

I Sistemi Informativi Territoriali liberi: un'opportunità tecnica e commerciale per i biologi italiani

E. Venturato, L. Lami, R. Petrini, P. Cavallini

Faunalia, Pontedera (PI)

Introduzione

La nostra specie vive a densità, e soprattutto con valori di biomassa, elevatissimi e crescenti su vaste aree del pianeta. Non è quindi un caso che lo spazio vitale e la sua gestione suscitino un interesse sempre maggiore; oggi le informazioni sul territorio, ai fini del suo monitoraggio e pianificazione, hanno un'importanza cruciale in molti processi. A conferma dell'attualità dell'argomento sono da considerare i notevoli investimenti che, le Amministrazioni Pubbliche in primo luogo, effettuano in questo campo, con interi settori dedicati (ad es. <http://www.rete.toscana.it/sett/territorio/carto/>).

In particolare in campo biologico i Sistemi Informativi Territoriali (in acronimo SIT, oppure GIS, per *Geographic Information System*) risultano utili per effettuare analisi approfondite (e visualizzazioni dei risultati) in numerosi ambiti, quali ad esempio:

- lo studio dell'impatto ambientale di una determinata opera o progetto, tramite la valutazione della vicinanza territoriale con aree protette, aree a vincolo paesaggistico, a pericolosità idraulica, ad ecosistemi ed emergenze naturalistiche da tutelare;
- la realizzazione di mappe di uso del suolo per ricerche di dinamica di popolazione, uso dell'*habitat*, risultati di censimenti e di radiotracking su specie animali e vegetali;
- la produzione di carte tematiche su indici di variabilità o qualità ambientale, vegetazione effettiva e potenziale, valore naturalistico;
- la realizzazione di mappe delle emergenze naturalistiche animali, vegetali e più in generale degli ecosistemi in una data provincia o regione; la realizzazione di mappe dei rilevamenti di specie rare o comunque oggetto di interesse.

Fra i biologi, l'utilizzo dei SIT è stato limitato, da una parte dalla scarsità delle competenze necessarie, spesso assenti dal *curriculum* universitario, dall'altra dall'elevato costo per l'acquisto delle licenze (varie decine di mi-

gliaia di euro).

Questa limitazione nell'uso dei SIT ha causato una serie di effetti negativi, alcuni con forti *feedback* ad es.:

- l'Università spesso non è in grado di acquistare un congruo numero di licenze, e conseguentemente di fornire una formazione efficace;
- anche quando i *software* siano disponibili nelle aule, gli studenti non hanno la possibilità di studiare autonomamente i programmi (in quanto non possono utilizzarli se non sui *computer* dotati di licenza e installati nelle aule laboratorio, in generale poco accessibili);
- i professionisti trovano difficile acquisire nuove competenze, in quanto mancano nuove professionalità nel settore, e in ogni caso la barriera d'ingresso (rappresentata soprattutto dal costo delle licenze) inibisce l'innovazione.

Questa situazione porta ad una generale perdita di senso etico per la pratica diffusa di copiare i programmi protetti da *copyright* (tema di grande attualità) e, conseguentemente, per la permanenza in situazioni di illegalità sanzionabili.

La maggior parte dei biologi si è quindi orientata su prodotti più economici e più facili da usare, almeno ad un primo approccio. Anche questa scelta non è priva di conseguenze negative, in quanto limita lo sviluppo di consulenze di qualità, e spesso pone i biologi in posizione subordinata rispetto ad altre categorie con più forte *background* (si veda ad es. quanto accade nel settore delle valutazioni di impatto ambientale).

I Sistemi Informativi Territoriali

I principali elementi di un SIT sono:

1. un sistema di immagazzinamento e gestione dei dati geografici (raster e vettoriali; Box 3)
2. un sistema per l'analisi e la modellistica ambientale
3. un sistema per l'interattività *desktop* (del singolo utente)

4. un sistema per l'interattività *web* (tramite internet). Mentre per le funzioni 1 e 2 le caratteristiche più importanti sono la potenza e l'affidabilità, per gli elementi 3 e 4 la facilità d'uso è il parametro fondamentale. A livello commerciale, un sistema completo di questo tipo ha un costo che supera gli 80.000 € (proibitivo nella maggior parte dei casi). Risulta quindi indispensabile valutare con attenzione quali alternative esistano.

I SIT liberi

La migliore via d'uscita da questa *impasse*, a nostro avviso, è rappresentata dal *software* libero (Box 1). Sono infatti oggi disponibili programmi che, senza il costo di acquisizione di una licenza, danno la possibilità di effettuare analisi, anche molto complesse, su grandi moli di dati. E' disponibile, anzi, un intero sistema operativo completamente libero utilizzabile, oltre che per i SIT, per tutte le altre applicazioni di analisi e di ufficio (Box 2). Una delle caratteristiche del *software* libero è di avere molte alternative per raggiungere gli stessi obiettivi; risulta quindi difficile ad un primo approccio orientarsi. Caratteristica che può, inizialmente, creare sensazioni di complicatezza, ma che mostra presto tutti i propri vantaggi è la modularità dello sviluppo di questi software. In generale, non si tende ad ottenere un unico programma che faccia tutto, dallo stoccaggio dei dati, alle analisi statistiche, a quelle territoriali, alla stampa, fino alla presentazione *web* dei risultati. Piuttosto il lavoro è ripartito in modo da fare sviluppare a programmi specifici (database, programmi di statistica, SIT vero e proprio, ecc.) le proprie funzionalità nel modo più appropriato e ottimale. Il lavoro di coordinamento tra i vari gruppi di sviluppo permette l'integrazione di tutto questo, grazie all'utilizzo di formati *standard* per lo scambio dei dati e, in vari casi, l'uso di un programma e dei suoi specifici linguaggi, direttamente dall'interno di un'altra applicazione così da far girare entrambi i *software* in modo integrato. In base alla nostra esperienza, presentiamo qui di seguito le applicazioni più efficaci per la messa a punto di un SIT funzionale, interamente basato su *software* libero, e pienamente compatibile con gli *standards* richiesti dalle pubbliche amministrazioni.

Gestione dati geografici

Il trattamento di grandi moli di dati (raster e vettoriali; Box 3), quali le informazioni geografiche, richiede un approccio adeguato; mentre i dati di tipo raster sono di solito gestiti direttamente dall'applicazione SIT principale, i dati vettoriali possono anche essere gestiti da un database relazionale, che tratta in maniera appropriata le rispettive proiezioni (Gauss-Boaga, UTM, latitudine-longitudine, ecc.); all'interno del database, è possibile effettuare le più comuni operazioni SIT (intersezioni, selezioni in base alla posizione, buffer, ecc.). Per questi scopi, la soluzione più adeguata è il database PostgreSQL (www.postgresql.org) con l'estensione spaziale PostGIS/GEOS. All'interno di PostgreSQL vengono gestiti i dati di tipo alfanumerico (valori e caratteri associati ad un determinato elemento geografico, quali il nome, la categoria, l'ente responsabile, ecc.). L'estensione PostGIS (in grado di leggere il formato *shapefile* di ESRI, lo *standard* del settore) permette inoltre lo stoccaggio anche dei dati geografici nel database relazionale (fig. 1).

Analisi ambientale

Questo è il "cuore" del sistema; il programma libero in assoluto più potente e versatile è GRASS (*Geographic Resources Analysis Support System*: www.grass.itc.it), nato oltre venti anni fa come SIT raster, oggi incorpora anche una sezione vettoriale topologica di grande potenza. Ha una struttura modulare (sono presenti centinaia di moduli, da quelli di base che si occupano dell'importazione ed esportazione dei dati fino a modelli molto specializzati, come ad esempio quelli che stimano parametri fondamentali dell'ecologia del paesaggio; una lista delle loro funzionalità si trova su: grass.itc.it/capabilities.html). Si collega ad un potente programma statistico, R (www.r-project.org) per effettuare analisi geostatistiche. È capace di: convertire raster e vettori fra di loro, e fra sistemi di coordinate diverse; importare dati da GPS; analizzare altitudini, pendenze ed esposizioni; realizzare visualizzazioni in tre e quattro dimensioni (anche animate, producendo filmati mpeg); ecc. Per le analisi complesse e ripetute, è particolarmente utile la possibilità di salvare le sequenze di comandi, che poi possono essere eseguite più volte, variando alcuni parametri o i dati a cui sono applicate.

oid	cat tipo	nome	the_geom
3577738	2 PR	ALPI APUANE	SRID=26591;MULTIPOLYGON(((1598978.98694221 488742...
3577784	34 RP	Riserva Naturale Acquerino-Cantagallo	SRID=26591;MULTIPOLYGON(((1666132.43834085 487210...
3577797	44 AP	ANPIL Lago e rupi di Porta	SRID=26591;MULTIPOLYGON(((1593815.57663773 487218...

Fig. 1 - Un estratto dei dati geografici contenuti in un database PostgreSQL con estensione PostGIS. I dati geografici sono stoccati come poligoni, multipoligoni, linee, punti (a seconda del tipo di dato) descritti con le coordinate xy dei punti o vertici della figura geometrica. Il sistema di coordinate geografiche da usare è contenuto nel codice SRID e permette l'eventuale conversione tra sistemi differenti.

L'uso approfondito delle funzione di GRASS permette di costruire, sulla base di dati disponibili, modelli predittivi importanti per tutti quei biologi impegnati nella gestione delle risorse territoriali e validi supporti alle decisioni degli enti che gestiscono il territorio.

"Usare" il SIT

L'uso più frequente (anche se non il più produttivo) dei SIT è quello "desktop", in cui l'utente seleziona le informazioni di interesse, attribuisce appropriate caratteristiche grafiche (colori, trasparenze, ecc.) e salva o stampa la mappa risultante. In questo settore il software libero aveva storicamente un ritardo rispetto agli analoghi commerciali, ma nel corso del 2004 un'applicazione (QuantumGIS: www.qgis.org) ha iniziato a colmare la lacuna. Si tratta di un programma ancora giovane, ma in sviluppo molto rapido, ed è previsto che nel corso dei prossimi mesi possa acquisire tutte le funzionalità necessarie per un uso professionale.

Mappe su Internet

Con l'uso crescente, registrato negli ultimi anni, di Internet risulta chiaro come questo mezzo sia una piattaforma ideale sia per la divulgazione dei risultati di ricerche, che per i rapporti con i cittadini, che per il coordinamento intra- e intergruppi di ricerca, ecc. I dati geografici non fanno eccezione in questo senso: la possibilità di visualizzare, interrogare, selezionare le informazioni in tempo reale è ormai indispensabile in un'ampia varietà di ambiti. In questo campo Mapserver (programma per la restituzione web di elaborazioni cartografiche sviluppato dall'Università del Minnesota: <http://mapserver.gis.umn.edu/>) consente di realizzare applicazioni anche molto complesse. Gli utenti possono visualizzare le mappe scegliendo i temi che interessano loro, l'area geografica specifica, la scala di visualizzazione, ecc., ed essere informati in tempo reale su fenomeni anche molto dinamici (una delle più grosse applicazioni, visitata da migliaia di utenti contemporaneamente, è quella che descrive i terremoti della faglia di S. Andrea, in California: <http://quake.usgs.gov/recenteqs/latestfault.htm>; fig. 2).

Alcuni esempi di analisi

1. Analisi della variabilità ambientale

Un primo esempio di uso dell'elaborazione territoriale è l'analisi della variabilità ambientale basata sulla misura locale dell'indice di Shannon. A partire dall'uso del suolo attribuito tramite l'analisi della cartografia aerea e sopralluoghi sul campo (fig. 3) è stato pos-

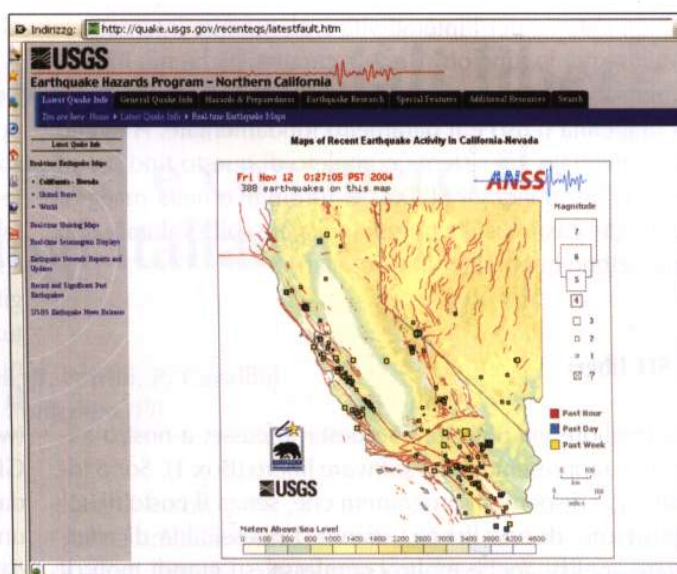


Fig. 2 - Un'immagine del sito che fornisce, in tempo reale, la situazione dei terremoti lungo la faglia di S. Andrea.

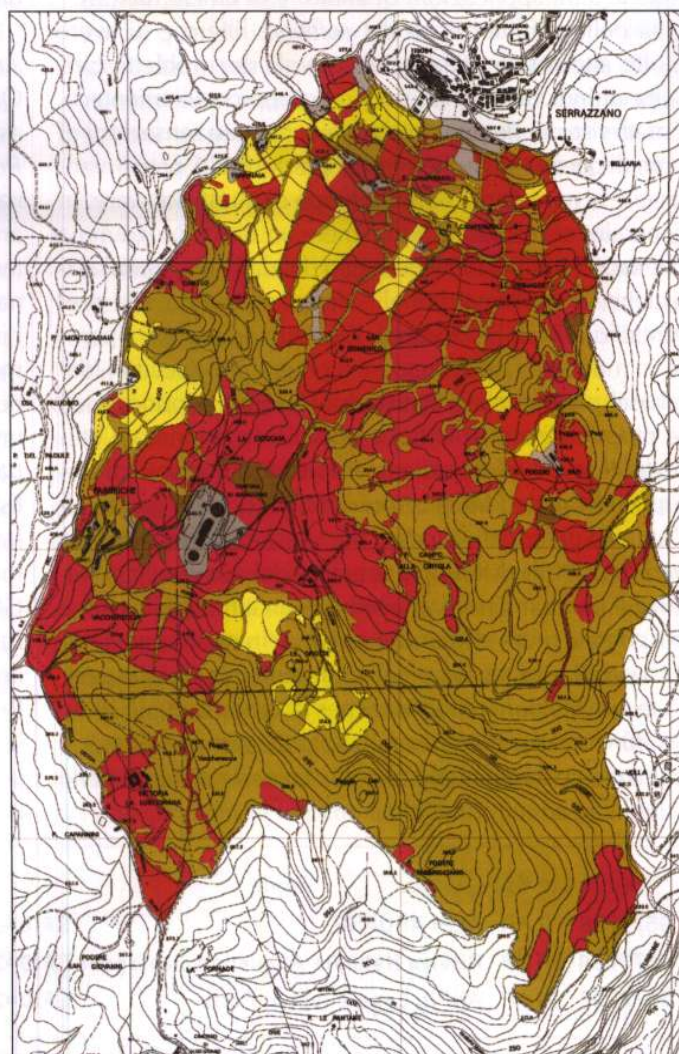


Fig. 3 - Mappa dell'uso del suolo. Ad ogni colore corrisponde un differente codice di uso del suolo attribuito sulla base del codice CORINE.

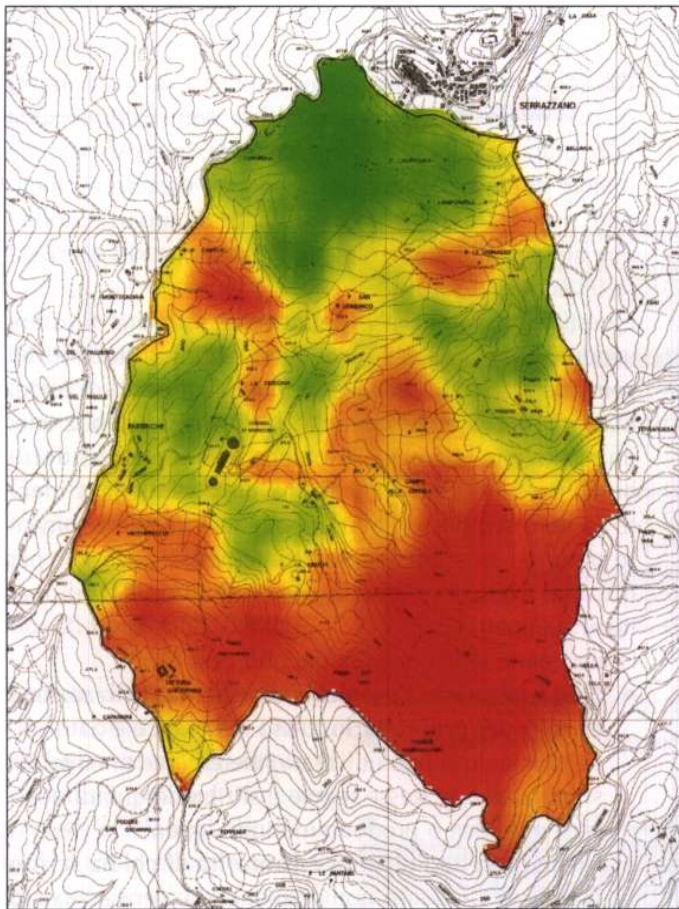


Fig. 4 - Indice di variabilità di Shannon applicato con il metodo della moving window all'uso del suolo. In rosso le aree più omogenee, in verde quelle caratterizzate da maggior diversità ambientale.

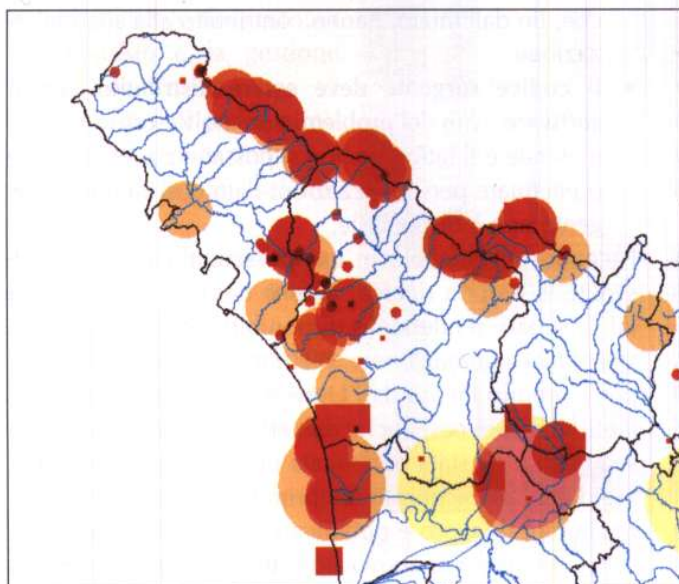


Fig. 5 - L'immagine mostra un esempio dell'analisi geografica svolta a partire da segnalazioni faunistiche e floristiche raccolte con metodi differenti nel corso di numerosi anni. L'elaborazione GIS ha permesso di effettuare una sovrapposizione delle segnalazioni tenendo conto della relativa importanza di ognuna e attribuendo il giusto peso ai differenti dati. Le aree più scure sono quelle in cui la specie oggetto di studio risulta presente con maggior certezza.

sibile determinare l'indice di diversità tramite la funzione "moving window", una tecnica che permette di calcolare, per ogni cella, il valore dell'indice di Shannon in un intorno definito a priori (fig. 4).

2. Distribuzione di specie di interesse conservazionistico

Per questo tipo di studio l'interazione tra GRASS e PostgreSQL ha permesso di analizzare i dati relativi ad un numero elevato di localizzazioni faunistiche e floristiche caratterizzate da estensione e forma dell'area interessata spesso anche molto diverse tra loro. L'analisi ha permesso di sovrapporre tutte le localizzazioni, tenendo conto sia della dimensione che della forma, ed attribuendo ad ognuna di esse un valore normalizzato sulla propria estensione. Questo metodo, oltre a valutare in modo più preciso l'area effettivamente interessata dalla segnalazione, ha permesso di far 'pesare' maggiormente le localizzazioni di estensione minore (più precise) ridimensionando invece l'importanza delle segnalazioni genericamente attribuite ad un'area vasta (fig. 5).

3. Modelli predittivi

A partire dall'analisi statistica della distribuzione dei danni da cinghiale alle colture agricole all'interno delle Zone di Ripopolamento e Cattura è stato possibile ottenere un'equazione logistica multivariata che, applicata alle caratteristiche territoriali di tutta la provincia, ha permesso di ottenere una mappa predittiva delle aree più a rischio di danneggiamento da parte di questa specie (fig. 6).

Conclusioni

Le potenzialità di questi strumenti sono evidenti, e tanto più importanti quanto più la complessità del problema di base si fa elevata. Naturalmente, il loro apprendimento e la messa a punto richiedono un certo investimento in termini di risorse umane (dal punto di vista dell'hardware non ci sono invece particolari esigenze), investimento sicuramente più proficuo di quello profuso in licenze destinate a diventare obsolete in pochi anni. Le competenze acquisite infatti si mantengono e si rivalutano negli anni.

Box 1: Cos'è il software libero?

Ogni software è costituito da un codice scritto in un linguaggio di programmazione che permette al sistema di compiere le operazioni e le funzioni richieste

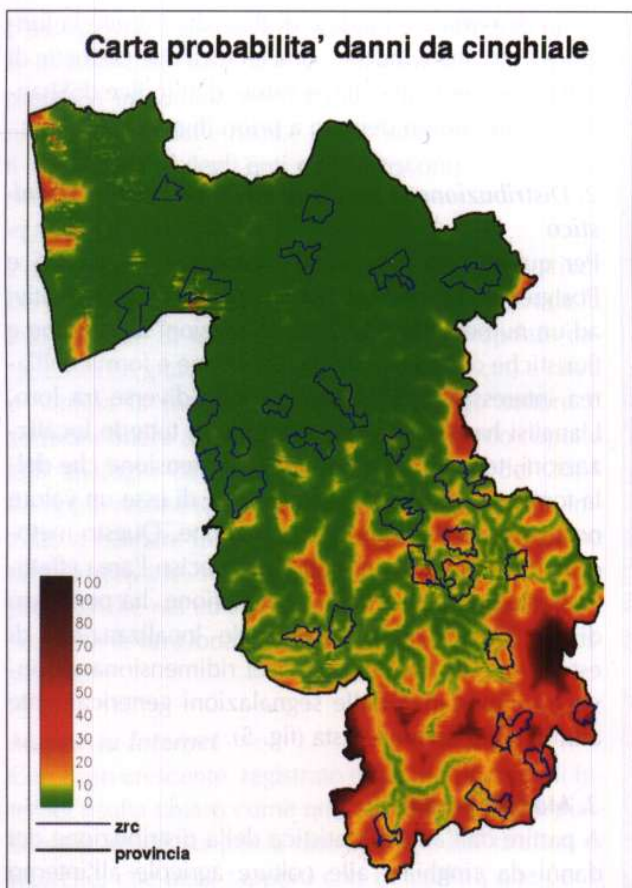


Fig. 6 - Mappa delle probabilità di danno da cinghiale per la Provincia di Pisa

dall'utente. Struttura portante del *software* è quindi il codice che può essere 'free' oppure no; il primo caso si verifica quando il codice viene messo a disposizione insieme al *software* e quindi può essere letto ed eventualmente modificato da chi sia in grado di programmare; è invece 'chiuso' o 'proprietario' quel *software* che viene reso disponibile senza fornire all'utente la possibilità di leggere il codice sorgente.

Quando si parla di *Free Software*, ci si riferisce alla 'libertà', non al 'prezzo'. Spesso si utilizza il termine "Open Source" per indicare sostanzialmente la stessa cosa.

Il *software* pubblicizzato come gratis spesso non è *free*; può essere vietato distribuirlo, e quasi sempre è vietato migliorarlo. *Software* e licenze gratis sono di solito un arma in una campagna di *marketing* per promuovere un prodotto correlato o per portare fuori dal *business* un piccolo concorrente. Non c'è garanzia che rimangano gratis e a disposizione di chiunque intenda usarli. Per il vero *software* libero sono garantite nel tempo, da apposite licenze, le quattro libertà fondamentali:

- la libertà di uso, per qualunque scopo;
- la libertà di studio e modifica per le proprie necessità;
- la libertà di copia e di redistribuzione del programma e delle sue modifiche;

- la libertà di modificare il programma e di distribuire le copie modificate.

Di conseguenza è possibile installare il *software* su quante macchine si vuole; qualsiasi numero di persone può utilizzarlo contemporaneamente; si possono fare quante copie del *software* si vogliono e darle a chiunque; non ci sono restrizioni sul fatto di apporre modifiche al *software*; non c'è alcuna restrizione nel distribuire o vendere il *software*.

Quest'ultimo punto, che permette al *software* di essere venduto sembra essere contrario all'idea globale del *software* libero. In realtà è uno dei suoi punti di forza. Dato che la licenza permette la libera redistribuzione, una volta che una persona ottiene una copia può distribuirla per conto suo o anche provare a venderla. L'utente ha però sempre la libertà di ottenere lo stesso *software* gratis (ad es. scaricandolo da internet); si paga solo per il valore aggiunto (come la comodità di un CD, di manuali stampati, ecc.).

D'altra parte le licenze impongono certi limiti:

- **Non è permesso l'uso del codice per la realizzazione di *software* proprietario.** L'uso del codice è dato sulla fiducia: lo si può utilizzare liberamente fintantoché si utilizzano le stesse regole e si rilascia quindi qualsiasi modifica o prodotto in modo 'open'.
- **Viene protetta l'identità e l'autorità di chi ha scritto il codice.** Gli autori non vogliono che venga rimosso il proprio nome dal *software* che hanno realizzato. Tale *software* quindi, anche se modificato e migliorato deve continuare a portare il nome di tutti coloro che, fin dall'inizio, hanno contribuito alla sua realizzazione.
- **Il codice sorgente deve essere distribuito con il *software*.** Uno dei problemi con molto *software* commerciale è il fatto che non si possano correggere *bug* o effettuare personalizzazioni dato che il codice sorgente non è disponibile.

Facendo un parallelo in campo scientifico, un autore svolge la propria ricerca, ottiene dei risultati, pubblica tali risultati rendendoli disponibili per la comunità scientifica internazionale. Nel far questo viene pagato per il proprio lavoro dall'Università, dagli Enti committenti, in alcuni casi anche dall'editore. Una volta che la sua ricerca sia stata pubblicata un altro autore può partire dalla base acquisita dal primo, ripetere, correggere, migliorare il lavoro e pubblicare a sua volta i propri risultati mettendo di nuovo a disposizione della comunità un tassello verso la comprensione di un dato fenomeno. Chi utilizza il lavoro di un collega può essere a sua volta retribuito per il lavoro svolto ed è tenuto a citare le fonti che gli hanno permesso di procedere nella ricerca. Questa regola è alla base di una qualunque pubblicazione scientifica. La scienza, e più in generale la cono-

scienza, si muovono da sempre secondo queste regole ed è quindi chiara per noi biologi la ricchezza culturale che si cela dietro il mantenimento di questo tipo di modello.

Il software libero non è quindi privo di regole ed è scorretto ritenere che non tuteli i diritti d'autore, anzi tende a tutelare il lavoro dell'autore vero e proprio piuttosto che dell'azienda per cui lavora. Da: www.debian.org/intro/free (modificato).

Box 2: Cos'è GNU/Linux?

Reduce da una "brutta fama", come sistema ostico, esclusivamente da linea di comando (senza finestre, mouse, ecc.), negli ultimi anni si è trasformato in un sistema operativo completo e facile da usare (Figure 7, 8 e 9). I vantaggi principali del suo impiego sono:

- abbattimento dei costi per l'acquisto e la gestione delle licenze;
- possibilità di provare migliaia di applicazioni;
- completa legalità nell'utilizzo;
- maggiore sicurezza, assenza di virus, trojan, backdoors, ecc.;
- ampio supporto da parte della comunità degli utenti e degli sviluppatori, tramite mailing lists;
- migliore integrazione fra programmi diversi;
- maggiore stabilità del sistema.

Gli svantaggi principali sono:

- la necessità di installare il sistema ex novo (raramente si possono reperire macchine con Linux preinstallato); questo può causare qualche difficoltà per gli utenti alle prime armi

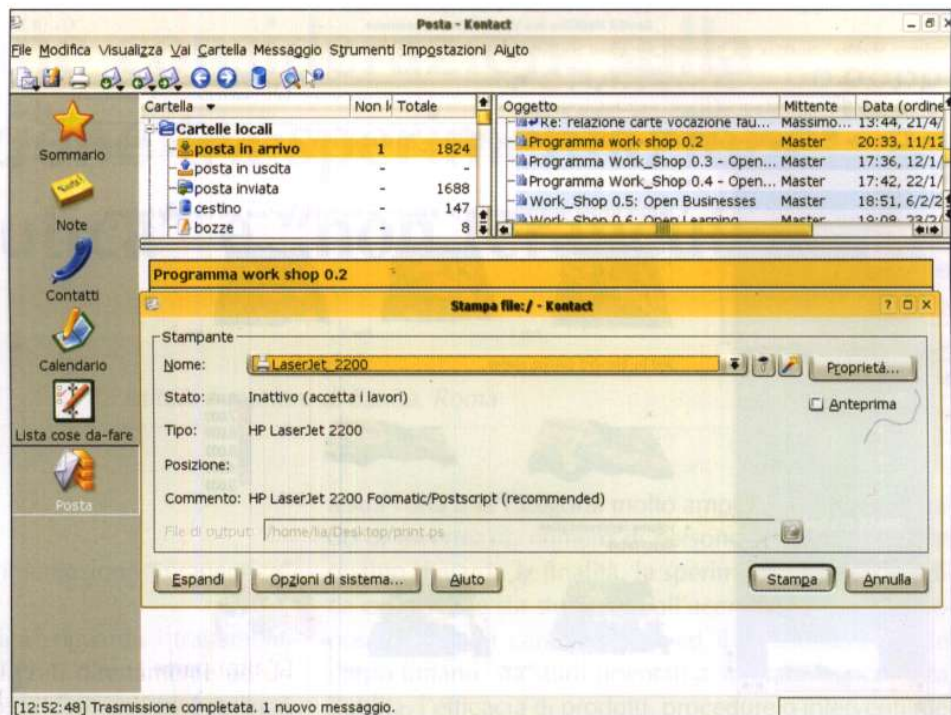


Fig. 7 - Uno dei numerosi programmi di posta elettronica (Kmail) libero e la finestra di stampa (Kprinter).

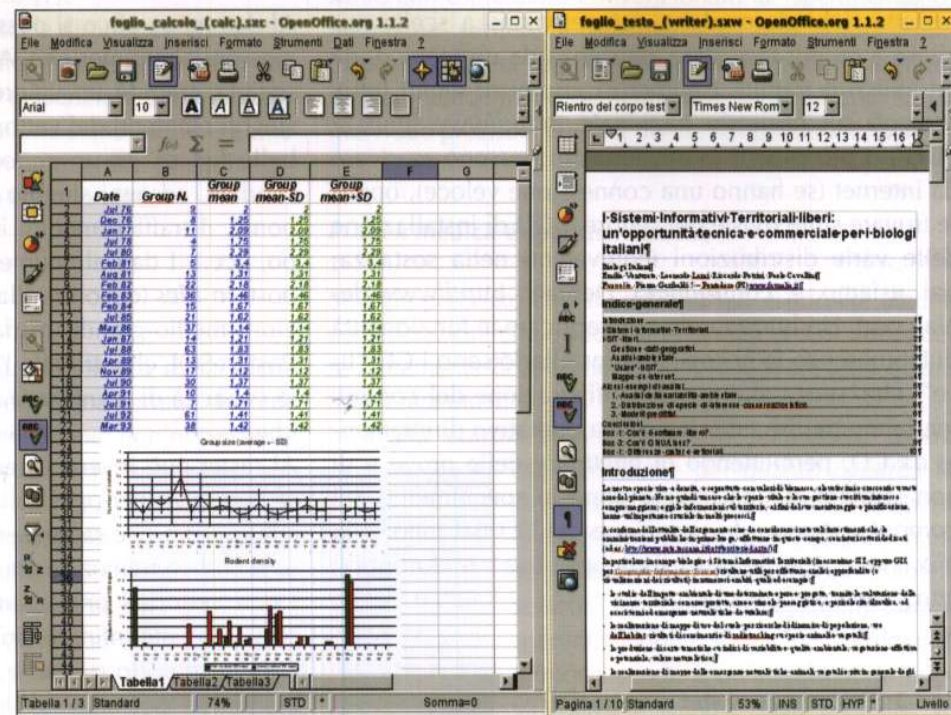


Fig. 8 - Foglio di calcolo e foglio di testo di OpenOffice.

- alcune applicazioni molto specifiche possono mancare o essere difficili da installare.

Installare Linux sul proprio computer, anche mantenendo Windows (con la possibilità di scelta fra i due sistemi all'avvio) è generalmente facile, ma una configurazione

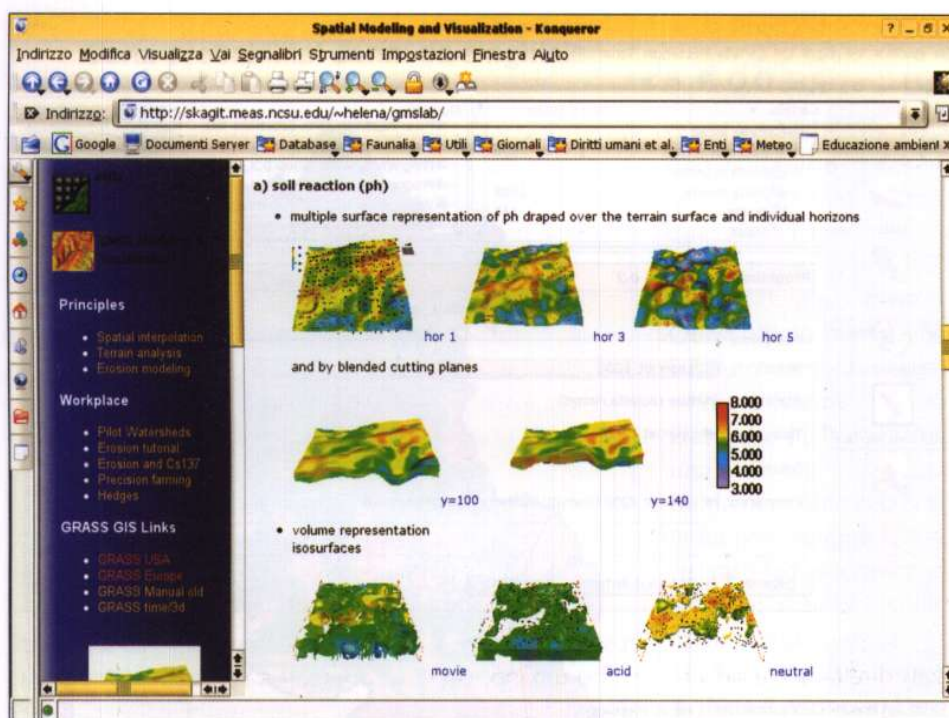


Fig. 9 - Konqueror. Uno dei numerosi browser disponibili fra i software liberi.

perfetta richiede qualche competenza. La scelta migliore è quindi farsi dare assistenza da un tecnico, da un amico o tramite i numerosi gruppi di utenti linux presenti sul territorio nazionale (<http://www.linux.it/LUG/>). I più esperti (o avventurosi) possono scaricare da internet (se hanno una connessione veloce), oppure trovare sulle riviste del settore, i CD di installazione delle varie distribuzioni (equivalenti nella sostanza; noi usiamo e consigliamo Debian: <http://www.debian.org>). Un buon modo per verificare le funzionalità di Linux senza la necessità di installarlo sono i CD "live", che, senza alterare la configurazione del computer, fanno partire tutto il sistema operativo direttamente da CD, permettendo di effettuare delle prove e di non lasciare alcuna traccia dopo lo spegnimento del computer stesso e la rimozione del CD dal lettore; la più nota tra le distribuzioni *live* è Knoppix: <http://www.knoppix.org>).

Box 3: Differenze raster e vettoriali

La struttura di una mappa raster è semplice, particolarmente adeguata per rappresentare dati continui (per es. l'altitudine) e per fare algebra delle mappe. Un raster può essere visto come una matrice, costituita da celle, in numero proporzionale alla risoluzione del raster stesso, ognuna delle quali assume un valore numerico.

I dati vettoriali sono utilizzati per la rappresentazione di punti, linee o poligoni e sono strutturati come un insieme di coppie di coordinate relazionate (linee e poligoni) o meno (punti) tra loro.

I vettori possono essere topologici, oppure no (*simple feature*). Il dato topologico ha una qualità intrinseca più elevata, a spese di una struttura più complessa e di un maggiore spazio occupato. Ad es., se una strada fa da confine fra bosco e un campo, un vettoriale non topologico dovrà avere tre linee (il margine del campo, quello del bosco e la strada), mentre il topologico avrà una sola linea che appartiene ai tre elementi.

Al dato geografico si possono, inoltre, associare informazioni alfanumeriche. Ad ogni vettore può essere associata, tramite un valore numerico caratterizzante i singoli elementi del vettore, un record di una o più tabelle (ad es. in un vettore che rappresenta una rete stradale, ad ogni singolo elemento corrisponde il suo nome, il traffico medio, il senso di circolazione, il tipo, ecc.). I dati alfanumerici associati possono essere posti in *files* (di solito dBase, dbf) oppure, per applicazioni medio-grandi, in databases; tra i più usati vi è PostgreSQL che, tramite l'estensione PostGIS/GEOS ha la capacità di mantenere anche informazioni geografiche.

Ai raster può invece essere associato un singolo valore numerico per ogni cella (ad es. ad una cella di uno strato raster che rappresenta l'uso del suolo è associato il codice numerico che lo descrive).

I dati vettoriali presentano una risoluzione maggiore (praticamente infinita), ad essi è possibile associare un numero elevato di dati e, di solito, occupano decisamente meno spazio sull'*hard disk*.

I dati raster hanno dimensioni direttamente proporzionali alla risoluzioni, contengono singole informazioni ma sono ideali per le analisi territoriali grazie alla estrema plasticità nelle funzioni algebriche.

Tutti i veri *software* GIS moderni includono in qualche misura entrambe le funzionalità. In particolare, negli studi ambientali complessi di solito vengono usati entrambi modelli di dati (raster e vettoriale).