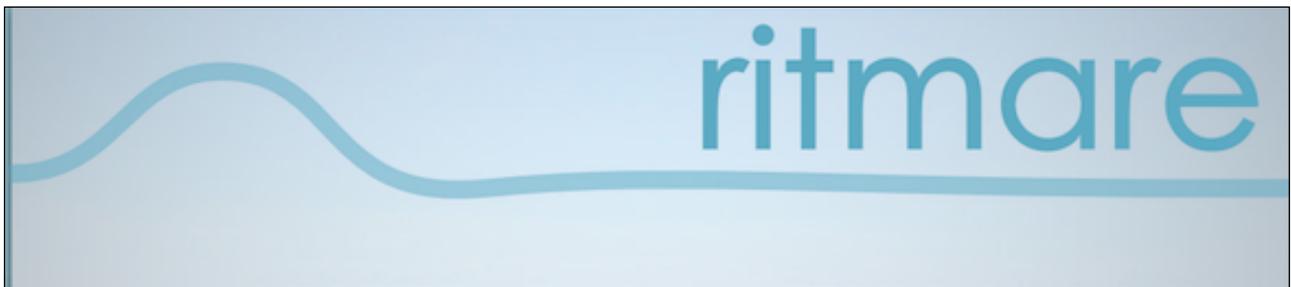

Progetto Ritmare

Utenti, Widget, uso - Possibile flusso di visualizzazione e gestione dati

Leonardo Marotta - 3 February 2017



Introduzione

Contratto di RICERCA Università IUAV di Venezia e CNR –IREA
Progetto RITMARE, Sottoprogetto 7

Descrizione di un possibile flusso informativo

Il widget di visualizzazione serve per visualizzare i risultati della ricerca Ritmare. Questi widget forniscono un mezzo per visualizzare i dati in immagine o forma grafica. Dai widget di visualizzazione deve essere possibile inserire nei documenti l'output di visualizzazione, o inserirli in altre applicazioni delle applicazioni utente (come in foglio di calcolo, in un database), o inserirli in applicazioni create con altri widget di Ritmare.

Il widget di visualizzazione deve avere le seguenti rappresentazioni minime che si citano di seguito:

Grafico lineare con unione punti; un grafico bidimensionale che mostra una sinistra a destra serie di punti di dati, con ogni due punti di dati consecutivi collegati da una linea retta. Anche se grafici lineari possono essere associati a qualsiasi tipo sfaccettatura, sono in genere utilizzati per mostrare una serie storica, con la linea che rappresenta il movimento cronologico.

Grafico a torta: un cerchio bidimensionale, diviso in diverse sezioni colorate, una per ogni valore di dati. percentuale di una sezione della superficie del cerchio è equivalente alla percentuale del contenuto del valore dei dati relativi all'intero dei dati (per esempio, se ci sono 16 riferimenti totali, e 8 di loro hanno valore 'temperatura t1', la sezione di grafico a torta corrispondente al valore 'temperatura t1' sarà la metà del cerchio). I valori dei dati possono essere sia non continui (salinità, plankton) o continui (salinità, temperatura, ossigeno disciolto).

Istogramma o grafico a barre: un grafico con barre rettangolari orizzontali o verticali che rappresentano la quantità di un particolare valore discreto. I valori possono essere sia non continui (salinità, plankton) che continuo (salinità, temperatura, ossigeno disciolto).

Punto su Mappa: una mappa (con varie possibilità ad es. di Google Earth) con i punti di dati che rappresentano i risultati di ricerca, ciascuno associato ad un valore geospaziale. Per esempio, supponendo che i dati sono nel database e il widget di cartografia è stato configurato per questo, un punto si trova nell'adriatico, per indicare il luogo di campionamento. È possibile ingrandire e ridurre la mappa, così come aprirla (Map pan) in tutte le direzioni.

Mappa di frequenza spaziale: una mappa con una frequenza geografica. I risultati possono essere rappresentati ad esempio come nuvole di punti o come chiazze colorate (attraverso una funzione geostatistica di media di punti per unità spaziale) situati in punti distinti sulla mappa. Per esempio, se ci fosse un gran numero di risultati, come dati di campionamento o sensori, si visualizzano una grande macchia più scura su quella posizione sulla mappa con dei colori che sfumano verso l'assenza di dati. La scelta possibile è l'unità spaziale di analisi.

Il problema della rappresentazione dei dati è che alcune rappresentazioni complesse come i digrammi ternari, i diagrammi ad albero, o l'analisi multiparametrica, o l'analisi di network o reti, ma anche la valutazione dell'errore da inserire nel diagramma richiedono funzionalità aggiuntive. Per questo gli scienziati e tecnici che saranno utenti di Ritmare devono utilizzare da subito una serie di software per visualizzare e trasformare i dati grezzi ed eseguire analisi statistiche su di essi.

Per cui riteniamo dal punto di vista dell'usabilità che non sia possibile sviluppare un widget di visualizzazione senza avere un widget di elaborazione collegato.



Alcuni strumenti per l'analisi dei dati sono già disponibili in R, che riteniamo l'ambiente di calcolo statistico open source più utile per lo scopo, inoltre la nostra analisi su 22 scienziati marini lo identifica come il più usato (anche se solo dal 15% del campione degli intervistati). Tuttavia, l'uso di R richiede la conoscenza minima di programmazione e non è di facile uso.

Esistono alcune utili interfacce:

- i. Red-R (<http://www.red-r.org>) è una interfaccia visiva open source user-friendly del flusso di dati basata su R. L'interfaccia utilizza principi della programmazione "pipeline programming" in cui le funzioni sono rappresentate da icone, widget, e flussi di dati.
- ii. Per la gestione dati si può usare anche Orange (<http://orange.biolab.si>) o per la visualizzazione rstudio (<https://www.rstudio.com>)

Red-R è in grado di svolgere compiti di analisi di dati comuni (test di ipotesi, analisi della varianza, regressioni, analisi delle componenti principali, trame di dati cloud, diagrammi a barre, biplot, etc.). E' stato sviluppato un pacchetto di Red-R che implementa il pacchetto R. Questo può essere utilizzato per eseguire operazioni sui dati composti, su dati grezzi e creare visualizzazioni ed analisi composte (come i diagrammi ternari, dendrogrammi, ecc) senza conoscenze di programmazione, dopo alcune operazioni di base. La rete di widget fornisce una procedura di facile seguire passo dopo passo per eseguire un gran numero di operazioni disponibili in R, e quindi facilitare i compiti di chi vuole si visualizzare che eseguire analisi di dati. Inoltre, l'intero flusso o rete di analisi possono essere salvati e riutilizzati. I report possono essere generati dalla rete widget per documentare e condividere i risultati. Questo pacchetto fa in modo che i non programmatori possano avere un facile accesso agli strumenti avanzati disponibili.