

Luca Gremese

L'ARCHIVIO INFORMATICO SYBArc

Genesi e Organizzazione



Roma 2005

1. Introduzione

L'uso degli strumenti digitali ha accompagnato fin dal suo principio il progetto SYBAP (Sybrita Archaeological Project – www.sybrita.icevo.cnr.it), ma il passaggio ad una gestione integralmente informatizzata dei dati archeologici si è verificato in contemporanea con il processo di vettorializzazione del materiale cartografico/topografico prodotto sul campo.

La necessità di organizzare e correlare globalmente la notevole quantità di dati, di natura anche profondamente differente, che una campagna di scavo genera, ha condotto al passo successivo: la creazione di un archivio informatizzato. Se è stato immediatamente chiaro che lo strumento adeguato era di tipo DataBase (DB), le peculiarità dello scavo hanno evidenziato diverse problematiche, in risposta alle quali sono stati adottati alcuni criteri di base che hanno informato l'intera progettazione dell'archivio. Innanzitutto la leggerezza: requisito fondamentale per una compilazione che si svolge prevalentemente sul campo, in tempo reale ed in condizioni non sempre agevoli dal punto di vista tecnologico. In questo senso è stato ovvio optare per un DB che potesse essere ospitato su una singola stazione di lavoro portatile con caratteristiche standard, in grado quindi di operare anche non collegata ad un server centrale, benché quest'ultima opzione rimanga sempre aperta e preferibile. In secondo luogo si è cercato di progettare l'archivio garantendo a priori la possibilità di modificare in modo semplice e poco dispendioso l'intera struttura, prevedendo che con il tempo, dall'impiego e dall'analisi di dati ancora non registrati, scaturiranno nuove necessità. Dato il carattere spiccatamente multidisciplinare della ricerca archeologica, risultava importante prevedere la possibilità di multiutenza e di accesso sincrono ai record, nonché la portabilità [1] verso altre piattaforme (PC). Infine, visto il profilo strumentale del DB – che viene impiegato sia per l'elaborazione scientifica dei dati di scavo, sia per l'organizzazione degli stessi ai fini della pubblicazione digitale e cartacea – si è scelta un'architettura ibrida (DB + archivio immagini ad alta risoluzione) che verrà estesamente descritta nel paragrafo relativo (2).

Il carattere internazionale del progetto, che può contare sulla collaborazione di specialisti di varia provenienza, ha portato alla scelta della lingua inglese come standard per la creazione e la compilazione dell'archivio.

La realizzazione dell'archivio risponde inoltre alla necessità di rapido reimpiego della struttura finale nell'ambito di progetti simili, ma non necessariamente legati allo stesso sito.

2. Scelta dello strumento: database relazionale

La scelta dell'applicazione di base per la compilazione del database, informata dalle condizioni precedentemente descritte, è ricaduta sulla versione 6 di FileMaker, programma con solide basi di sviluppo sia in ambiente Mac che Wintel.

La previsione di un database di tipo relazionale [2] è stata determinata dalla necessità di porre in stretta connessione tutte le classi di dati prese in esame, così da ottenere

il quadro piu' completo possibile a partire dai risultati raggiunti sul campo. FileMaker, nella versione scelta, risponde ovviamente anche a questa ulteriore esigenza.

Oltre all'applicazione propriamente impiegata per la costruzione e compilazione dell'archivio, si e' reso necessario l'utilizzo di una serie di strumenti di supporto per la manipolazione delle informazioni nella fase precedente alla loro integrazione in SYBArc. Tra queste un ruolo preminente hanno certamente le applicazioni per la gestione delle immagini digitali sia raster che vettoriali (Adobe Photoshop ed Illustrator, VectorWorks). L'impiego prevalente della piattaforma Mac, con il suo sistema OS X, ha reso infine possibile la generazione di uno script di batch in shell bash [3] per una rapida ridefinizione ed ordinamento dei nomi dei files-immagine.

3. Organizzazione concettuale dei dati

In un primo tempo la suddivisione tra i diversi gruppi di documenti ha trovato fondamento nelle (evidenti) differenze sostanziali tra dati: i reperti di scavo, ovvero i materiali, i documenti cartacei e quelli digitali che peraltro rappresentavano i tre diversi aspetti operativi della ricerca archeologica. Questo approccio e' tuttavia risultato immediatamente poco funzionale ai fini dell'ordinamento dei dati e delle sintesi finali.

Si e' passati pertanto ad una suddivisione tematica delle informazioni. L'archivio comprende ora due gruppi principali di "raccolte-dati": documenti visivi (grafici o fotografici) e testi. Al primo gruppo appartengono i record che raccolgono foto e disegni dei reperti, del sito, dell'ambiente naturale, e la documentazione topografica. Il secondo gruppo è invece costituito dall'insieme delle schede di unità stratigrafica e quelle dei reperti archeologici e bioarcheologici.

Un'ulteriore problematica è posta dalla schedatura del cartaceo (che comprende foto e diapositive, carte topografiche, piante di scavo etc.). Il materiale cartaceo via via prodotto dev'essere infatti archiviato seguendo criteri affini a quelli adottati per il settore digitale.

L'archivio SYBArc costituisce così anche un registro preciso dei documenti cartacei, ciascuno dei quali avrà in una scansione ad alta risoluzione ed una scheda di riferimento i propri "alias" [4] informatici, con relativo codice.

Poiché il vasto patrimonio d'immagini digitali relative allo scavo non può essere impiegato all'interno del DB nel formato nativo (TIFF ad alta risoluzione, RAW o vettoriale), si è effettuata una copia di ciascun documento, convertendolo in bassa risoluzione e formato compresso JPG. Il file immagine così ottenuto può essere agevolmente inserito nel DB e non risulta appesantire il pacchetto finale, anche nel caso di uso in rete di SYBArc. In un'ottica d'impiego non limitata al solo ordinamento dei dati, parallelamente al DB vero e proprio si è inteso creare un "serbatoio" di tutte le immagini disponibili alla massima risoluzione, correlando ciascuna di esse ad una scheda tramite il codice univoco (che sarà pertanto identico per l'immagine ad alta risoluzione e per la corrispondente scheda) nell'ambito del progetto SYBArc. Tale serbatoio d'immagini ad alta risoluzione può essere facilmente trasferito, aggiornato e protetto (backup) su server e/o su media

trasportabili (dvd, cd-rom). Qualora si renda necessario le immagini ad alta risoluzione stivate nel cosiddetto “serbatoio” possono essere rintracciate rapidamente a partire dal loro codice ed utilizzate, ad esempio, per opere di stampa di qualità’.

L’integrazione nel database del materiale testuale ha comportato problemi di minore entità, dato che larga parte del lavoro pregresso era già disponibile in formato digitale (talvolta addirittura già in forma di database). Il processo di informatizzazione delle schede di scavo e testi di altro tipo, è stato quindi un passo ovvio, anche se non del tutto indolore.

4. SYBArc: definizione dei codici

La fase decisiva per l’ordinamento dei documenti è risultata quindi la definizione di gruppi omogenei rispetto ad un criterio di comodo. A ciascun gruppo è stato attribuito un codice alfabetico, in qualche misura evocativo, in riferimento alla lingua di compilazione del database (inglese).

Si hanno pertanto sigle relative alle immagini: ad esempio, DP ad indicare le fotografie digitali (Digital Picture), PH per le fotografie analogiche (PHotograph), SL per le diapositive (SLides) (vd. tabella in fig. 1).

CLASSIFICAZIONE DATI	
CODICE	DESCRIZIONE
MAP123/thk	cartografia e foto aeree
DPS123thk	foto digitali del sito
PHS123thk	foto analogiche del sito (scansioni)
SLS123thk	diapositive del sito (scansioni)
VM123thk	disegni vettoriali US/ES (planimetrie/elevati/sezioni di fase)
PM123thk	scansioni disegni US/ES (planimetrie/elevati/sezioni di fase)
DP123thk	foto digitali dei reperti
PH123thk	foto analogiche dei reperti (scansioni)
SL123thk	diapositive dei reperti (scansioni)
DR123thk	disegni dei reperti (scansioni)
US123thk	schede US
ES123thk	schede ES
PR123thk	schede equivalenza US/ES = Prokopiou-Rocchetti
RA123thk	schede RA
TMA123thk	schede TMA
PR_TMA_thk	schede equivalenza TMA = Prokopiou-Rocchetti
FLC123thk	schede flottazione - resti organici
ABC123thk	schede contesto reperti ossei
AB123thk	schede reperti ossei
EIMG123thk	immagini dell’ambiente

fig. 1 - Legenda codici SYBArc

L'inserimento di dati ed immagini non segue un ordine cronologico lineare. Il patrimonio d'informazioni da ordinare è per larga parte costituito da documenti generati nel corso di campagne di scavo e ricerca precedenti alla formazione dell'archivio digitale, che vengono digitalizzati ed accorpati nel momento in cui il lavoro di ordinamento lo permette o richiede. Diviene quindi fondamentale non attribuire valore temporale al codice di ciascun record, che segue invece una progressione numerica assoluta, nell'ambito del gruppo di appartenenza, come illustrato in precedenza e come risulta chiaro dalla tabella in fig.1.



The screenshot displays the SYBArc interface for finding images. At the top left is the SYBArc logo, and at the top right is the text "Findings Images". The central area features a photograph of a reddish-brown pottery fragment, identified by the code "thk00/13" in the top left corner of the image area. Below the image, a metadata table provides details for the record.

SYBArc_CODE	DP86thk
Origin	PIT54
Inventory Code	thk00/13
Description	fondo di pentola
Author	Ierman M.
Record Date	22-06-2004

At the bottom of the interface, there is a navigation bar with buttons labeled RA, US, AB, Drawing, MAP, and MAIN.

fig.2 - esempio di record

Tramite l'impiego di codici assoluti è possibile integrare nell'ambito di SYBArc qualunque tipo di documento in qualsiasi momento. Il riferimento temporale relativo al documento sarà conservato all'interno della scheda digitale, in appositi campi, insieme alla data di digitalizzazione ed archiviazione.

Il codice del singolo record è una sigla alfanumerica che tiene conto della categoria a cui esso appartiene e della posizione di inserimento. Esso riporta inoltre la sigla del sito intorno al quale il DataBase è stato costruito (nel nostro caso thk – corrispondente a Thronos Kephala). Un esempio può chiarire meglio i criteri di generazione del codice di un record:

DP345thk risulta essere il codice della 345^a immagine digitale (**D**igital **P**icture) inserita nel database relativo al progetto di scavo sulla sommità della Kephala (**T**Hronos Kephala).

Ovviamente la scheda relativa a questo documento riporterà al suo interno dati sostanziali, come il soggetto dell'immagine e la data di creazione dell'originale, ma il codice non risulta descrittivo in questo senso (vd. fig. 2)

5. SYBArc: definizione delle relazioni

La nervatura dell'intero lavoro, ciò che ne permette un proficuo impiego ai fini della ricerca, è costituita dalle relazioni tra i record. Considerando le procedure fondamentali della disciplina archeologica ed osservando lo schema in fig. 3, risulta chiaro come sia possibile, a partire da uno qualsiasi dei record, consultare altri records, appartenenti a categorie differenti, correlati al documento di partenza. Per rendere più chiara la sostanza delle relazioni, prendiamo l'esempio di un'immagine digitale di un reperto (DP---thk).

DP---thk presenta collegamenti immediati al record del disegno del reperto, DR---thk, al record della scheda RA (scheda Reperti Archeologici), al record US---thk (scheda Unità Stratigrafica) e tramite quest'ultimo ai record relativi alle immagini ed ai disegni del sito (VM---thk – Vector Maps; PM---thk – Paper Maps e via dicendo). Viceversa, a partire da una scheda US---thk, sarà possibile consultare una serie di record della categoria DP (Digital Picture), ovvero tutti quelli che hanno al proprio interno un riferimento all'unità stratigrafica cui il record di partenza US---thk si riferisce. In tal modo si genera un flusso continuo di informazione, base per qualunque sintesi, che deriva dalla organizzazione stessa, primaria, dei dati. Le relazioni sono state formalizzate e visualizzate sotto forma di pulsanti siglati sulla scheda di ogni singolo record, per permetterne un rapido impiego (vd. fig.2 parte inferiore della scheda).

La realizzazione di un portale d'accesso ai DB (e link al sito/DB on-line), infine, permette la generazione di relazioni occasionali, passaggi non previsti dal progetto, ma che possono sorgere in corso d'uso, offrendo quindi all'utente la possibilità di associare i record in maniera non limitata alle sole relazioni prestabilite. Il portale, come si può evincere dalla fig.4, offre la possibilità diretta di accesso, oltre che alla legenda dei codici (KEY), ai relativi gruppi di records, suddivisi nei due grandi comparti dell'archivio come descritto nel paragrafo 3.

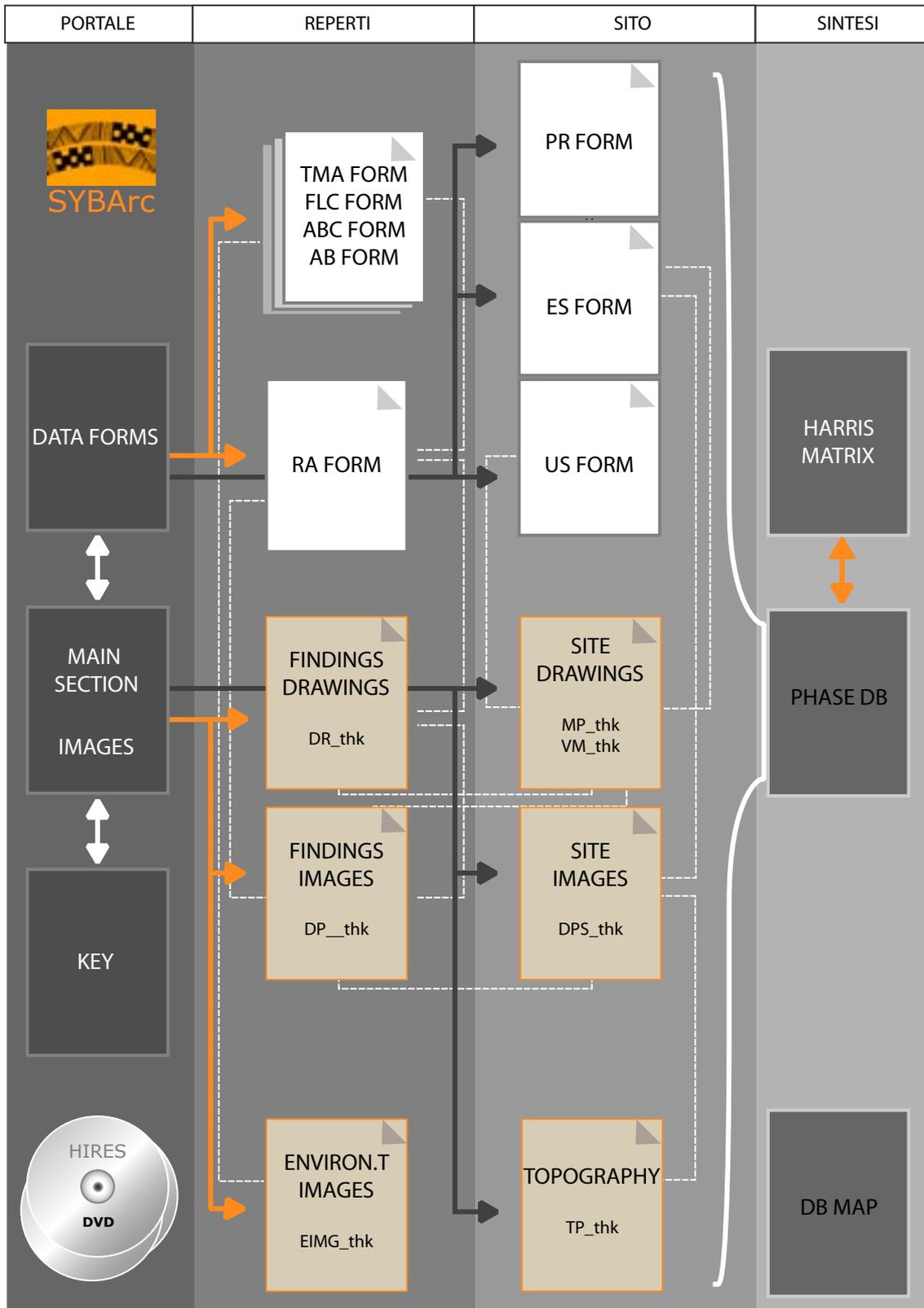


fig.3 - La mappa dell'archivio SYBArc con indicazione delle relazioni



fig.4 - il portale del database

6. SYBArc: il progetto grafico

L'aspetto grafico del database SYBArc risulta chiaramente caratterizzato cromaticamente. Si è inteso in tal modo fornire agli utenti un sistema di riconoscimento visivo rapido dei vari comparti o gruppi omogenei di dati. La scelta di formati e caratteri richiama inoltre il sito web relativo alla ricerca, in direzione di un'uniformità generale per gli esiti pubblici della ricerca.

Le singole schede dei records presentano quattro differenti layout [5], destinati ad altrettanti impieghi previsti. Di seguito una breve descrizione di ciascuna delle modalità nelle quali sono presentate le informazioni dell'archivio:

COMPLETE: modalità di consultazione integrale della scheda (record) di un documento. riporta tutti i campi descrittivi ed identificativi relativi al documento, l'immagine, dove richiesto, ed i collegamenti relazionali di cui al paragrafo 5, in forma di pulsanti.

LIST: modalità di consultazione seriale dei record appartenenti ad uno dei database dell'archivio. Destinata a fasi di lavoro di confronto o consultazione rapida. L'elenco dei record consta solamente dei campi principali, come ad esempio il codice SYBArc o la data di registrazione.

ENLARGE: modalità di visualizzazione relativa ai record che presentano un'immagine. In tale maniera si intende favorire eventuali confronti o analisi visive dei documenti, si tratti di fotografie digitale o elaborati grafici. La scheda presentata in tale modalità non reca alcun tipo di informazione aggiuntiva, oltre all'immagine ed al codice identificativo SYBArc.

PRINT: modalità di visualizzazione ottimizzata per la stampa di reports e singoli records.

APPENDICI

A. Shell Script

Considerata soprattutto l'enorme mole di disegni e fotografie dei reperti, i cui relativi files riportano originariamente un nome generico (attribuito in modo a volte non facilmente controllabile dall'applicazione di scansione, ad esempio, o dall'apparecchio fotografico), e' stato affrontato e risolto il problema di convertire ciascun nome-file alle esigenze dell'archivio, serialmente, con quantita' di files anche molto alte: un semplice file batch (scritto appositamente per lo scopo) viene richiamato in shell bash (disponibile su OS X) con il comando "rinomina". Il programma batch esegue la conversione su tutti i files indicati dall'operatore, che puo' interagire passo passo indicando la categoria, il numero progressivo da cui iniziare il conteggio del codice e, ove occorresse, l'estensione dei file (ad es. jpg o tif). Il processo di batch si serve di due directory appositamente create, contenenti i file di partenza e quelli di destinazione, ovvero i file da rinominare e quelli rinominati. E' chiaro come l'esecuzione di tale processo manualmente, file per file, risulterebbe oltremodo dispendiosa in termini di tempo, nonche' rischiosa per l'importantissima univocità dei codici.

B. Backup e protezione dell'archivio

Semplici procedure di backup periodico (copia su supporto ottico ad alta capacità – DVD) si affiancano all'impiego di hardware specificamente pensato per offrire la massima sicurezza possibile. Una coppia di dischi identici è infatti installata sulla macchina che ospita SYBArc. Sui due dischi le informazioni vengono stivate in parallelo, in tempo reale all'atto della compilazione del database (mirror raid), garantendo così una probabilità di perdita delle informazioni prossima allo zero.

C. Accesso, aggiornamento e revisione del DB in simultanea (rete)

E' possibile che più utenti necessitino di un accesso al database, contemporaneamente, ponendo così una serie di problemi diversi dall'ordinario. Primo fra tutti quello relativo alle postazioni di lavoro. La macchina che usualmente ospita il database diviene in questi casi l'equivalente di un server di rete. L'accesso simultaneo, sincrono al database da differenti postazioni di lavoro personali richiede quindi la presenza di una rete (network) LAN [6] o WAN [7] e di codici identificativi personali per l'accesso ai documenti o addirittura alla macchina/server stessa ai fini della sicurezza dei dati. In alternativa è possibile sfruttare una caratteristica del sistema, la multiutenza, che tuttavia consente solo un accesso asincrono al database ed in locale, essendo in tal caso solamente una la macchina impiegata. Il MacOS X offre infatti la possibilità di generare un profilo personale per ciascun utente, che avra' così la possibilità di apportare gli aggiornamenti di sua competenza all'archivio SYBArc.

Il database è strutturato anche per sessioni di lavoro sul campo. In tali frangenti, data la probabile distanza geografica, non sempre è facile accedere alla macchina che ospita il database (l'operazione potrebbe essere possibile, ad esempio, tramite una linea dati). In tal caso è indicato effettuare preventivamente una copia integrale del database su un

elaboratore portatile e bloccare l'archivio d'origine fino al momento in cui l'accesso al server centrale da parte degli amministratori non sarà nuovamente garantito.

Considerate le previsioni di intenso impiego in rete, anche in associazione con il sito web della ricerca, e i problemi di esportazione dei dati che le nuove versioni del programma FileMaker presentano, si valuta la possibilità di una transizione dell'intero SYBArc verso una differente base, che potrebbe essere individuata nell'abbinamento della piattaforma SQL con il linguaggio PHP.

GLOSSARIO

[1] **portabilità**: possibilità di esportare o convertire in altro formato/piattaforma i documenti creati con un'applicazione specifica (es: filemaker --> html).

[2] **relazionale**: e' cosi' definito un insieme di database in relazione tra loro secondo regole prestabilite.

[3] **script, batch, shell, bash.**

Script: file eseguibile che raggruppa una serie di istruzioni in linguaggio interprete (ad es. il linguaggio della shell di un sistema operativo)

Batch: file che automatizza una serie di operazioni ripetitive.

Shell: ambiente di lavoro a linea di comando, tipicamente di sistemi unix ed affini - tra cui anche il MacOS X (alt: linguaggio interprete dei comandi).

Bash: Bourne-Again SHell - una delle shell piu' comunemente utilizzate in ambito unix ed affini.

[4] **alias**: corrispettivo virtuale di un documento. Nel caso di un ambiente ibrido digitale analogico

Nell'ambito del sistema operativo MacOS, invece, si tratta di un collegamento ad un documento.

[5] **layout**: aspetto grafico, impostazione visiva.

[6] **LAN**: acronimo di Local Area Network. Rete tra stazioni di lavoro locali, in cui ogni nodo puo' svolgere sia la funzione di server che di client.

[7] **WAN**: acronimo di Wide Area Network. Rete tra stazioni di lavoro remote (da cui il termine wide), in cui ogni nodo puo' svolgere sia la funzione di server che di client.