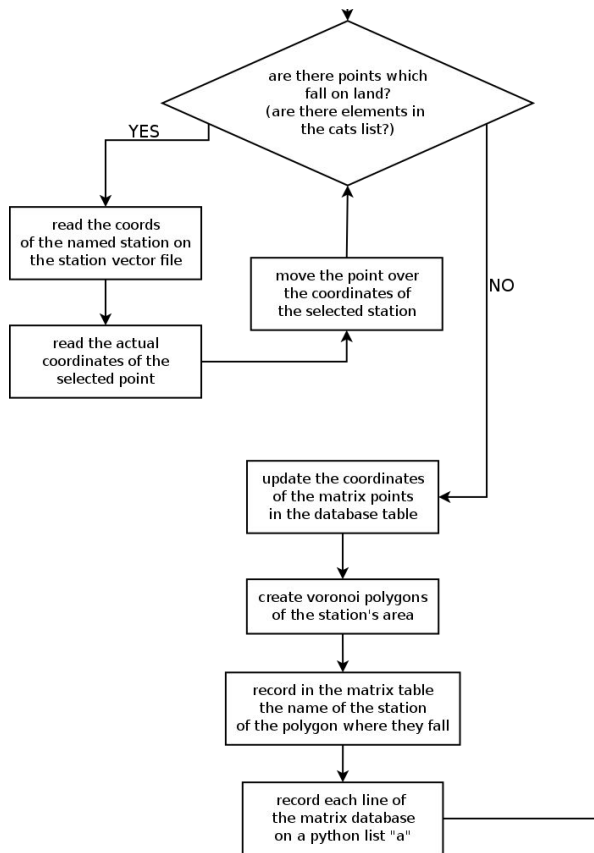
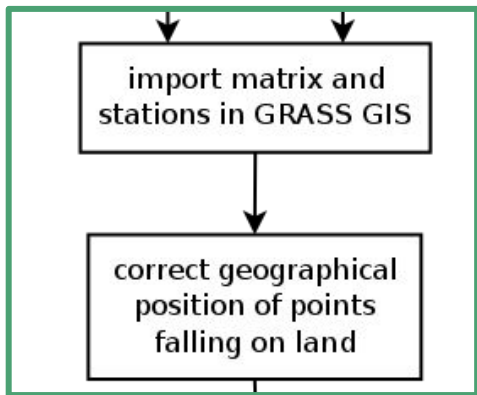
# Armonizzazione sintattica del database

Lo script Python

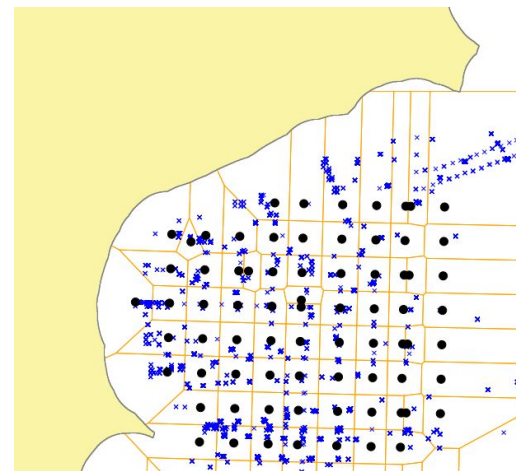


# Armonizzazione sintattica del database

## Lo script Python

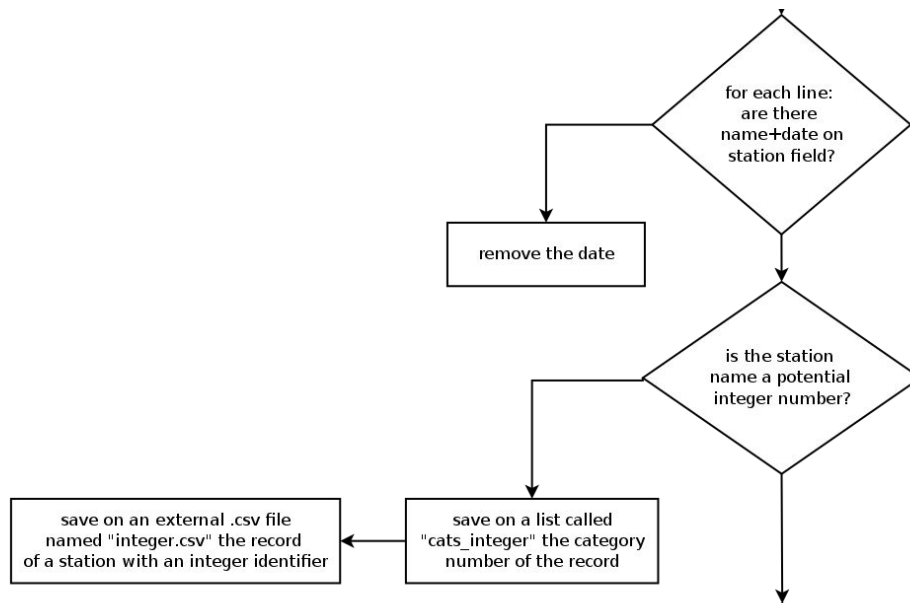
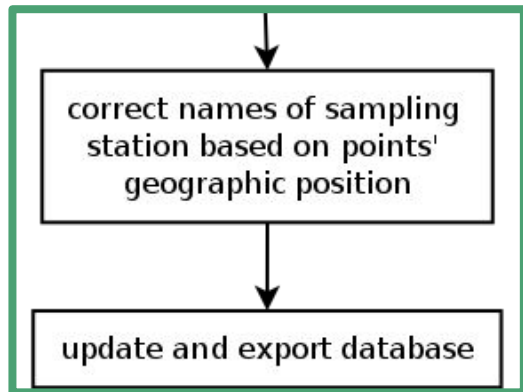


long	lat	name	periods
12:29E	45:24N	A	summer_1965,autumn_1965
12:52E	45:16N	C	summer_1965,autumn_1965
12:46E	45:30N	1/1	summer_1965,autumn_1965
13:14E	45:20N	3/2	summer_1965,autumn_1965
12:46E	45:20N	Y/2	autumn_1965
12:46E	45:25N	04/2A	1978,1979
13:21E	44:45N	09/2E	1978,1979
13:13E	44:30N	08/0G	1978,1979



# Armonizzazione sintattica del database

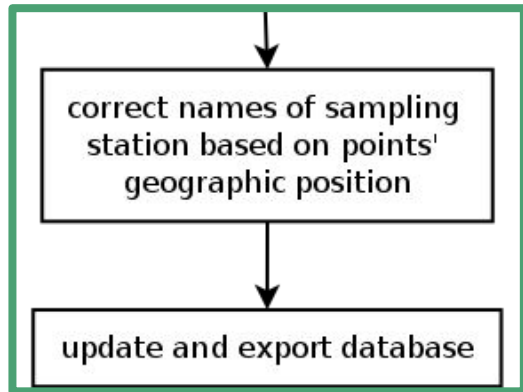
Lo script Python



N	CRUISE	STATION	SHIP	dd/mm/yy	hh:mm	Long [degrees_east]	Lat [degrees_north]
2814	PFP.04	03.0A_01/12/1978	DANCONA	01/12/1978		12.65	45.49
3184	PFP.05	03.1F_17/05/1979	DANCONA	17/05/1979	10.00	12.65	44.62
3185	PFP.05	03.1F_17/05/1979	DANCONA	17/05/1979	10.00	12.65	44.62
3186	PFP.05	03.1F_17/05/1979	DANCONA	17/05/1979	10.00	12.65	44.62

# Armonizzazione sintattica del database

Lo script Python



GRASS GIS Attribute Table Manager - <matrice\_3D\_matching@ismar72>

1 / Table matrice\_3D\_matching

Attribute data - right-click to edit/manage records

long	lat	depth	progressive	cruise	station	boat	date	time
13.4666666667	44	-0.5	805	PAAL...	2/10	BANNOCK	01/02/1966	13.2
13.4666666667	44	-1	806	PAAL...	2/10	BANNOCK	01/02/1966	13.2
13.4666666667	44	-5	807	PAAL...	2/10	BANNOCK	01/02/1966	13.2
13.4666666667	44	-10	808	PAAL...	2/10	BANNOCK	01/02/1966	13.2
13.4666666667	44	-20	809	PAAL...	2/10	BANNOCK	01/02/1966	13.2
13.4666666667	44	-30	810	PAAL...	2/10	BANNOCK	01/02/1966	13.2
13.4666666667	44	-40	811	PAAL...	2/10	BANNOCK	01/02/1966	13.2
13.4666666667	44	-50	812	PAAL...	2/10	BANNOCK	01/02/1966	13.2
12.4833333333	45.4	-0.5	883	PP/19	A	VERCELLI	01/04/1966	7.09

SQL Query

Simple Builder

SELECT \* FROM matrice\_3D\_matching WHERE

cat

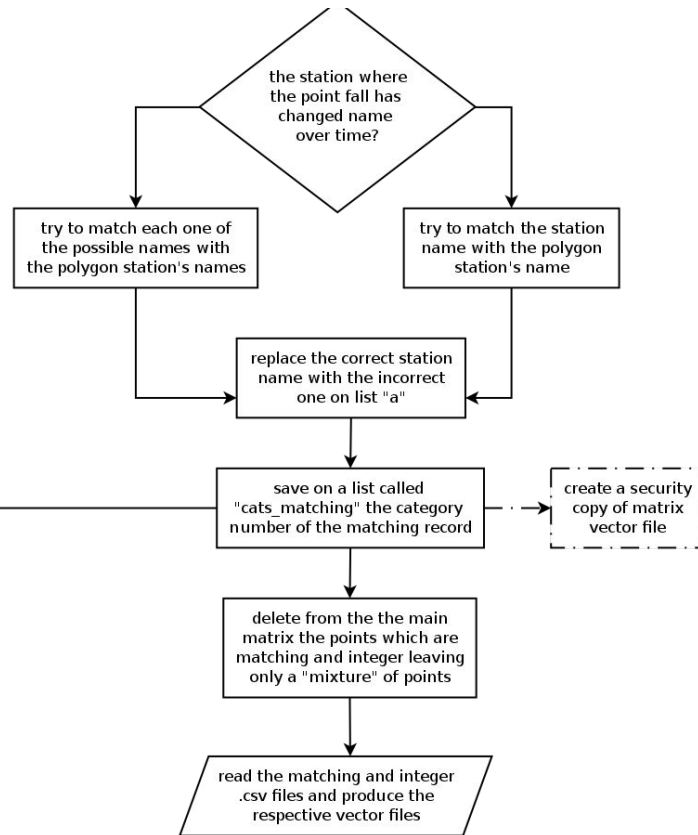
Apply

Browse data | Manage tables | Manage layers

Number of loaded records: 4262



fonte: <https://pixabay.com>

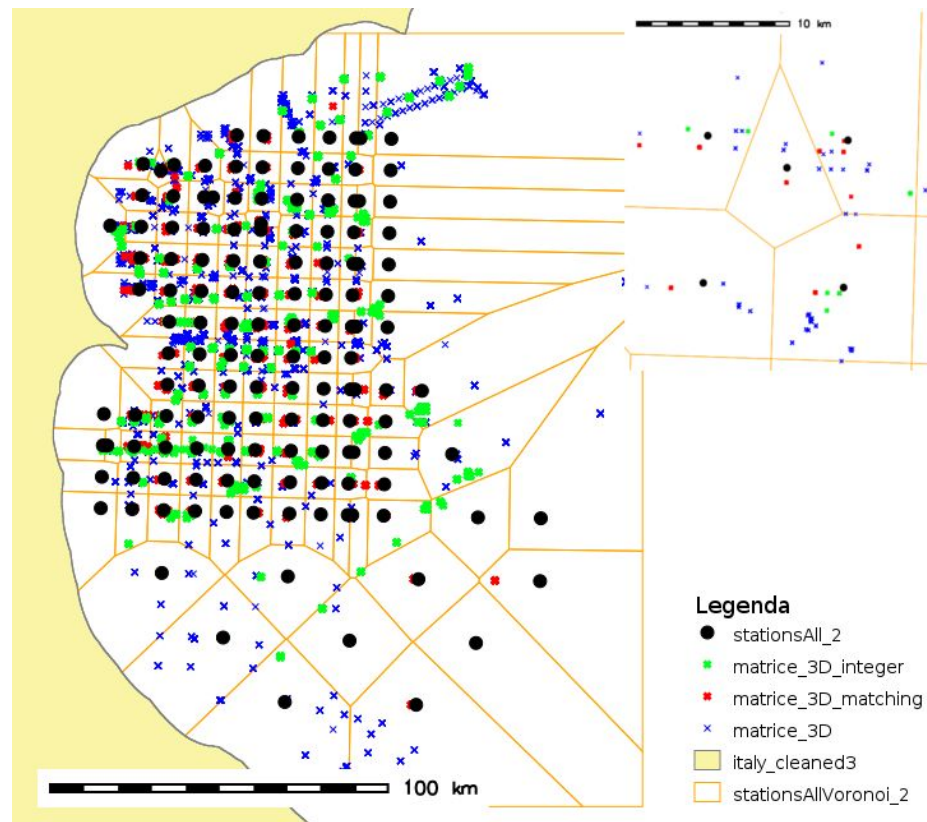
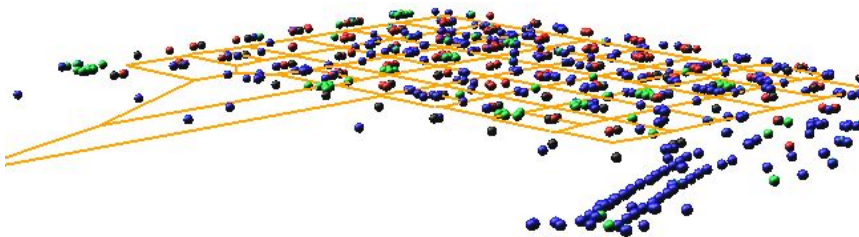




# Armonizzazione sintattica del database

## Gli output:

- 2 shapefiles 3D con nomi delle stazioni corretti nel database
  - nomi delle stazioni combacianti con quelli dei punti di campionamento “predefiniti”
  - nomi delle stazioni sono “numeri interi”
- 1 shapefile 3D dei restanti punti (nomi delle stazioni vari o non presenti)



# Armonizzazione sintattica del database

## Osservazioni:

- i punti campionati hanno nomi delle stazioni combacianti col nome delle stazioni standard fino a circa i primi anni '90 (introduzione del GPS a bordo)
- i punti campionati su stazioni standard rappresentano circa il 7% del campione (~4200 punti)

## Prospettive:

- l'armonizzazione sintattica degli altri parametri è auspicabile
- sarebbe interessante utilizzare (rappresentazione e analisi) il dato nella sua completezza, il dato è infatti in 4 dimensioni: X, Y, Z, t.

# Invito all'interoperabilità

## Condivisione VS Funzionalità



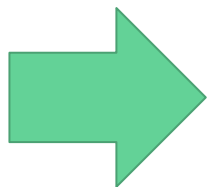
### Osservazioni:

- condividere non deve significare un ingiustificato aggravio di lavoro per i ricercatori
- proposta di processo: non si devono stravolgere le abitudini dei ricercatori

# Invito all'interoperabilità

## Uso interno:

- il database non è grande: può essere ancora gestito tramite **fogli di calcolo**
- per visualizzare rapidamente dati oceanografici e preparare output con una precisa formattazione si può continuare ad usare **Ocean Data View**
- per visualizzare l'evoluzione temporale di alcuni parametri può essere utile l'uso di **InfluxDB**, software di storage, analisi e visualizzazione di serie temporali



per rendere veramente “open” il nostro lavoro è utile aprire i dati (e tutto il processo di ricerca) ad un pubblico più ampio utilizzando standard e formati **condivisi** dalla comunità scientifica ed **interoperabili**

# Invito all'interoperabilità

## Cosa propone RITMARE:

- un database **PostGres**
- un servizio **SOS** a cui i dati sono raggiungibili, scaricabili ed interrogabili
- un client: **GET-IT**
- un editor di metadata: **EDI**

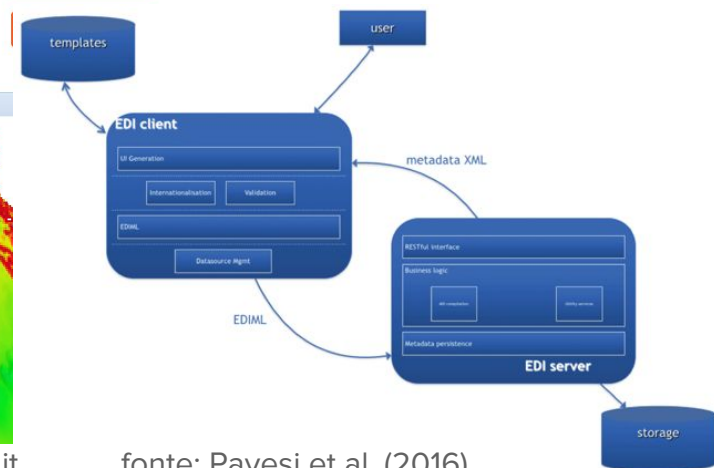


ID	Google Title	Type
3621	"SOS"	sos
7265	"SOS"	sos
3595	"SOS"	sos
1233	"cormp.ocpl.buoy OGC:SOS 1.0.0 NONE NONE http://secoora..."	sos
7278	"SOS"	sos
3597	"SOS"	sos
1232	"enp.canfl.met OGC:SOS 1.0.0 NONE NONE http://secoora.or..."	sos
3614	"SOS"	sos
3594	"SOS"	sos
7276	"SOS"	sos
7273	"SOS"	sos
110...	"SOS - It works - developer! 139.124(squid)"	sos
9255	"SOS"	sos
7280	"SOS"	sos
7282	"SOS"	sos

da: <http://bolegweb.geof.unizg.hr>

A screenshot of the RITMARE web interface. The top navigation bar includes 'HOME', 'LAYERS', 'SENSORS', 'DOCUMENTS', 'VIEWS', 'PEOPLE', and 'SEARCH'. Below the navigation is a search bar. The main content area is titled 'RITMARE - DEMO SERVIZI OSSERVATIVI' and displays a map with a color-coded overlay (red, yellow, green) representing data. A table of metadata is visible on the left side of the map area.

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



fonte: Pavesi et al. (2016)

# Grazie!

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).
- Pavesi, Fabio, et al. "EDI—A Template-Driven Metadata Editor for Research Data." *Journal of Open Research Software* 4.1 (2016).

