Buone pratiche dal mondo OSS ?

Federico Giuntoli - UCS Livorno

Terni, 18 giugno 2015

il problema

una soluzione possibile ?

prima valutazione della soluzione ipotizzata

il problema

lista dei desideri. . .

Elenco dei lavori statistici che vorremmo fare :

- tavole di mortalità
- tavole di fecondità
- migrazioni
- proiezione demografica
- previsioni demografiche
- la città ai censimenti 91-01-11 - edifici ed alloggi
- la città ai censimenti 91-01-11 - popolazione
- la città ai censimenti 91-01-11 - lavoro ed istruzione
- la città ai censimenti 91-01-11 - pendolarismo
- la città ai censimenti 91-01-11 - imprese ed addetti
- la città ai censimenti 91-01-11 - agricoltura
- la città ai censimenti 91-01-11 - deprivazione

lista dei desideri. . . (continua)

- redditi - individui e famiglie
- povertà e disuguaglianze
- stranieri
- statistiche di genere
- incidenti
- nuzialità, separazioni e divorzi
- flussi elettorali
- prezzi
- BES
- pressione fiscale
- altro ?

ma tra il dire ed il fare. . .

- la struttura organizzativa del nostro UCS non e' adeguata per la realizzazione della maggior parte dei "desideri" elencati
- il carico istituzionale futuro (censimento permanente ?) toglierà probabilmente risorse a questo tipo di lavori
- sopratutto : necessità ogni volta di "**re-inventare la ruota**" per ogni nuovo prodotto/desiderio

risultato : sensazione di "**isolamento**"

e gli altri UCS del Sistan ?

Se ampliamo lo sguardo da una singola città all'insieme dei Comuni :

- città metropolitane - grandi Comuni : quasi sicuramente non hanno nessun problema (hanno risorse e know-how)
- medi Comuni : come noi ? forse in situazione migliore ?
- piccoli Comuni : forse hanno ulteriori ostacoli ?

risultato : il "sistema globale" degli UCS ha **capacità produttiva ridotta** e limitata solo ad alcuni Comuni

una soluzione possibile ?

idea : mutuiamo l'esperienza del mondo OSS

- da parecchi anni, gli sviluppatori del software open source collaborano a (grande) distanza mediante piattaforme tecnologiche condivise
- l'idea e' : proviamo a replicare la loro organizzazione di lavoro
- possiamo pensare ai nostri lavori statistici come :

(lavoro/rapporto/studio/analisi)= desiderio = **dati + codice**

- i dati sono nostri, ma spesso hanno formati comuni
- il codice potrebbe essere condiviso

idea : mutuiamo l'esperienza del mondo OSS (continua. . .)

- allora una possibile soluzione e' :

(sogno =) un deposito congiunto di "**semi-lavorati statistici**" disponibile in rete, da cui prelevare secondo necessità e da ri-applicare ai nostri dati

- requisito fondamentale : **costo nullo**
- perchè **semi-** : se il semi-lavorato scaricato dal deposito mi fa risparmiare il 70-75-80% del tempo di lavoro totale, va bene; il restante 30-20% sara' personalizzazione (riduzione, ampliamenti, modifiche, miglioramenti. . . .)
- il prototipo iniziale del server "deposito" può essere ospitato da un UCS
- l'accesso al deposito e' consentito ai soli membri UCS/Sistan (almeno per una prima fase), con credenziali fornite a richiesta

componenti di un semi-lavorato

Un semi-lavorato dovrebbe/potrebbe contenere :

- codice (script principale, parametri di elaborazione + eventuali altri scripts)
- documentazione testuale operativa, compreso le istruzioni di esecuzione
- eventuali dati comuni di servizio (ad es. : classificazioni)
- eventuali dati di partenza : solo fittizi, creati ad arte, a scopo di test
- eventuale bozza schematica di rapporto finale

convenzioni necessarie

Saranno opportune/necessarie alcune convenzioni di lavoro :

- tutto il materiale di un semi-lavorato e' contenuto in una **specifica directory** del server "deposito"
- la directory specifica e' organizzata con una struttura largamente comune
- ogni semi-lavorato e' interamente automatizzato, ottenuto tramite l'elaborazione di "codice statistico (script)" : analisi, grafici e tabelle sono ottenuti dalla corsa dello script
- il codice dovrebbe essere, per quanto possibile, auto-documentante
- esiste uno script principale che esegue e/o coordina l'elaborazione
- i dati di partenza (microdati o altri) non sono mai editati
- tutte le manipolazioni/preparazioni dei dati sono eseguite dal codice

componenti tecnologici

stazione di lavoro individuale (presso UCS)

- linguaggio di elaborazione : R
- ambiente di sviluppo : RStudio (o altro IDE)
- sw per “deposito locale” : Git

server in rete

- gestione del “deposito centrale” con interfaccia web : GitLab (community edition)

perchè R

- perche' il sistema funzioni (ed i semi-lavorati possano essere utilizzati da altri), occorrono strumenti comuni
- strumenti non costosi (requisito iniziale)
- R e' OSS, diffuso, maturo, con tante estensioni
- normativa C.A.D. sulle soluzioni software per P.A.

prima valutazione della soluzione ipotizzata

rischi potenziali

principalmente uno : rimane una **esperienza isolata**

quindi **irrilevante a livello di sistema**

vantaggi potenziali

miglioramento della capacità complessiva del sistema

- disponibilità di una sorta di "ufficio virtuale" per ogni UCS
- possibile aumento della capacità produttiva della maggior parte degli UCS
- miglioramento della qualità dei prodotti (tramite l'osservazione del "come" gli altri lavorano); per tutti ?

scenari di utilizzo

- piccolo UCS : dall' "ufficio virtuale" può ri-applicare alcuni semi-lavorati, anche solo ad un livello minimo di funzionalità
- medio UCS : può sfruttare l' "ufficio virtuale" per aumentare la capacità produttiva e di analisi
- grande UCS : capacità di sfruttare eventuali miglioramenti nella propria produzione, sulla base degli apporti dagli altri UCS

requisiti per il successo ?

Prime idee su condizioni e fattori necessari per il successo della soluzione ipotizzata :

- un prototipo di server funzionante con almeno 1 semi-lavorato (anche banale) a scopo dimostrativo e di test
- 2-3 UCS cha avviano il processo, ciascuno caricando un proprio semi-lav
- altri 3-5 UCS che "testano/verificano" i semi-lavorati ospitati nel deposito
- **spirito di collaborazione**
- accettazione che colleghi di altre città vedano non solo cosa facciamo, ma anche **come** lavoriamo
- adeguamento a metodiche di lavoro condivise
- creato e testato il primo nucleo di semi-lavorati, opportune iniziative di diffusione (eventi USCI ? blog dedicato ? ....)

a che punto siamo ?

- prototipo di server : in allestimento presso UCS di Livorno (attualmente : visibile solo internamente, sono in corso tests di configurazione e sicurezza)
- semi-lavorato dimostrativo : quasi pronto (legge un file csv estratto da 8milaCensus, prepara i dati, costruisce 1 tabella e 1 grafico)

conclusione

- qualche UCS e' disposto a collaborare ?
- domande ?
- suggerimenti su parola alternativa per “semi-lavorato” ?
- contatto :

mail = `f.giuntoli@comune.livorno.it`

telefono = 0586 820 461