

# データ解析の電子ジャーナル EJDA の実働化

—— registered ftp の提案と実装 ——

統計数理研究所 丸 山 直 昌

(1994 年 3 月 受付)

## 1. 序

近年、コンピュータとコンピュータ通信の技術が急速に発展するなかで電子的な学術雑誌(電子ジャーナル)の必要性和実現可能性が多くの人々から指摘されている。すなわち、学術雑誌を従来のような印刷物としてではなく、コンピュータ上で編集しコンピュータネットワークを利用して電子的に配布するのである。

学術雑誌を電子化することは、配布を敏速化する、記事内容を検索可能とする、紙資源の節約になる、論文の長さの制約を緩くできる、などの利点が考えられる。この中で統計学の場合には最後の論文の長さの制約の問題は極めて重大な影響がある。従来の印刷形態の雑誌ではこの長さの制約のために、研究対象となっている統計データを掲載せずにデータの解析手法と解析結果のみを掲載するということになりがちであった。しかしそれでは論文を読んだ第三者が解析結果を再検証できない。また仮に長さの制約を緩めてデータを掲載してみたところで、多量のデータの場合には印刷されたものを再入力するのに多大の労力がかかるので、やはり第三者による再解析は実際上困難と言わざるを得ない。データを個人的な努力によって磁気テープなどを用いて配布している統計学者もいるが、統計学において電子ジャーナルを始めることは、このような個人レベルではなく、組織立ったデータの流通路を開くという重大な意義を持つ。

統計数理研究所においては、このような視点に立った電子ジャーナル発刊に向けての研究は早くから行なわれていた。一連の共同研究(課題番号 62-共研-6, 63-共研-1, 1-共研-6, 2-共研-6, 3-共研-14, 4-共研-19, 5-共研 A-22)において研究された EJDA (Electronic Journal of Data Analysis) である。慶應大学の渋谷政昭、柴田里程両氏を中心として展開されてきたこれらの共同研究に筆者は途中から参加し、主としてジャーナル実現のための技術的側面を担当した。本稿では実働段階に入った EJDA について、電子的配布方式の技術的な面に焦点をあてて述べる。特に筆者が創案して EJDA に初めて採用した新しい方式の ftp (registered ftp と名付けた) の考え方および実現方法を解説する。なお「EJDA 利用マニュアル」を付録として添えた。

## 2. EJDA の構想

EJDA はデータ解析の電子ジャーナルとして慶應大学の渋谷政昭、柴田里程両氏によって構想された(渋谷・柴田(1987))。この構想は、ジャーナルの配布に世界的な広がりを持つインターネット(The Internet)を利用すること、さらにデータの流通という観点から、データの標準記述方式として D&D (Data and Description) を採用すること(渋谷・柴田(1991))、の 2 点

を骨子としている。

### 3. インターネット

EJDA がジャーナルの配布路として選んだインターネット (The Internet) というコンピュータネットワークについて少し述べる。

インターネットは通信規格として TCP/IP (Transmission Control Protocol/Internet Protocol) を採用した全世界規模のネットワークである。1980 年代の初頭に TCP/IP はアメリカ国防総省の財政的支援を受けた DARPA (Defense Advanced Research Projects Agency) によって開発され、ARPANET (Advanced Research Projects Agency NETwork) 上で初めて使用された。これがインターネットの出発点であった。その後、1980 年代後半には全米科学財団 (NSF, National Science Foundation) の財政支援のもとで TCP/IP を用いた NSFNET の運用が始まり、インターネットはアメリカで大きく発展してゆく。

日本においても 1980 年代の終り頃から WIDE (WIDE Internet), TISN (国際理学ネットワーク, Today International Science Network), JAIN (Japan Academic Inter-university Network) などの TCP/IP のネットワークが生まれて日本国内で相互に接続され、またアメリカとも接続されてインターネットに仲間入りしている。同様の動きはヨーロッパでも着実に進行している。

ここ 1, 2 年は世界中で爆発的とも言える広がりを見せ、1993 年 9 月の時点で世界 92 箇国以上に IP 接続が延び、接続コンピュータ数は少なくとも 200 万台 (1993 年 10 月)、電子メール到達国数 137 箇国 (1993 年 8 月)、推定利用者数は 1500 万人以上 (1994 年始め) と見積もられている (Internet Society (1994))。このようにインターネットは、学術研究交流の基盤のコンピュータネットワークとして既に揺るぎ無い地位を築いたと言える。

統計数理研究所は 1990 年 12 月に TISN に加盟してインターネットの一員となった。これにより世界中のインターネット参加組織と電子メールの交換や ftp (File Transfer Program) によるデータ交換が可能となった。

### 4. EJDA のインターネット接続

EJDA は統計数理研究所内に専用のマシンを一台設置し、これをサーバーマシンとしてジャーナルの配布サービスを行なうこととした。サーバーマシンには現在 Sparc Station 1+ (ホスト名 `ejda.ism.ac.jp`, SunOS 4.1.2) を充てている。サーバーマシンは研究所内のネットワークに接続されているので、TISN を通じてインターネット全域と通信可能となっている。ジャーナルの配布は TCP/IP のアプリケーションプログラムである電子メールと ftp を応用して実現した。その詳細を以下に述べる。

#### 4.1 電子メールによるジャーナル配布

電子メールによるジャーナル配布は、電子メールの自動応答プログラムを用いて行なう。すなわち電子メールアドレスを決めておいて、そのアドレスに一定の書式で要求メールを出すとそれに応じてジャーナルの一部分を自動的に送り返すプログラムである。このような自動応答プログラムはインターネット内ではいくつか例があるが、EJDA では AT&T の Bell 研究所の Netlib を用いた。電子メールアドレスは `ejda@ism.ac.jp` であり、利用方法を付録の「EJDA 利用マニュアル」に示した。

## 4.2 ftp によるジャーナル配布

EJDA の ftp による配布は、今までとは少し違った ftp を用いて行なうことにした。これは従来の ftp では配布先の把握に難があるという考えからである。(配布先の把握は利用状況調査のため、料金を取る意図では無い。)この点を解決するために registered ftp と名付けた新しい方式の ftp を提案し実装したが、この間の事情を説明するために、まず従来からある anonymous ftp について解説する。

### 4.2.1 anonymous ftp

TCP/IP における ftp プログラムは通常、インターネット上の二つのホスト A と B の両方に利用資格を持つ人が A にログイン中に A と B の間でファイルの複写を行なうというかたちで使われる。利用者は A にログインするために A での登録名と暗証(パスワード)を入力するが、B との間で ftp コマンドを使う時には B での登録名と暗証も入力する。さてインターネット上では無償ソフトウェアの流通に anonymous ftp と言われる方式が以前から盛んである。これはホスト間のファイルの複写に通常の ftp コマンドを使うが、通常のユーザ登録名ではない特別のユーザ登録名“**anonymous**”あるいは“**ftp**”を使う。するとこの場合も普通と同じように暗証の入力を要求されるが、実際には暗証の検査は行わずに無条件で接続を許し、ファイルの複写を認める。インターネット上のあるホスト X の管理者が、X のディスクのうちの一部を anonymous ftp によって公開したいと思う場合、管理者は **ftp** というユーザ名を `/etc/passwd` に登録し、**ftp** のホームディレクトリの下にいくつかファイルを設定する。すると **ftp** のホームディレクトリ以下にあるファイルが anonymous ftp による公開の対象となる。設定作業の詳細は、例えば SunOS 4.1 の場合ならば、オンラインマニュアルの `ftpd(8C)` に書かれているが、ともかく管理者が必要な設定を行えば anonymous ftp はすぐにでも始められる。

このようにしてホスト X で anonymous ftp が設定されていると、インターネット上の他のホスト A の利用者は、X の利用資格が無くとも X の管理者が公開の対象としたファイルは A に複写することができる。一連の手続きで、X の管理者は anonymous ftp の開始にあたっていくらかの設定作業が必要だが、A の方では利用人も、また A の管理者も特別な準備を必要とせず、通常の ftp の利用と同じようにしてできる。anonymous ftp はこのように簡便なものなので、配布に制限が設けられていない無償ソフトウェアを広める手段としてインターネットの世界では現在広く普及している。

一方 anonymous ftp の短所として、X の側からはホスト A のどのユーザであるかを特定できない、という問題がある。これは単に ftp の問題と言うよりは TCP/IP の基本設計にかかわる問題で、安易に評価できる事ではないが、インターネット上ではこれに対して一つの行動規範が推奨されている。それは anonymous ftp においては暗証の入力要求に対して自分の電子メールアドレスを答える、というものである。暗証の検査は anonymous ftp では実際には行なわれないので何を答えても ftp 接続が拒まれることはないが、ここで自分が誰であるか名乗ってください、というわけである。これはあくまで推奨であって強制ではないので、守られるかどうかかわからない。また守る意思があっても暗証入力機構を利用しているために打鍵がユーザ側にはエコーバックされず、このため見えないのが普通であり、しばしば打鍵誤りによって不正確な情報となる。システムを改造して anonymous ftp のときだけエコーバックされるようにすれば良い、と思うかもしれないが、それは最早ホスト X の改造だけでは済まない問題で、ホスト A の方を直す必要が出てくる。anonymous ftp を行なうホスト X の管理者が世界中に百万台以上もあるホスト A、A'、A''... の改造を行なうことは不可能である。

#### 4.2.2 registered ftp の提案と実装

EJDA では ftp による配布については、当初 anonymous ftp の利用も考えた。しかしジャーナルの配布先を確実に把握しておきたいという意見は強く、anonymous ftp の持つ短所の克服は一つの課題となった。そこでこの点を解決するために筆者は registered ftp と名付けた新しい方式の ftp を提案して EJDA サーバマシンに実装した。

registered ftp では、利用者はジャーナル閲覧の前に登録申請を電子メールで行なう。この登録はジャーナルの閲覧専用の登録で、サーバマシンの通常の意味でのユーザ登録とは全く別のものとし、サーバマシンの自動処理プログラムが申請の受け付け処理を行なう。ジャーナル閲覧は anonymous ftp と同じ手順を取るのであるが、暗証の入力要求に対して打鍵された電子メールアドレスと利用者が使っているホストの IP アドレスは登録簿と照合される。メールシステムが正しく設定されているホストから発信される電子メールの発信人アドレス欄は比較的信頼性が高いと考えられるので、閲覧者の特定という目的には、registered ftp は十分な効果を持つ。

利用者にとって、registered ftp は通常 anonymous ftp に比べて著しく使い勝手が悪くなっているわけではない。なぜならば、インターネット内では、ftp が支障なく使える条件下では電子メールも十分な速さで配達されるので、登録手続きが必要と言ってもせいぜい 1, 2 分程度の待ち時間で済む。例えば閲覧希望者が新方式の ftp と知らずに EJDA のサーバマシンに ftp 接続を試みて、そこで表示されるメッセージを見て通常 anonymous ftp とは違うことに気が付いてから登録申請メールを出したとしても、たいした手間ではない。しかもこれは始めの一回だけで、次回以後は既に登録済みであるから anonymous ftp と同じ労力で済む。付録の「EJDA 利用マニュアル」ではこの一連の手順を実例で示している。

ここで重要なことは、registered ftp は、これを行なうサーバマシンの改造だけで実施可能であって、インターネット上の他のホストは全く変更なしで今まで通りの ftp コマンドの使用で済むことである。

registered ftp の実装は EJDA サーバマシン (SunOS 4.1.2) の `/usr/etc/in.ftpd` の変更によってできる。SunOS の in.ftpd のソースプログラムは公開されていないのでパケレイ版 Unix の中にある無償の ftpd (Version 5.60) を改造してこの機能を実現した。改造の細部をここでいちいち述べることは繁雑に過ぎるので、代わりに EJDA サーバマシン `ejda.ism.ac.jp` の ftp 公開ディレクトリに `SYSTEM/ftpd` というサブディレクトリを設けて、ソースプログラムを registered ftp によって公開する。

### 5. 今後の課題

registered ftp の実装により EJDA は、ジャーナル閲覧のシステムとしては体制が整った。今後はジャーナルの中身を充実していく必要がある。そのためには当面次の二つの課題に取り組む必要がある。

一つは D&D のサポートソフトウェアの充実である。EJDA ではデータの記述方式として D&D というデータ記述文法を採用することにした (渋谷・柴田 (1991))。これはジャーナルの内容をコンピュータソフトウェアやハードウェアに依存しないものにするという趣旨であり、非常に重要な思想である。ところが D&D の普及の遅れが EJDA 発展の足かせとなっている面もある。編集者としてはデータの投稿は D&D 形式であって欲しいが、投稿者の立場から見れば D&D 形式のデータを扱うためのソフトウェア——D&D サポートソフトウェア——が充分になれば D&D 形式でデータを記録する価値は少ないということになる。その結果 D&D 形

式でないデータが投稿され、それを D&D 形式に直す作業が EJDA 編集関係者の労力負担となっている。D&D サポートソフトウェアの充実はその意味で当面の大きな課題である。

もう一つの課題は審査体制である。学術雑誌を標榜する以上、EJDA も当然のことながら投稿論文の審査を行なう必要がある。しかし、現時点では十分な数の審査員が確保できているとは言えない。最大の原因は、D&D 文法が十分に普及していないために D&D で書かれたデータを審査できる人が極めて少ないという点であろう。当面は審査を無しにしても D&D 形式のデータを多く集積し、D&D の普及に努める必要がある。

## 6. 謝 辞

本稿の内容は、統計数理研究所共同研究「“データ解析の電子ジャーナル (EJDA)” の創造」(課題番号 3-共研-14)、「“データ解析の電子ジャーナル (EJDA)” の実働化」(課題番号 4-共研-19) に端を発したものである。ここで論ずる registered ftp の創案と実装に当たっては、これら共同研究の研究代表者である柴田里程氏、並びに共同研究員である渋谷政昭氏に多くの助言を頂いた。ここに感謝の意を表したい。

## 参 考 文 献

- Internet Society (1994). Abstract 等 ([cnri.reston.va.us](http://cnri.reston.va.us) の anonymous ftp より入手可能).
- 渋谷政昭, 柴田里程 (1987). 電子ジャーナル “Electronic Journal of Data Analysis” の構想, 統計数理, 35, 81-87.
- 渋谷政昭, 柴田里程 (1991). EJDA におけるデータ記述, 統計数理, 39, 85-96.
- 統計数理研究所共同研究レポート 54 (1994). “データ解析の電子ジャーナル (EJDA)” の実働化.

## 付録

### EJDA 利用マニュアル

第七版 1994.9

EJDA を使うには詳しい説明資料は必要ありません。とっかかりさえわかればあとはオンラインで情報がとれるからです。ここでは、1. インターネットの電子メールを使う方法、2. インターネット上の UNIX の ftp コマンドを使う方法、にわけてなるべくわかりやすく説明をします。

1. インターネットの電子メールを使える方は以下に示すように「send index」というたった一行の本文を含むメールを ejda@ism.ac.jp 宛にお送りください。

```
% mail ejda@ism.ac.jp
Subject:                ← Subject は不要です。リターンのみ入力
send index              ← 一行のみの本文
^D                       ← control-D で終了して送信
%
```

あとは、詳しい資料が自動的にメールで返送されますので、それをご覧ください。

2. インターネット上で ftp コマンドを使える方は ejda.ism.ac.jp (IP アドレス 133.58.14.13) に ftp コマンドで接続して EJDA を利用することができます。但しこの EJDA の ftp はジャーナルの閲覧者を正確に把握するために、読者事前登録方式 (registered ftp) を取っています。登録の方法も ftp コマンドで接続すると表示されますので、とりあえず登録以前でも試しに ftp コマンドで接続してみてください。

```
% ftp
ftp> open ejda.ism.ac.jp
Connected to ismspc3.ism.ac.jp.
220-
```

```
Welcome to Electronic Journal of Data Analysis (EJDA) FTP server.
To subscribe the Journal, please use "ftp" or "ejda"
as the "Name"(User name).
```

```
220 ismspc3.ism.ac.jp FTP server (Version 5.60) ready.
Name (ejda.ism.ac.jp:maruyama):
```

ここで, ejda と入力して Return キーを押します. すると

331-

```
-----
Entering the Journal subscribe mode.
-----
```

To maintain list of subscribers of the journal, here we provide "registered FTP service"; a subscriber needs to make registration via e-mail before FTP access. If you have not yet been registered, please send the following registration form to ejda-ftp@ism.ac.jp:

```
add UID MACHINE IP-ADDRESS EMAIL-ADDRESS
```

where UID is your uid, MACHINE is your machine's name in internet full domain format, IP-ADDRESS is it's IP address in internet dot notation, and EMAIL-ADDRESS is your email address.

Now you are accessing from sunnm.ism.ac.jp (133.58.1.18).  
331 Type your mail address or UID as password.  
Password:

となります. 以下登録済みのメールアドレス或は UID を入力すれば

```
230 Member maruyama on sunnm.ism.ac.jp (133.58.1.18) login succeeded.
ftp>
```

となり, get コマンドによるジャーナルの取得が可能です.  
上のメッセージからもわかるように, 登録依頼のメールは

```
% mail ejda-ftp@ism.ac.jp
Subject:          ← Subject は不要です. リターンのみ入力
add UID MACHINE IP-ADDRESS EMAIL-ADDRESS
^D              ← control-D で終了して送信
```

とします. ここに UID はユーザ識別名, MACHINE はマシンの名前 (インターネット完全ドメイン名—Internet full domain name), IP-ADDRESS は IP アドレス (インターネットドット記法) です. sunnm.ism.ac.jp (133.58.1.18) というマシンの maruyama という UID の人で, maruyama@ism.ac.jp というメールアドレスの人ならば

```
% mail ejda-ftp@ism.ac.jp
Subject:
add maruyama sunnm.ism.ac.jp 133.58.1.18 maruyama@ism.ac.jp
^D
```

です。

### 3. 電子メールアドレス

1, 2に述べた `ejda@ism.ac.jp` と `ejda-ftp@ism.ac.jp` を含めて次のようなメールアドレスを EJDA では使用します。

<code>ejda@ism.ac.jp</code>	EJDA の電子メールによる閲覧
<code>ejda-ftp@ism.ac.jp</code>	EJDA の Registered ftp の利用登録
<code>ejda-submission@ism.ac.jp</code>	EJDA への投稿
<code>ejda-editor@ism.ac.jp</code>	EJDA 編集長への手紙
<code>ejda-request@ism.ac.jp</code>	EJDA への提案, 事務連絡

### 4. FAQ(Frequently asked questions— 想定問答)

**問** ftp で `ejda.ism.ac.jp` とするホスト名を指定すると `host unknown` と言って叱られる。

**答** お使いのマシンでは Domain name server の resolver routine の設定がされていないためです。この設定は一般ユーザには無理なので、IP アドレス (133.58.14.13) を指定してください。

**問** 今自分が使っているマシンのメールアドレスは何？

**答** メールシステムの設定方法は多種多様で一般性のある調べ方はありません。システム管理者に聞くのが一番確実ですが、メールをどこかに送ってみて、そのメールの From 行を見るのがかなり確実性が高い方法です。ejda を利用して、1. の方法で資料請求して返事が返ってくれば、その返事の To: 行があなたのメールアドレスと考えて良いでしょう。返事が返ってこなければ、メールシステムが正しく設定されていない可能性を疑う必要があります。

**問** 今自分が使っているマシンの IP アドレスは何？

**答** UNIX のマシンであれば、`netstat` というコマンドで調べられます。

```
% netstat -in
```

詳しくは `man netstat` を見てください。

**問** 今自分が使っているマシンのホスト名は何？

**答** UNIX では `hostname` というコマンドでホスト名が表示されますが、表示されるホスト名がピリオド "." を含んでいなければこれは同一組織内でのみ通用する省略名で、インターネット完全ドメイン名ではありません。インターネット完全ドメイン名は domain name server というデータベースに登録されて組織外に公表されているもので、もし `nslookup` というコマンドが正しく設定されていて使えれば、

```
% nslookup
```

```
Default Server: ismspc2.ism.ac.jp
```

```
Address: 133.58.1.12
```

```
> sunnm
```

← ここでホストの省略名を入力

```
Server: ismspc2.ism.ac.jp
```

```
Address: 133.58.1.12
```



Name: sunnm.ism.ac.jp

Address: 133.58.1.18

のように、省略名を与えるとインターネット完全ドメイン名と IP アドレスを教えてください。この手が使えなければ、2. の方法で EJDA の ftp サーバに接続すると、domain name server を調べて

Now you are accessing from sunnm.ism.ac.jp (133.58.1.18).

のように教えてくれますので、これが確実です。ただしここでもし、

Now you are accessing from 133.58.1.18 (133.58.1.18).

のように表示されるようでしたら、Domain name server への登録が不完全で、これ以上はわかりません。この場合 ejda-ftp@ism.ac.jp へは、ホスト名も 133.58.1.18 であるかのようにして登録して差し支えありません。

**問** Registered ftp の登録依頼メールの書き方を間違えたときにはどうしたら良いか？

**答** 登録依頼メールを出し直してください。

Implementation of the Electronic Journal of  
Data Analysis EJDA  
— Proposal for Registered Ftp and Its Implementation —

Naomasa Maruyama

(The Institute of Statistical Mathematics)

EJDA (Electronic Journal of Data Analysis) was first planned by Sibuya and Shibata (1987), and was carried out as a series of Cooperative Research Programs of the Institute of Statistical Mathematics (87-ISM.CRP-6, 88-ISM.CRP-1, 89-ISM.CRP-6, 90-ISM.CRP-6, 91-ISM.CRP-14, 92-ISM.CRP-19, 93-ISM.CRP-A22). There are two essential ideas in the plan. The first is to use the Internet, and the second is to use a formal data description language D&D (Sibuya and Shibata (1991)). The author joined the EJDA program in 1990, and have contributed to the implementation of journal server system. In this paper we present how we implemented EJDA. Especially, we describe a new type of ftp server system based upon our idea on ftp protocol, which we named 'registered ftp'.

Access from the Internet to EJDA is already available by two methods. One is by e-mail, and the other is ftp. For e-mail access, we have set up an automatic replying system at the address **ejda@ism.ac.jp** using NETLIB of AT&T Bell Laboratory. For ftp access, we have implemented the above mentioned 'registered ftp' on the server machine **ejda.ism.ac.jp**.

As is well known, so-called anonymous ftp is very much successful in distributing copy-free software in the Internet. However, anonymous ftp has one weak point in identifying the user. It is possible to identify from which host the access is made, but it is impossible to identify who on the host having the access. To maintain the correct list of subscribers of the journal, we have implemented registered ftp on EJDA server machine. Registered ftp server program checks pre-registered member list at each journal subscribing access. Membership registration must be made before ftp access by an e-mail. After the registration, ftp access procedure works just like in case of anonymous ftp.

The implementation was done by replacing **/usr/etc/in.ftpd** of the server machine with our version, which we made from free source of ftpd version 5.60 in Berkeley UNIX. We made the program little bit loquacious so that one may get enough information about the registration procedure by ftp access to the host. The source code of the program is accessible from the server machine itself by registered ftp service in the directory **SYSTEM/ftpd**.