DandD Environment for financial data

Daisuke Yokouchi, Hitotsubashi Univ.
yokouchi@ics.hit-u.ac.jp

Utilization of the Japan stock market data

- The data contains information of all stocks listed on all markets in Japan from January 1977 to March 2007.
- The data file is organized as a text file and its size is over 2GB.

(Most of software in average PC can not even open the data file !)

Organization of the data into a DandD instance

Once the data is converted into a DandD instance,

- users can obtain background information about the data through the DandD Browser.
- data can be imported into the statistical software R through the DandDR.

The users can easily start data analysis anytime they want to.

Data size problem

We organized the data into a DandD instance through the DandD Editor.

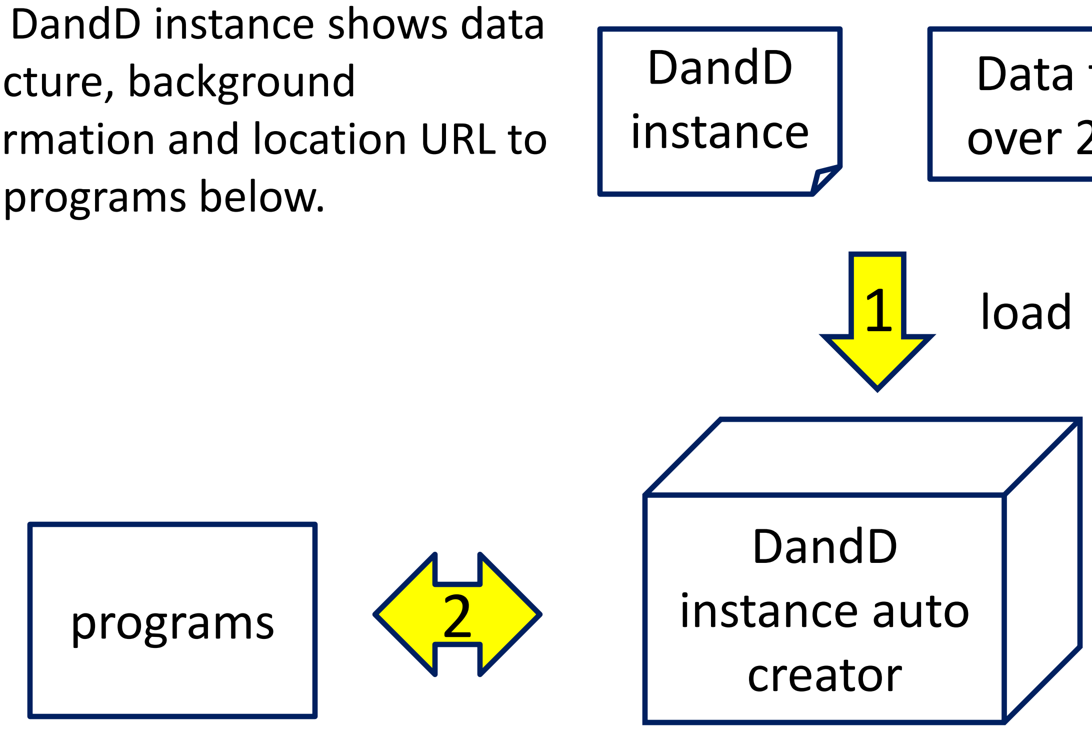
However, the size prevents us from using the created DandD instance.

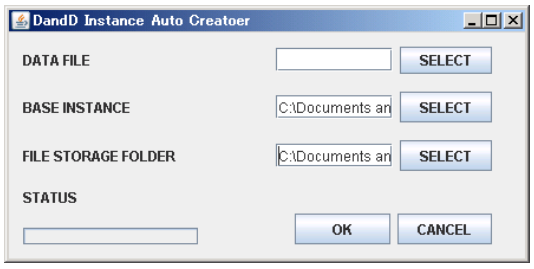
A solution of the size problem

Most of the users are interested in the behavior of prices of a stock.

We created an application program that divides the DandD instance into a number of DandD instances each of which contains individual stock information .

The DandD instance shows data structure, background information and location URL to the programs below.




DandD Instance Auto Creatoer

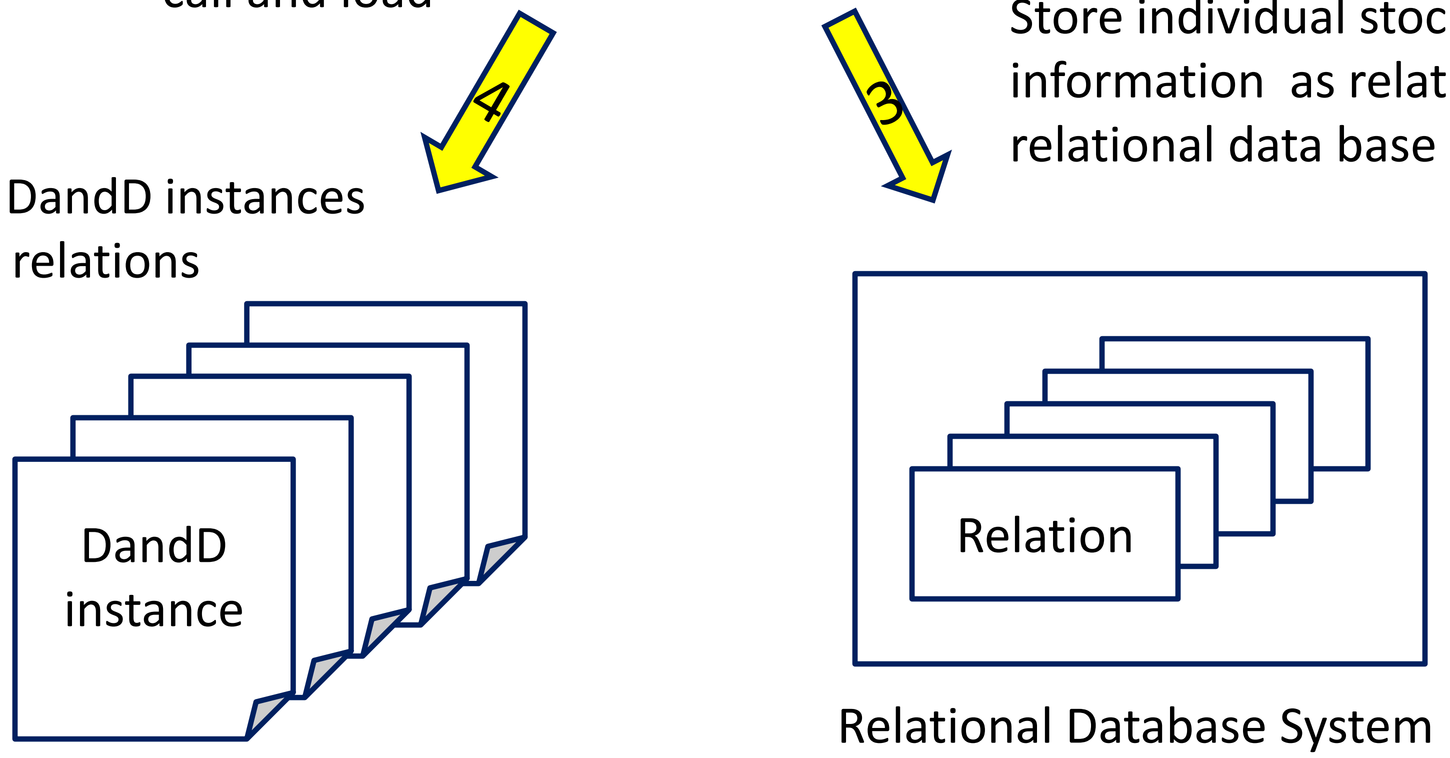
DATA FILE

BASE INSTANCE C:\Documents an

FILE STORAGE FOLDER C:\Documents an

STATUS

call and load



Store individual stock information as relations in a relational data base system.

Create DandD instances to each relations

Relational Database System

Conclusion

Using DandD Environment gives us two advantages below.

Data Auditing

Everyone can verify the data processing to create the DandD instance by examining the program.

Easiness of database maintenance

To update of the Japan stock market data, all the users have to do is to replace the data file as long as the structure of the updated data is invariant. Then all the created DandD instances can be updated automatically.

DandD Project home page

The latest DandD rule and several software are downloadable freely from the DandD project homepage

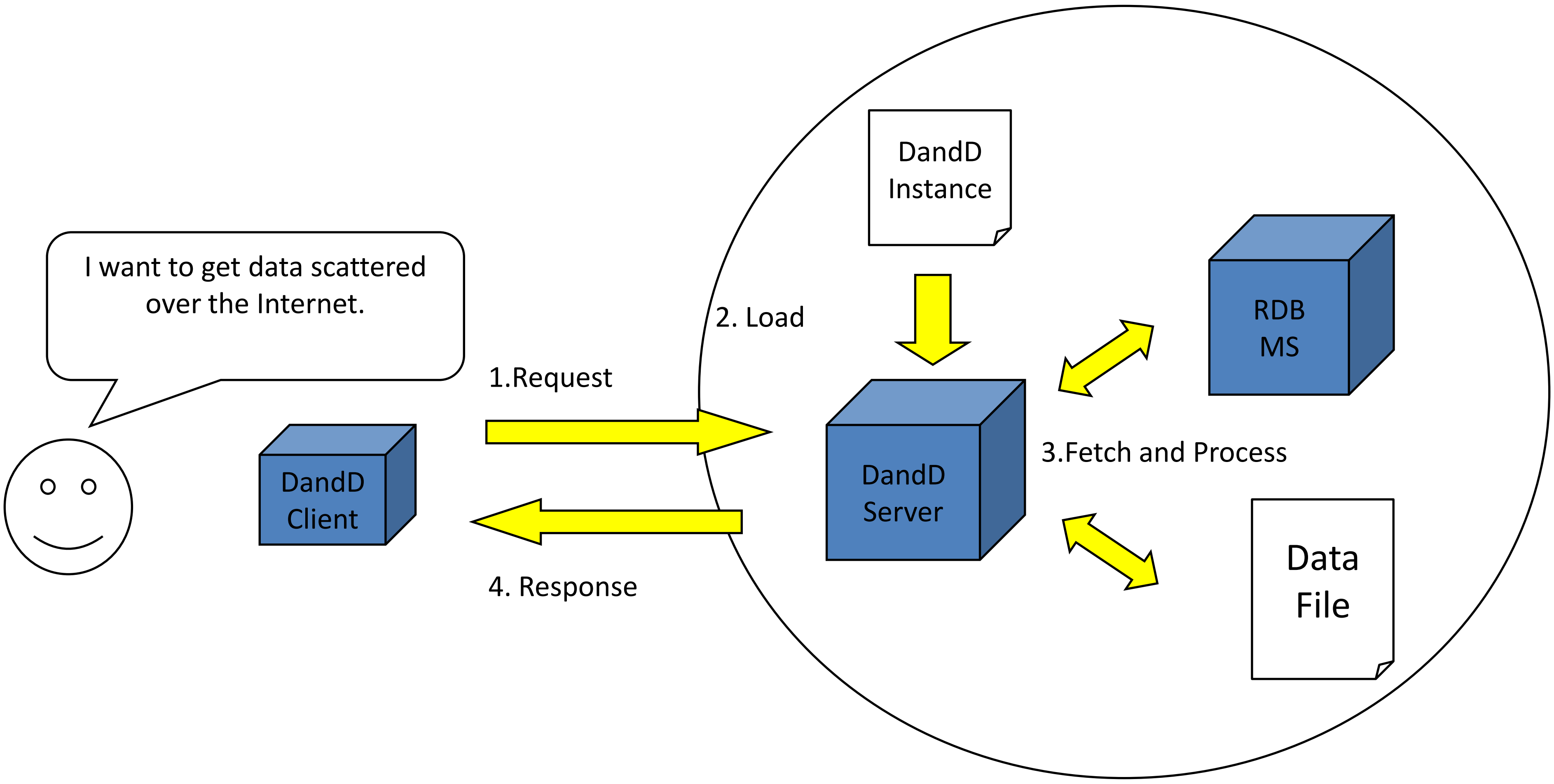
http://www.stat.math.keio.ac.jp

APPENDIX

DandD Client Server System

- Advantage of developing support software using client server system
  - Easiness of programming support software
    - Any language can be used as far as it supports socket handling.
  - Flexibility
    - When DandD rule is modified, the client programs works as same as before, in most cases
  - Mobility
    - The size of client program can be reduced, so it is installable on low ability machine like PDA or cellular phone.

An example of use of data scattered over the Internet using the DandD environment



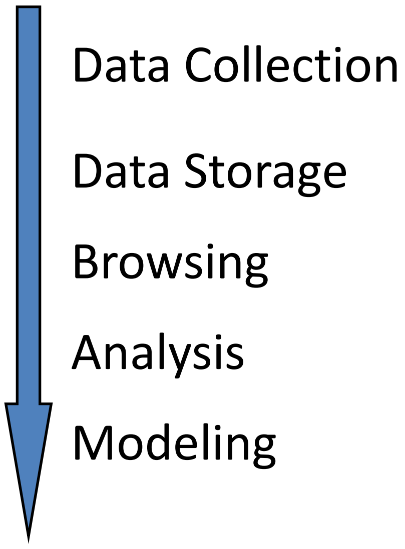
We can easily acquire data scattered over the Internet by using DandD instance and DandD Client Server System.

DandDServer

- Purpose
  - Provision of sufficient methods for manipulating DandD instance for clients.
- Recieve from clients
  - DOM (Document Object Model) methods & original methods
- Send to clients
  - Flag + Size + String
    - Flag which tells us whether execution of method is successful or not
    - Size which is a size of String
    - String which is a result of the executed method
- Implementation
  - Language
    - Java
  - API for handling XML document
    - Xerces for Java2 (Apache Project)
  - Access to relational database management system on the network
    - JDBC (Sun Micro Systems)
  - Interpreter
    - Pnuts (Sun Micro Systems)

DandD Client Programs

- DandD Editor
  - is the editor for creating a new DandD instance and editing existing DandD instances.
- DandD Browser
  - is the software for browsing a DandD instance.
- DandDR
  - is an interface between DandD Server and R which is a statistical software for data analysis and modeling



Data Collection

Data Storage

Browsing

Analysis

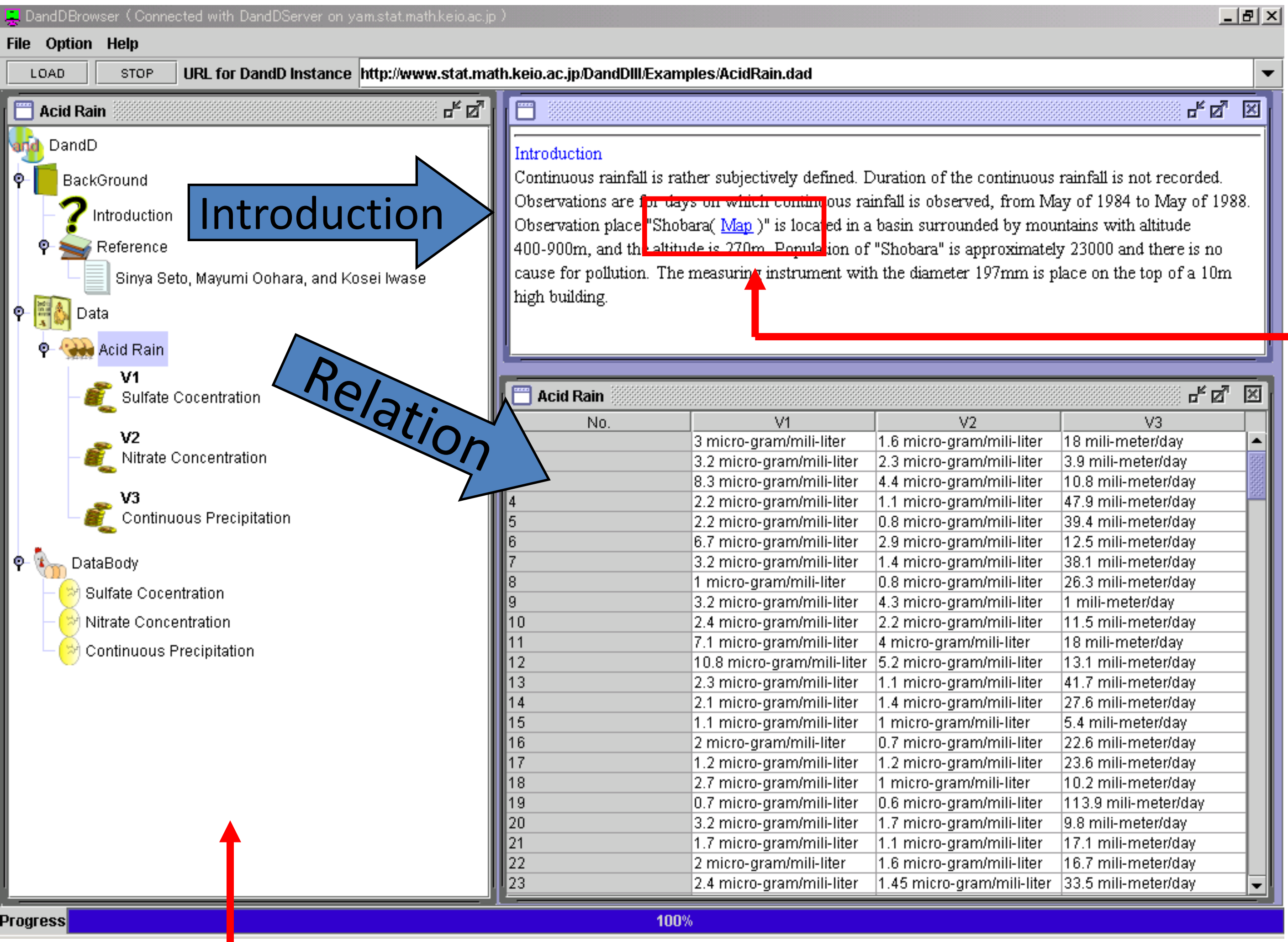
Modeling

DandD Server and client programs can be downloadable freely together with their source codes from

http://www.stat.math.keio.ac.jp/DandD/

DandD Browser

- Purpose
  - To help users to understand what kind of data are described in a DandD instance.
- Implementation
  - Java language



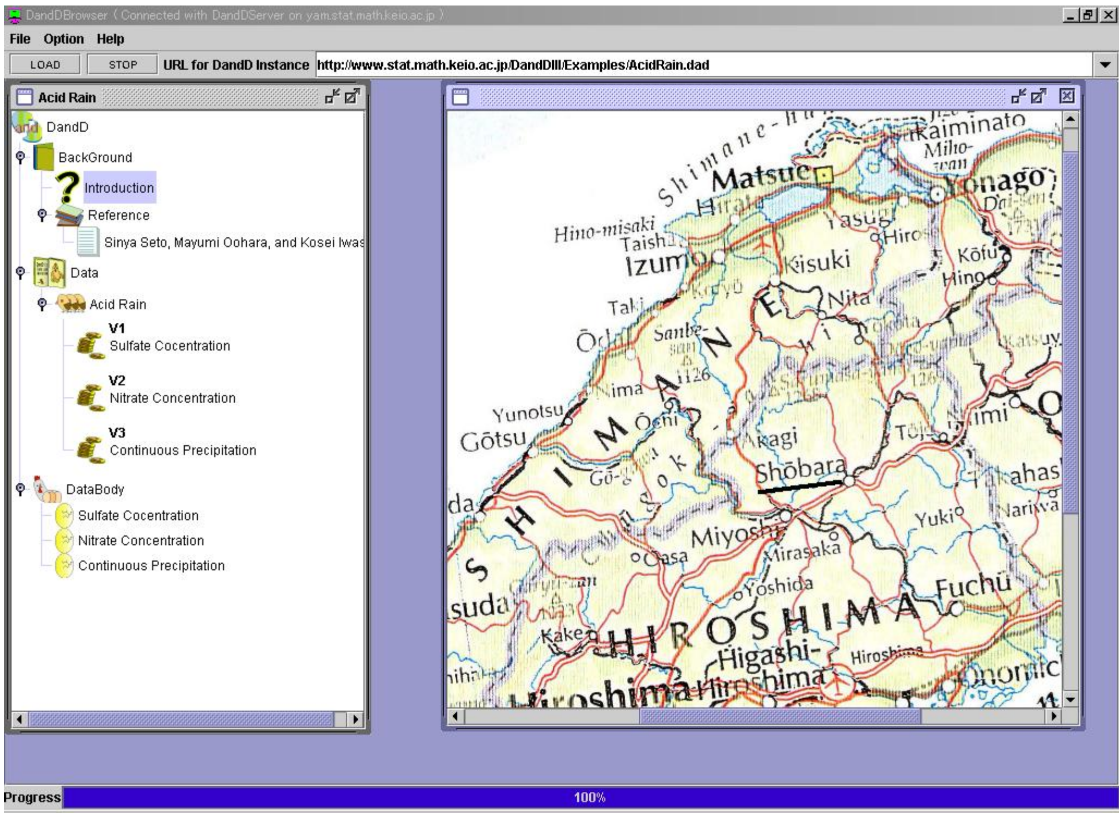
Introduction

Continuous rainfall is rather subjectively defined. Duration of the continuous rainfall is not recorded. Observations are for days on which continuous rainfall is observed, from May of 1984 to May of 1988. Observation place "Shobara(Map)" is located in a basin surrounded by mountains with altitude 400-900m, and the altitude is 270m. Population of "Shobara" is approximately 23000 and there is no cause for pollution. The measuring instrument with the diameter 197mm is place on the top of a 10m high building.

Acid Rain

No.	V1	V2	V3
	3 micro-gram/mili-liter	1.6 micro-gram/mili-liter	18 mili-meter/day
	3.2 micro-gram/mili-liter	2.3 micro-gram/mili-liter	3.9 mili-meter/day
	8.3 micro-gram/mili-liter	4.4 micro-gram/mili-liter	10.8 mili-meter/day
4	2.2 micro-gram/mili-liter	1.1 micro-gram/mili-liter	47.9 mili-meter/day
5	2.2 micro-gram/mili-liter	0.8 micro-gram/mili-liter	39.4 mili-meter/day
6	6.7 micro-gram/mili-liter	2.9 micro-gram/mili-liter	12.5 mili-meter/day
7	3.2 micro-gram/mili-liter	1.4 micro-gram/mili-liter	38.1 mili-meter/day
8	1 micro-gram/mili-liter	0.8 micro-gram/mili-liter	26.3 mili-meter/day
9	3.2 micro-gram/mili-liter	4.3 micro-gram/mili-liter	1 mili-meter/day
10	2.4 micro-gram/mili-liter	2.2 micro-gram/mili-liter	11.5 mili-meter/day
11	7.1 micro-gram/mili-liter	4 micro-gram/mili-liter	18 mili-meter/day
12	10.8 micro-gram/mili-liter	5.2 micro-gram/mili-liter	13.1 mili-meter/day
13	2.3 micro-gram/mili-liter	1.1 micro-gram/mili-liter	41.7 mili-meter/day
14	2.1 micro-gram/mili-liter	1.4 micro-gram/mili-liter	27.6 mili-meter/day
15	1.1 micro-gram/mili-liter	1 micro-gram/mili-liter	5.4 mili-meter/day
16	2 micro-gram/mili-liter	0.7 micro-gram/mili-liter	22.6 mili-meter/day
17	1.2 micro-gram/mili-liter	1.2 micro-gram/mili-liter	23.6 mili-meter/day
18	2.7 micro-gram/mili-liter	1 micro-gram/mili-liter	10.2 mili-meter/day
19	0.7 micro-gram/mili-liter	0.6 micro-gram/mili-liter	113.9 mili-meter/day
20	3.2 micro-gram/mili-liter	1.7 micro-gram/mili-liter	9.8 mili-meter/day
21	1.7 micro-gram/mili-liter	1.1 micro-gram/mili-liter	17.1 mili-meter/day
22	2 micro-gram/mili-liter	1.6 micro-gram/mili-liter	16.7 mili-meter/day
23	2.4 micro-gram/mili-liter	1.45 micro-gram/mili-liter	33.5 mili-meter/day

By double-clicking the hyper link, we can easily get an auxiliary material of introduction.

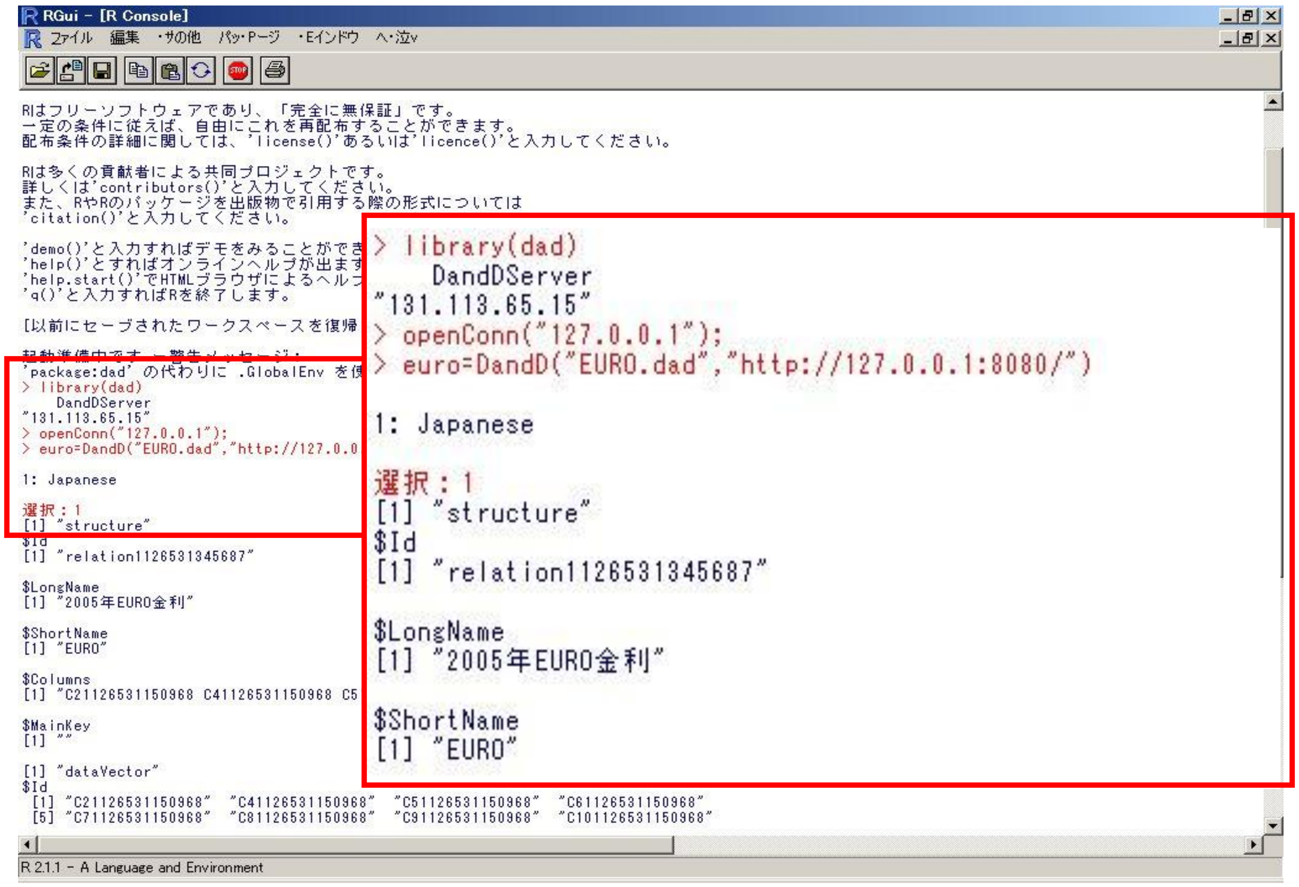


The tree in this window shows the structure of DandD instance.

By double-clicking or right-clicking each node, we can acquire data and its background information.

DandDR

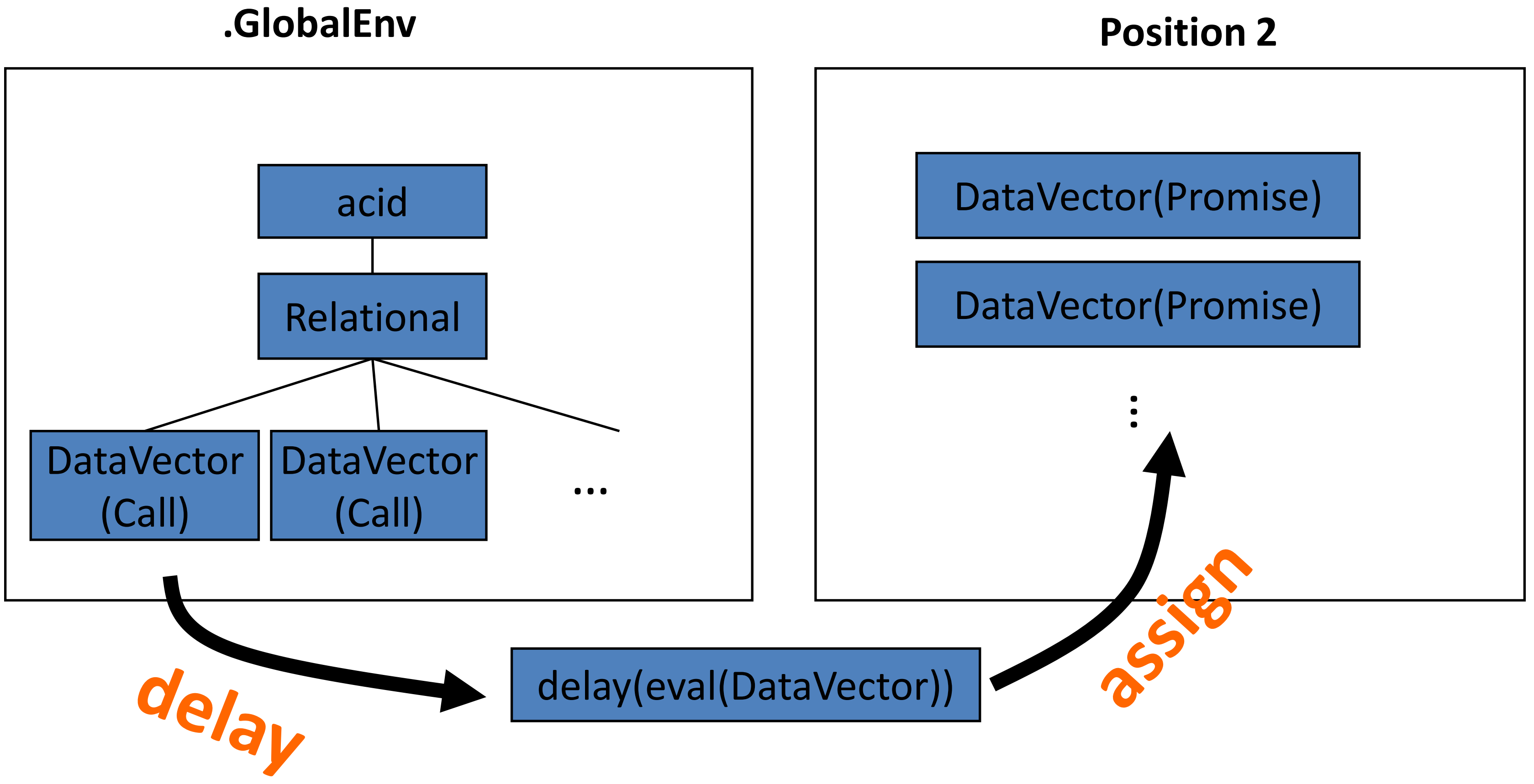
- Purpose
  - To support data analysis and modeling
  - An Interface between DandD Server and R
- Implementation
  - C language
  - Iconv (GNU)


RGui - [R Console]
ファイル 編集 ・サの他 パッ・P-ジ ・Eインドウ ヘ・泣v
R 2.1.1 - A Language and Environment

```
Rはフリーソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()'あるいは'licence()'と入力してください。

Rは多くの貢献者による共同プロジェクトです。
詳しくは'contributors()'と入力してください。
また、RやRのパッケージを出版物で引用する際の形式については
'citation()'と入力してください。

'demo()'と入力すればデモをみることができ
'help()'とすればオンラインヘルプが出ます
'help.start()'でHTMLブラウザによるヘルプ
'q()'と入力すればRを終了します。

[以前にセーブされたワークスペースを復帰

起動準備中です - 警告メッセージ:
'package:dad' の代わりに .GlobalEnv を使
> library(dad)
    DandDServer
"131.113.65.15"
> openConn("127.0.0.1");
> euro=DandD("EURO.dad","http://127.0.0

1: Japanese

選択：1
[1] "structure"
$Id
[1] "relation1126531345687"

$LongName
[1] "2005年EURO金利"

$ShortName
[1] "EURO"

$Columns
[1] "C21126531150968 C41126531150968 C5

$MainKey
[1] ""

[1] "dataVector"
$Id
 [1] "C21126531150968"  "C41126531150968"  "C51126531150968"  "C61126531150968"
 [5] "C71126531150968"  "C81126531150968"  "C91126531150968"  "C101126531150968"
```

```
> library(dad)
    DandDServer
"131.113.65.15"
> openConn("127.0.0.1");
> euro=DandD("EURO.dad","http://127.0.0.1:8080/")

1: Japanese

選択：1
[1] "structure"
$Id
[1] "relation1126531345687"

$LongName
[1] "2005年EURO金利"

$ShortName
[1] "EURO"
```

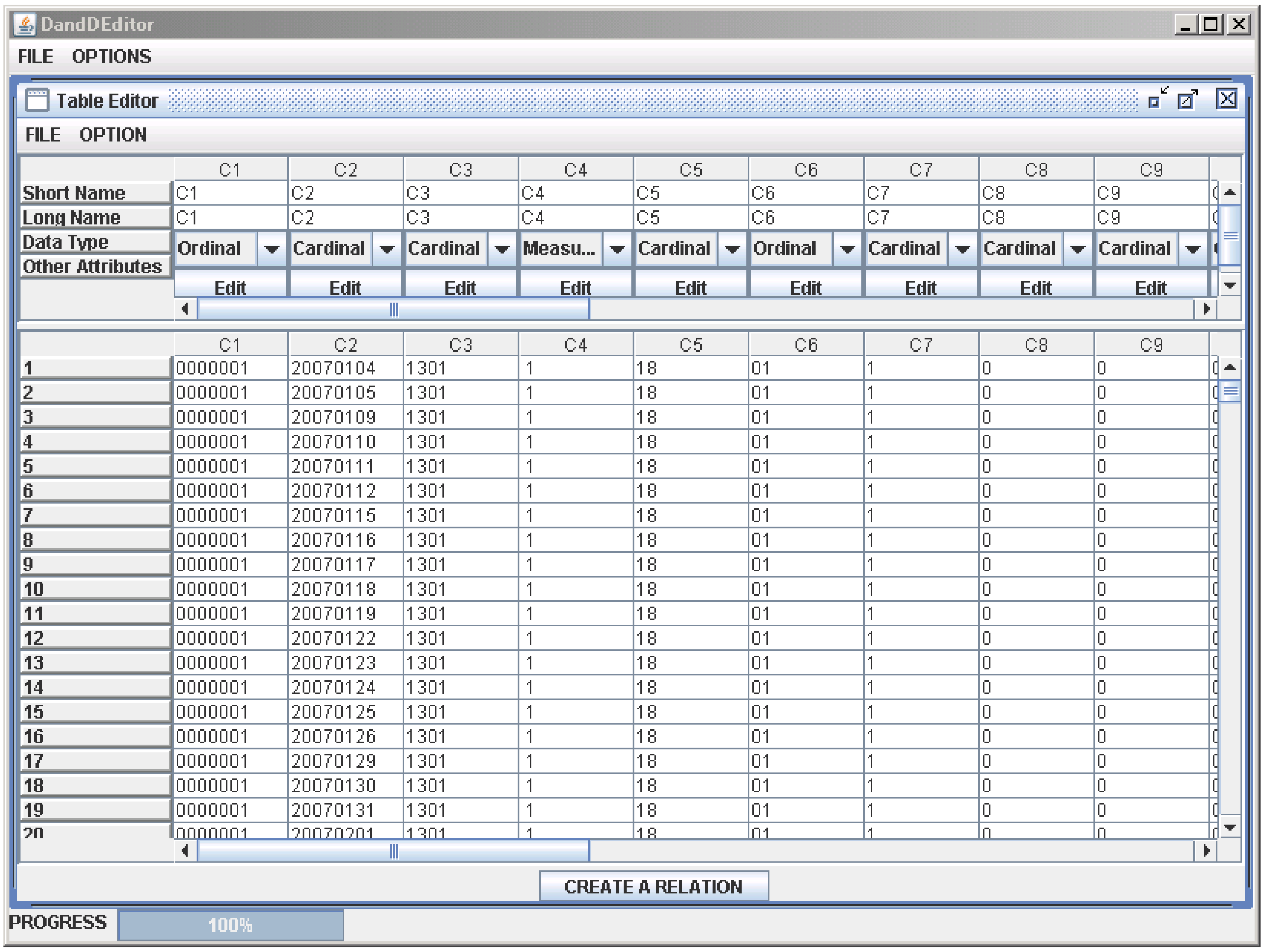
Mechanism of loading DandD instance



```
> library(dad)
> euro=DandD("EURO.dad")
> euro #`print.dad' is called
```

DandD Editor

- Purpose
  - Creating a DandD instance from data in text files
  - Editing an existing DandD instance
- Supported Format
  - Plain text file
    - comma separated, tab separated, space separated, fortran format
- Data Storage
  - Use of relational database management system
    - DandD Editor automatically stores data in relational database management system.
  - Advantage
    - Data stockpile of relational database management system is larger than that of statistical analysis software or spread sheet software.
- Implementation
  - Java language



	C1	C2	C3	C4	C5	C6	C7	C8	C9
Short Name	C1	C2	C3	C4	C5	C6	C7	C8	C9
Long Name	C1	C2	C3	C4	C5	C6	C7	C8	C9
Data Type	Ordinal	Cardinal	Cardinal	Measu...	Cardinal	Ordinal	Cardinal	Cardinal	Cardinal
Other Attributes	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit

	C1	C2	C3	C4	C5	C6	C7	C8	C9
1	0000001	20070104	1301	1	18	01	1	0	0
2	0000001	20070105	1301	1	18	01	1	0	0
3	0000001	20070109	1301	1	18	01	1	0	0
4	0000001	20070110	1301	1	18	01	1	0	0
5	0000001	20070111	1301	1	18	01	1	0	0
6	0000001	20070112	1301	1	18	01	1	0	0
7	0000001	20070115	1301	1	18	01	1	0	0
8	0000001	20070116	1301	1	18	01	1	0	0
9	0000001	20070117	1301	1	18	01	1	0	0
10	0000001	20070118	1301	1	18	01	1	0	0
11	0000001	20070119	1301	1	18	01	1	0	0
12	0000001	20070122	1301	1	18	01	1	0	0
13	0000001	20070123	1301	1	18	01	1	0	0
14	0000001	20070124	1301	1	18	01	1	0	0
15	0000001	20070125	1301	1	18	01	1	0	0
16	0000001	20070126	1301	1	18	01	1	0	0
17	0000001	20070129	1301	1	18	01	1	0	0
18	0000001	20070130	1301	1	18	01	1	0	0
19	0000001	20070131	1301	1	18	01	1	0	0
20	0000001	20070201	1301	1	18	01	1	0	0