BANCA D'ITALIA
EUROSISTEMA

Indagine sulle imprese industriali e dei servizi

BIRD
Bank of Italy Remote access to micro Data

Istruzioni per l'uso
Versione 3.4

giugno 2019

Sommario

Utilizzo del sistema di elaborazione a distanza "BIRD" 3
1. Introduzione 3
2. Fasi di utilizzo del sistema 3
3. Abilitazione degli utenti 4
4. Ambienti supportati 4
5. Accesso al sistema tramite e-mail 5
6. Revisione e risposte del sistema tramite e-mail 5
7. Limitazioni del sistema 6
8. Riservatezza 6
9. Suggerimenti generali di utilizzo 7
10. Segnalazione dei lavori di ricerca svolti 7
11. Assistenza e supporto 7
12. Appendice 1: indicazioni specifiche per il linguaggio Stata 8
12.1. Versione supportata e default di sistema 8
12.2. Indicatore di fine riga 8
12.3. Accesso ai dataset 8
12.4. Suggerimenti di utilizzo 8
13. Appendice 2 : Indicazioni specifiche per il linguaggio R 9
13.1. Versione supportata e default di sistema 9
13.2. Accesso ai dataset 9
13.3. Suggerimenti di utilizzo 9

Utilizzo del sistema di elaborazione a distanza "BIRD"

1. Introduzione

La Banca d'Italia, tramite il sistema BIRD, offre la possibilità di svolgere elaborazioni sui dati raccolti con le indagini sulle imprese industriali e dei servizi, nel rispetto della riservatezza dei dati individuali.

L'utente svolge le sue analisi statistiche senza avere accesso diretto ai singoli micro dati: invia una e-mail contenente un programma di elaborazione in uno dei linguaggi ammessi e il sistema fornisce, sempre via e-mail, l'output dell'elaborazione.

Per facilitare la definizione dei programmi di elaborazione è stato creato un dataset fittizio che replica la struttura dell'archivio Indagine sulle imprese industriali e dei servizi. Il dataset è diffuso in formato ASCII con campi separati da punto e virgola. I dati contenuti in questo dataset di prova sono generati casualmente e non hanno nessuna relazione con i dati effettivamente rilevati dall'indagine.

L'utilizzo del sistema BIRD è subordinato all'accettazione, da parte della Banca d'Italia, della richiesta di rilascio di un'utenza.

Nelle pubblicazioni che facciano riferimento alle elaborazioni ottenute tramite il sistema BIRD, l'autore è tenuto a citare la fonte (Banca d'Italia, Indagine sulle imprese industriali e dei servizi o sondaggio congiunturale sulle imprese industriali e dei servizi), ma non può coinvolgere in alcun modo questo Istituto, essendo l'utilizzo dei dati sotto la piena e sola responsabilità dell'autore medesimo.

2. Fasi di utilizzo del sistema

L'utilizzo del sistema prevede le seguenti fasi successive:

a. gli utenti si registrano, riempiendo un modulo in cui si qualificano e sottoscrivono le condizioni d'uso;

b. una volta autorizzati, per svolgere le elaborazioni, gli utilizzatori mandano un messaggio di posta elettronica al sistema (con un'intestazione che permette l'identificazione del mittente) e un programma da eseguire in uno dei linguaggi previsti;

c. se le credenziali vengono riconosciute, il sistema effettua una serie di controlli formali sui comandi utilizzati. In caso di esito positivo elabora il programma e, dopo una serie ulteriore di verifiche sul contenuto, i risultati vengono inoltrati all'indirizzo di posta elettronica abilitato.

Per una descrizione della metodologia e le informazioni necessarie allo sviluppo di programmi di elaborazione si rimanda alla specifica documentazione presente sul sito internet del sistema BIRD e, per quanto non in essa reperibile, ai Supplementi al Bollettino Statistico – Indagini campionarie, annualmente dedicati all'Indagine sulle imprese industriali e dei servizi e al sondaggio congiunturale sulle imprese industriali e dei servizi.

3. Abilitazione degli utenti

L'accesso al sistema viene offerto tramite una procedura di autenticazione basata sull'attribuzione all'utente di alcuni parametri individuali che devono essere sottoposti insieme al programma per ciascuna elaborazione.

Gli utenti interessati propongono una richiesta di abilitazione tramite il modulo disponibile sul sito internet di BIRD. Tale richiesta deve essere compilata, sottoscritta e corredata della copia di un documento d'identità del richiedente in corso di validità.

Questa richiesta deve essere inviata **via email** al seguente indirizzo: **bird_assist@bancaditalia.it**.

Le domande di abilitazione devono contenere allegati (in formato pdf) le scansioni elettroniche **leggibili** della documentazione cartacea richiesta.

L'abilitazione dell'utenza viene comunicata mediante un messaggio di posta elettronica, inviato alla casella e-mail indicata nel modulo di domanda, che riporta i parametri individuali seguenti:

- **user** (nome, scelto dall'utente). Si raccomanda di usare una sequenza continua di lettere o numeri non interrotta da caratteri non alfabetici quali ".", "/", etc.
- **password** (parola chiave, scelta dall'utente)
- **project** (progetto di abilitazione - attribuito dall'amministratore del sistema)

La modifica dell'indirizzo e-mail abilitato è possibile mediante apposita comunicazione all'indirizzo e-mail **bird_assist@bancaditalia.it** (vedi paragrafo "Assistenza e supporto"). Su richiesta è anche possibile abilitare all'invio ulteriori indirizzi e-mail, purché riferiti al medesimo richiedente.

L'utenza abilitata è assolutamente individuale e non cedibile nemmeno temporaneamente; essa viene abilitata alle elaborazioni per la durata di un anno, trascorso il quale viene automaticamente disabilitata; è possibile richiederne l'estensione di validità tramite comunicazione all'indirizzo e-mail **bird_assist@bancaditalia.it** (vedi paragrafo "Assistenza e supporto").

4. Ambienti supportati

I linguaggi di elaborazione ammessi sono Stata e R. La scelta del linguaggio di elaborazione è fatta tramite lo statement ***package =** delle intestazioni di identificazione. Ad esempio, per sottomettere un programma in Stata sarà necessario scrivere tra le intestazioni lo statement

***package = stata**

per sottomettere un programma in R sarà necessario scrivere invece tra le intestazioni lo statement

***package = R**

Ogni altra parola che compaia nello statement ***package =** genererà un messaggio di errore e il programma non verrà sottomesso al sistema.

Maggiori dettagli sulle piattaforme disponibili al sistema sono reperibili nelle Appendici di questo documento.

5. Accesso al sistema tramite e-mail

L'utente accede al sistema BIRD predisponendo un programma scritto in uno dei linguaggi supportati. Il programma deve essere contenuto all'interno del testo di un messaggio di posta elettronica. Tale messaggio deve rispettare i seguenti requisiti:

- deve essere in "**testo normale**" ("plain text"), **con esclusione di qualsiasi tipo di formattazione (sono quindi da evitare, tra l'altro, colori, grassetti, Vcard, richieste di conferma di lettura, allegati, disclaimer, firme automatiche)**; si raccomanda inoltre di evitare caratteri quali apostrofi o caratteri specifici di alcune lingue (ü ß æ Å ç ñ e simili);
- deve **contenere nelle prime quattro righe un'intestazione**, di seguito esemplificata per un utente il cui nome è "mario", la cui password è "rossi", il cui progetto di assegnazione è "invind" che intende sottomettere un programma in linguaggio Stata:

  ***user = mario**
  ***password = rossi**
  ***project = invind**
  ***package = stata**

Ciascuna delle quattro righe di autenticazione va preceduta, come si vede, da un asterisco. Tali righe devono essere consecutive e separate l'una dall'altra dal carattere "a capo".

Tutto ciò che è contenuto nel messaggio a partire dalla riga successiva alle intestazioni viene passato all'ambiente elaborativo e deve pertanto essere strettamente nel linguaggio prescelto.

Maiuscole, minuscole e spazi **non** sono rilevanti per le intestazioni, anche se potrebbero esserlo per il linguaggio utilizzato (si vedano al riguardo le Appendici).

Si raccomanda di verificare che non vi siano spazi in eccesso dopo il termine di ciascuna delle righe del programma da elaborare. In caso contrario potrebbero verificarsi errori nell'elaborazione.

Il messaggio di posta elettronica inviato deve avere come mittente l'indirizzo e-mail abilitato ed associato permanentemente alle credenziali di autenticazione fornite. **Il sistema respinge sottomissioni con credenziali esatte, ma provenienti da caselle e-mail non corrispondenti all'utenza abilitata.**

L'indirizzo e-mail al quale inviare il messaggio contenente le elaborazioni richieste è: **bird@bancaditalia.it**

Il server di posta non elabora messaggi provenienti da utenti non abilitati, o che non rispettino le regole sopra indicate, o il cui contenuto non consista di un programma scritto nei linguaggi supportati.

6. Revisione e risposte del sistema tramite e-mail

L'elaborazione del programma viene svolta dal sistema in via automatica nei tempi necessari in relazione alle risorse disponibili e alla domanda degli altri utenti. Al termine delle elaborazioni, l'output viene sottoposto ad un **processo di revisione**.

I gestori del sistema esaminano l'output per verificarne la conformità alle politiche di riservatezza e di uso del sistema; al termine del processo, l'output viene inoltrato all'utente all'indirizzo e-mail abilitato; se opportuno, l'output viene trattenuto e l'utente riceve invece un messaggio motivato di rifiuto.

Il tempo necessario all'elaborazione dipende dalla sua complessità e dal concorrente utilizzo delle risorse, incluse quelle relative alla gestione della posta in entrata e in uscita. Il tempo necessario al processo di revisione dipende dalla complessità dell'elaborazione e da eventuali profili di riservatezza emergenti. Ci si può aspettare che il tempo complessivo di attesa della risposta sia in media assai breve: tuttavia si prega di attendere almeno 2 giorni lavorativi prima di chiedere informazioni.

Il sistema può restituire automaticamente i risultati delle elaborazioni che non tentino di violare la riservatezza dei dati elementari: tuttavia, le eventuali revisioni di sicurezza manuali vengono effettuate esclusivamente nei giorni lavorativi dal lunedì al venerdì. La Banca si riserva sempre la possibilità di controllare manualmente le elaborazioni. I lavori che dovessero risultare bloccati in coda di revisione durante un fine settimana o un periodo festivo verranno esaminati appena possibile.

7. Limitazioni del sistema

Non è possibile sottomettere più programmi in linguaggi diversi nella medesima e-mail. È possibile, invece, sottomettere più programmi nello stesso linguaggio nella medesima e-mail.

La Banca si riserva di stabilire limiti in termini di tempo macchina, memoria dinamica, memoria di massa, numero di statement del programma, linee di output, in funzione delle risorse disponibili a tutela della stabilità e della massima fruibilità del sistema.

Gli output molto lunghi vengono sempre sottoposti a revisione di sicurezza; oltre una certa lunghezza, attualmente posta pari a 5.000.000 byte, vengono automaticamente cancellati.

Non è possibile salvare dataset permanenti. Il salvataggio di dataset temporanei è consentito secondo modalità specifiche di ciascun pacchetto applicativo, reperibili nelle Appendici.

Per motivi di riservatezza, come analizzato nel paragrafo seguente, un interprete dedicato esclude dall'elaborazione i programmi che contengono taluni comandi o parole vietate.

Si richiama l'attenzione sulle caratteristiche e limitazioni specifiche di ciascuna piattaforma supportata, reperibili nelle Appendici.

8. Riservatezza

Al fine di garantire al contempo la riservatezza dei dati individuali e adeguate funzionalità di elaborazione statistica ed econometrica, sono stati stabiliti presidi per garantire la riservatezza dei dati individuali.

Sono pertanto vietate tutte le elaborazioni che possono rivelare i dati delle singole imprese. Ad esempio è proibita la stampa di singoli record o di parte di essi (ad esempio, è stato inibito il comando "list" in Stata). Il sistema è progettato per impedire l'esecuzione di alcuni comandi che potrebbero comportare la violazione della riservatezza delle imprese del campione.

Il controllo sulla riservatezza dei dati individuali è svolto anche a livello di logica del programma sottomesso. A questo fine la Banca d'Italia può controllare il contenuto delle elaborazioni effettuate e conservarle senza limiti di tempo. La Banca d'Italia si riserva il diritto di non trasmettere tutto o parte del risultato di un'elaborazione al fine di tutelare la riservatezza delle imprese.

È vietato l'utilizzo della casella di e-mail **bird@bancaditalia.it** per l'invio di messaggi diversi dalla sottomissione di programmi elaborativi secondo gli standard ed i formati prescritti in questo manuale. La casella è gestita in modo automatico e parimenti in modo automatico i messaggi vengono letti e smistati.

Abusi nell'utilizzo del sistema di elaborazione a distanza possono portare alla sospensione o alla revoca dell'abilitazione.

9. Suggerimenti generali di utilizzo

Al fine di ottimizzare l'utilizzo delle risorse e ridurre al minimo la probabilità di una revisione di sicurezza e il ritardo ad essa conseguente, o la cancellazione totale del lavoro sottomesso, si raccomanda di:

- **non tentare di stampare dati elementari;**
- **non creare tavole statistiche con celle contenenti un numero esiguo di osservazioni;**
- **non creare distribuzioni di frequenze di variabili continue;**
- **evitare di creare output eccessivamente lunghi.**

10. Segnalazione dei lavori di ricerca svolti

Per consentire la raccolta del materiale bibliografico riguardante l'indagine, l'autore è invitato a far pervenire una copia dei lavori di ricerca condotti utilizzando il sistema BIRD all'indirizzo e-mail: **bird_assist@bancaditalia.it**. Per gli utilizzatori di LaTex, ove disponibile, si prega di allegare anche un file di testo contenente le intestazioni BibTex[1] del lavoro.

11. Assistenza e supporto

L'assistenza è dedicata all'uso del sistema e degli archivi ed è fornita di preferenza via e-mail, tramite la casella **bird_assist@bancaditalia.it**. Si raccomanda di specificare l'oggetto nel campo apposito. **Questa casella <u>non può</u> essere usata per sottomettere le elaborazioni.**

[1] Il file di testo contenente le intestazioni BibTeX dovrà contenere tutte le necessarie informazioni per identificare il lavoro tra le quali il titolo, l'anno di pubblicazione, gli autori e la rivista o la collana di working paper in cui è stato pubblicato.

12. Appendice 1: indicazioni specifiche per il linguaggio Stata

12.1. Versione supportata e default di sistema

Versione supportata: 15.0 per Windows. Il sistema contiene tutte le .ado comunemente presenti nel pacchetto Stata di pari versione.

Comandi automaticamente posti in testa a ciascuna elaborazione e non personalizzabili:

set more off , permanently;
set maxiter 300 , permanently;
set mem 300m;

12.2. Indicatore di fine riga

Nei programmi in Stata da sottomettere al sistema BIRD è necessario usare come indicatore di fine riga il carattere ; (punto e virgola). A tal fine è necessario premettere il comando "#delimit;" a ciascuna elaborazione. Il semplice "a capo" può dare origine a errori nel riconoscimento dei comandi. Si ricorda inoltre che ciascuna riga non può contenere più di 72 caratteri.

12.3. Accesso ai dataset

La libreria contenente i dati dell'indagine sulle imprese industriali e dei servizi viene invocata con il nome simbolico **$db_001.**

Il sondaggio congiunturale è accessibile usando il nome simbolico **$db_sondstor** per l'archivio storico, mentre per quelli relativi ai singoli sondaggi occorre usare i nomi **$db_sond*xxxx***, dove *xxxx* è l'anno di effettuazione dell'indagine.

Per un'allocazione ottimale delle risorse, in particolare per l'indagine sulle imprese industriali e dei servizi, è richiesto che per ciascuna elaborazione siano selezionate le variabili strettamente di interesse per l'elaborazione, operando al contempo le necessarie selezioni orizzontali. La sintassi da utilizzare è la seguente:

```
use    <elenco delle variabili d'interesse>
       if <condizioni sulle variabili>
        using $db_001;
```

Per dei casi concreti si rimanda alla documentazione riguardante gli esempi.

12.4. Suggerimenti di utilizzo

Non usare comandi in grado di violare la riservatezza dei dati individuali (ad esempio, **list**) o l'integrità del sistema. Non tentare di usare comandi di sistema quali **exit** o il prefisso "**!**". I lavori contenenti comandi od opzioni vietate vengono cancellati automaticamente. L'uso ripetuto di comandi od opzioni vietate può condurre alla revoca dell'abilitazione, salvo ogni ulteriore effetto di legge.

Il linguaggio Stata distingue le maiuscole dalle minuscole. Si raccomanda l'uso dei caratteri minuscoli salvo ove espressamente richiesto.

Alcuni comandi Stata, come ad esempio **probit** e **logit**, possono produrre output molto lunghi. **Le elaborazioni il cui output è troppo lungo possono essere sottoposte a revisione di sicurezza o cancellate.** Per taluni comandi, l'opzione "**, nolog"** è in grado di abbreviare l'output. In generale, l'utilizzo adeguato dei comandi "**noisily**" e "**quietly**" può abbreviare la lunghezza dell'output complessivo.

Per salvare dataset intermedi usare il riferimento simbolico **$mydata** indicando di seguito il nome utente, come nel seguente esempio riferito all'utente "maria" che salva il file temporaneo **myfile.dta**:

save $mydata\maria\myfile, replace;

I dataset intermedi vengono cancellati al termine di ciascuna elaborazione.

13. Appendice 2 : Indicazioni specifiche per il linguaggio R

13.1. Versione supportata e default di sistema

Versione supportata: 3.3.1 per Windows.

Nel linguaggio R non è consentito al momento memorizzare dataset intermedi o altri file, come ad esempio grafici o file PDF.

13.2. Accesso ai dataset

Per l'**Indagine annuale**, il file contenente i dataset accessibili si chiama **indann.rdata**. Il dataset contenente l'indagine sulle imprese industriali e dei servizi si utilizza con l'istruzione:

```
invind <- get(load("indann.rdata"))
```

la quale carica nella variabile denominata "invind" il relativo dataset.

Per il **Sondaggio congiunturale**, l'archivio storico si utilizza con l'istruzione:

```
sstor <- get(load("sondstor.rdata"))
```

mentre gli archivi annuali del sondaggio possono essere indirizzati con l'istruzione:

```
sdXXXX <- get(load("sondXXXX.rdata"))
```

dove *xxxx* è l'anno di effettuazione dell'indagine.

13.3. Suggerimenti di utilizzo

Non usare comandi in grado di violare la riservatezza dei dati individuali o l'integrità del sistema. I lavori contenenti comandi od opzioni vietate vengono cancellati automaticamente. L'uso ripetuto di comandi od opzioni vietate può condurre alla revoca dell'abilitazione, salvo ogni ulteriore effetto di legge.

Le elaborazioni il cui output è troppo lungo possono essere sottoposte a revisione di sicurezza o cancellate. Tra le opzioni di ciascun comando, si suggerisce di selezionare quelle che garantiscono l'output di lunghezza minima compatibile con il livello di dettaglio richiesto.

Si ricorda che un lavoro può essere cancellato se l'utilizzo delle risorse, in termini di tempo macchina, memoria dinamica o memoria di massa, risulta eccessivo. Per ridurre lo spazio in memoria si consiglia di lavorare su *data frame* contenenti le sole variabili di interesse.