



L'ARCHIVIO STATLIB: UN NUOVO PUNTO DI ACCESSO PER L'EUROPA

Gianfranco Galmacci

Dipartimento di Scienze Statistiche, Università degli Studi di Perugia.

1. L'ARCHIVIO STATLIB

Con l'avvento della rete Internet, basata sul protocollo TCP/IP e costituita originariamente da elaboratori con sistema operativo UNIX, è invalsa la consuetudine di creare archivi contenenti programmi, dati, informazioni ecc. che, ritenuti di pubblica utilità, vengono messi a disposizione degli interessati e tramite la rete possono essere acquisiti liberamente. La tecnica con cui ciò è reso possibile consiste nel creare ambienti di lavoro, privi di parole chiave per l'accesso e solitamente identificati come *anonymous*, ai quali è possibile connettersi tramite rete utilizzando la funzione *ftp* di TCP/IP. In tal modo chiunque è in grado di entrare in tali ambienti, esaminarne i contenuti e trasferire sulla propria macchina tutto ciò che può essere interessante.

La diffusione del protocollo TCP/IP a tutti gli ambienti, dai *personal computer* ai *main frame* qualunque sia il sistema operativo da essi utilizzato*, ha allargato in maniera rapidissima l'interesse per questo tipo di servizio; oggi questo è il modo più semplice ed efficiente di mettere a disposizione della comunità scientifica tutti i

* Attualmente si calcola che i computer connessi in Internet siano circa 1.300.000 con oltre 10 milioni di utenti.

prodotti software che non sono commercializzati e che vengono distribuiti gratuitamente dagli autori (*public domain software*).

La maggior parte degli archivi di questo tipo è disponibile negli Stati Uniti, tuttavia quelli più importanti, e per i quali si rileva anche il maggior traffico, vengono duplicati in varie parti del mondo per evitare intasamenti della rete con conseguenti disservizi.

L'archivio di maggiore interesse per la statistica è STATLIB, creato e mensilmente aggiornato da Michael M. Meyer della Carnegie Mellon University. Esso è molto esteso e contiene una grande quantità di programmi di vario tipo (tra cui tutte le procedure apparse su *Applied Statistics*), procedure per ambienti di calcolo orientati alla statistica, serie di dati di interesse rilevante da un punto di vista didattico o scientifico e svariate informazioni (tra cui alcuni archivi contenenti gli indirizzi di ricercatori nel campo della statistica di vari paesi).

Un'idea di come l'archivio è strutturato e dei suoi contenuti la si può ottenere dalla tabella inclusa nel file GET.ME.FIRST, presente nella *directory* principale di STATLIB e di seguito riportata.

apstat	– Selected algorithms transcribed from <i>Applied Statistics</i>
asascs	– Material related to the Statistical Computing Section of the American Statistical Association.
blss	– Macros and fixes for the BLSS statistical package.
crab	– Kodiak Island king crab survey data.
cmlib	– Core Mathematics Library from NIST. Fortran source for lots of statistics and numerical procedures.
datasets	– Various datasets. (see also <i>jasadata</i>)
designs	– Designs and algorithms and programs for creating designs.
directory	– Lists of addresses and e-mail addresses of statisticians.
disease	– National Notifiable Diseases Data
general	– software of general statistical interest.
glim	– Macros and software for <i>glim</i> .
griffiths-hill	– <i>Applied Statistics</i> algorithms contained in the book by Griffiths and Hill.
jasadata	– Data published in <i>JASA</i> .
jcgs	– Algorithms and abstracts from the <i>Journal of Computational and Graphical Statistics</i> .
jqt	– Algorithms published in the <i>Journal of Quality Technology</i> .
meetings	– Programs for forthcoming meetings.
mis	– Minitab Industrial Statistics macros
mug	– Minitab macros from the Minitab Users Group
multi	– <i>Multivariate Analysis and Clustering</i> . An annotated directory and selected algorithms.
S	– S functions, device drivers and related software.
s-news	– Archives of the S-news mail, in digest format.
xlispstat	– Luke Tierney's <i>XlispStat</i> system, updated March 1991.

Recentemente questo archivio è stato completamente duplicato presso l'Università degli Studi di Perugia con l'intento di creare un servizio più efficiente per gli utenti europei, ed italiani in particolare. Nella versione disponibile presso l'Università di Perugia l'archivio Statlib è perfettamente identico a quello originale tranne che per il sistema XlispStat (prodotto da Luke Tierney della University of Minnesota) che è stato completato delle versioni per Macintosh, DOS e Amiga.

Statlib contiene attualmente un totale di circa 4000 file (che occupano oltre 80Mb di spazio disco) e si articola in 22 directory principali. Descrivere nel dettaglio i contenuti dei vari ambienti sarebbe molto lungo e comunque inutile, la via più diretta è quella di accedere all'archivio e seguire la documentazione di cui esso è provvisto (in ogni directory è presente un file *index* e/o *README* che ne illustra il contenuto); tuttavia si ritiene utile fornire una brevissima descrizione degli ambienti più rilevanti.

Il più ricco ed esteso degli ambienti è quello contenuto nella directory *cmllib*, una biblioteca di *routine* Fortran nota come National Bureau of Standards (NBS) Core Math Library, contenente oltre 50 package di programmi relativi a metodi matematici e statistici. Tra i prodotti più noti di questo archivio vanno ricordati *Linpack* ed *Eispack* realizzati presso l'Argonne National Laboratory (si vedano ad esempio Dongarra, 1978, e Smith, 1976).

Gli ambienti *S*, *glim*, *mis*, *mug* (gli ultimi due si riferiscono a Minitab) contengono procedure relative a metodologie statistiche non previste nei package originali e, nel caso di *S*, una serie di programmi di supporto per rendere più flessibile l'ambiente di programmazione. Nel caso di *glim*, sono state rese disponibili tutte le procedure pubblicate in GLIM Newsletter.

Le directory *datasets*, *disease*, *jasadata*, *crab* contengono invece i dati utilizzati come esempi in 26 diversi volumi, oltre a quelli inseriti in articoli, studi monografici ecc.

La directory *meetings*, infine, contiene informazioni e programmi relativi a Convegni e manifestazioni scientifiche internazionali concernenti discipline statistiche.

2. SERVIZIO ANONYMOUS FTP PRESSO L'UNIVERSITÀ DI PERUGIA

Presso l'Università degli Studi di Perugia è stato attivato un servizio per la distribuzione di *public domain software* tramite una macchina *anonymous**; essa è operativa attualmente su una workstation il cui indirizzo INTERNET è *ftp.unipg.it*** e consente un accesso libero via *ftp*, una delle funzioni del protocollo TCP/IP.

Molti sono i prodotti che possono essere acquisiti tramite questo servizio, tra questi può essere interessante menzionare in questa sede il sistema per la composizione di testi T_EX, l'editore Gnu-emacs e, naturalmente, tutto quanto è contenuto in Statlib. Per accedere a Statlib è necessario entrare nella *directory pub/stat* e quindi nella *sub-directory statlib*; questo ambiente contiene a sua volta un insieme di *directory* molto articolato, che comunque è descritto nei file *index* che sono disponibili ai diversi livelli.

Poiché la macchina *ftp.unipg.it* è collegata alla rete INTERNET, coloro che hanno una connessione con questa rete possono accedere direttamente al distributore. Un servizio analogo viene offerto anche a coloro che fossero connessi unicamente alla rete EARN/BITNET; in tal caso si dovrà utilizzare un *server* realizzato dalla Princeton University che provvede ad inoltrare sulla rete INTERNET, con modalità *batch*, una sequenza di comandi *ftp* alla macchina *anonymous*. Nei paragrafi successivi verranno illustrate le modalità da seguire nei due casi.

Una ampia documentazione generale sull'utilizzo della rete INTERNET e sui servizi che tramite essa vengono resi disponibili è contenuta nella guida di Kehoe (1992) ed in quella curata dal Network Service Center (1989), entrambe disponibili come file PostScript, in formato compresso, presso *anonymous@ftp.unipg.it* della *directory/pub/stat/network-doc*.

2.1. ACCESSO DA UNA STAZIONE CONNESSA IN INTERNET

In questo paragrafo viene esemplificata una sessione-tipo per illustrare l'accesso alla macchina *anonymous@ftp.unipg.it* da un elaboratore direttamente collegato con la rete INTERNET. Si noti che i comandi da inviare sono stati indicati in neretto mentre le risposte della macchina sono riportate in corsivo.

* Si ringrazia il dott. Giuseppe Vitillaro della IBM-SEMEA per la collaborazione prestata avendo egli provveduto ad installare il software necessario per la definizione dell'ambiente *anonymous*.

** Questo nome è in realtà un alias della macchina *teseo.unipg.it*, una workstation IBM RISC 6000 mod. 560 sulla quale è stata realizzata l'applicazione. Questa tecnica è stata utilizzata al fine di mantenere invariato l'indirizzo dell'*anonymous* anche nel caso in cui il sistema che lo ospita dovesse essere cambiato.

ftp ftp.unipg.it

Connected to teseo.unipg.it.

220 teseo FTP server (Version 6.6 Fri Mar 5 10:04:08 MET 1993) ready.

Name (ftp.unipg.it): **anonymous**

331 Guest login ok, send e-mail address as password.

Password: inserire qui l'indirizzo dell'utente che fa la richiesta, ad es.:

glm@fisher.stat.unipg.it

230 Guest login ok, access restrictions apply.

ftp> **dir**

200 PORT command successful.

150 Opening ASCII mode data connection for /bin/ls.

total 24

dr-xr-xr-x	2	root	system	512	Apr 06 10:52	bin
dr-xr-xr-x	3	root	system	512	Apr 06 10:52	etc
Drwxr-xr-x	27	bin	bin	1024	May 15 21:49	pub
drwxr-xr-x	2	root	system	512	May 15 22:03	tmp

226 Transfer complete.

ftp> **cd pub**

250 CWD command successful.

ftp> **cd stat**

250 CWD command successful.

ftp> **dir**

200 PORT command successful.

150 Opening ASCII mode data connection for /bin/ls.

total 2

Drwxr-xr-x	22	bin	bin	10240	Feb 08 08:50	statlib
Drwxr-xr-x	2	bin	bin	384	Sep 29 15:19	susam
Drwxr-xr-x	2	bin	bin	384	Jan 22 19:22	network-doc
Drwxr-xr-x	8	bin	bin	1024	Jan 22 19:51	xlispstat

226 Transfer complete.

ftp> **cd statlib**

250 CWD command successful.

ftp> **dir**

200 PORT command successful.

150 Opening ASCII mode data connection for /bin/ls.

total 11

Frw-r-r-	1	bin	bin	4291	May 31 08:48	GET.ME.FIRST
Drwxr-xr-x	2	bin	bin	2048	Jun 04 06:25	S

```

Drwxr-xr-x  2 bin  bin  4096 May 29 18:48 apstat
Drwxr-xr-x  2 bin  bin   384 May 29 18:56 asascs
Drwxr-xr-x  9 bin  bin  3072 Jun 04 07:41 cmlib
.....
.....

```

226 Transfer complete.

ftp> **get GET.ME.FIRST**

200 PORT command successful.

150 Opening ASCII mode data connection for GET.ME.FIRST (4291 bytes).

226 Transfer complete.

4387 bytes received in 0.3878 seconds (11.05 Kbytes/s)

ftp> **quit**

221 Goodbye.

Come si può dedurre dall'esempio, mediante la direttiva **get** è possibile trasferire sulla propria macchina tutti i file desiderati. Si noti che nella lista che si ottiene mediante il comando **dir** il primo carattere a sinistra specifica se il nome, che figura all'estrema destra della riga, si riferisce ad un file (carattere F), ad una directory (carattere D) oppure ad un logic link con un altro file (carattere L).

Si tenga inoltre presente che è sempre necessario scrivere i nomi dei file rispettando rigorosamente caratteri maiuscoli e minuscoli.

2.2. ACCESSO DA UNA STAZIONE CONNESSA IN RETE EARN/BITNET

Se si dispone di una macchina collegata alla rete EARN, e non anche ad INTERNET, l'accesso al distributore è sempre possibile ma tramite una via diversa. Presso il nodo PUCC.BITNET (Princeton University) esiste una macchina che funge da server tra le reti BITNET ed INTERNET e provvede ad inoltrare richieste di ftp a macchine anonymous consentendo di ottenere informazioni e software in modo analogo a quanto è possibile fare con un collegamento diretto.

A tal fine è necessario inviare un messaggio (E-mail) alla macchina *bitftp@pucc.bitnet* inserendo come testo una serie di comandi ftp, così come mostrato di seguito.

```

ftp ftp.unipg.it
user anonymous
cd pub/stat/statlib
dir
get GET.ME.FIRST
quit

```

I file richiesti verranno inviati automaticamente al richiedente nel giro di un periodo di tempo relativamente breve, dipendente comunque dal traffico di rete. In particolare nell'esempio sono stati inseriti i comandi da inviare per prelevare dal distributore il file GET.ME.FIRST.

Per un corretto uso dei comandi disponibili per *ftp* è comunque necessario consultare un manuale specifico (la guida di Kehoe, 1992, costituisce un eccellente supporto anche per utenti esperti).

BIBLIOGRAFIA

- Dongarra J.J., Bunch J.R., Moler C.B., Stewart G.W., 1978, *LINPACK Users Guide*, SIAM Publications, Philadelphia, PA.
- Kehoe B.P., 1992, *Zen and the Art of Internet, a Beginner's Guide to Internet*. Widener University. Network Service Center, 1989, *Internet Resource Guide*. BBN System Technologies Corporation, Cambridge, Ma.
- Smith B.T. et al., 1976, *Matrix Eigensystem Routines – EISPACK Guide*, 2nd, ed., Springer-Verlag, New York.

SUMMARY

Stablit is the most important network based retrieval system for statistical software. Originally created at the Carnegie Mellon University, it is now available at the University of Perugia, Italy, (anonymous@ftp.unipg.it) via ftp, one of the functions of the TCP/IP protocol of Internet.

A short overview of the archive has been given as well as an example of an ftp interactive session from Internet and a batch sample request to retrieve files from Bitnet.