

Procedures for Using and Maintaining the User Registration and File Download System

The Health and Retirement Study

Michael A. Nolte, MILS
Research Area Specialist Lead

July 2012

Table of Contents

- I. HRS File Distribution System: User Perspective 2
 - 1. Overview 2
 - 2. Privacy and Security: Additional Issues 2
 - 2a. IP Logging 2
 - 2b. Cookies 3
 - 2c. Registration Form 3
 - 2d. Security 3
 - 3. User Actions 3
 - 3a. Login Process 3
 - 3b. Product Selection 4
 - 3c. File Downloads 4
 - 4. Registration/Login to Download – Step by step 6
- II. HRS File Distribution System: Management Perspective 7
 - 1. Overview 7
 - 2. User Management 8
 - 3. Adding a Product and Files 9
 - 3a. New Products 9
 - 3b. Product Files 10
 - 4. Installing Product Files 10
- III. Special Procedures for Limited Access Files 13
 - 1. User Registration Procedures 13
 - 1a. Procedures for Sensitive Health Data Users 13
 - 1b. Procedures for Production Data Reviewers 13
 - 1c. Download Process for All Limited Access Files 13
 - 2. Administrative Procedures 13
 - 2b. Accgroup Table Modification 14
 - 2b. User permission change 15
- Appendix A: How the Server System Works 16
 - Starting Point: LAMP Software 16
 - Linux 16
 - Apache HTTP Server 16
 - MySQL 16
 - PHP: Hypertext Preprocessor 17
- Appendix B: MySQL Database Design Overview 18

I. HRS File Distribution System: User Perspective

1. Overview

The Health and Retirement Study File Distribution System is designed to provide researchers with access to all the HRS public limited access data products listed on the *What's Available* page. HRS data are provided at no cost; the only input required from users is identification information such as email address and institutional affiliation. Information on user identities gives HRS the ability to provide funding agencies with usage information; it also lets us keep in touch with users.

To access HRS public release data, users begin by filling out a registration form to obtain a username and password. This username/password combination must be used whenever the user wishes to download data files. Users can change their registration information at any time by logging in to the file distribution site. By registering, the user agrees to the *Conditions of Use*¹ governing access to HRS public data. Users have privacy rights to the information they provide: it will not be used for commercial purposes, and will not be redistributed to third parties. See the *Privacy Statement*² for more information.

[Data Products](#) » [Access to Public Data](#)

[Registered Users](#) [New Users](#)

Why You Must Register...

In order to access Health and Retirement Study public release data, you must first register and obtain a username and password. This username/password combination must be used whenever you wish to download data files. By registering at this site, you are agreeing to the [Conditions of Use](#) governing access to Health and Retirement public release data. You can change your registration information at any time by simply logging in to this site.

Through your registration, we are able to convey to our sponsors the size and diversity of our user community, allowing us to continue to collect these important data. Registered users will receive user support, as well as information related to errors in the data, future releases, workshops, and publication lists. The information you provide will *not* be used for commercial purposes, and will *not* be redistributed to third parties. For more information concerning privacy issues, please read our [Privacy and Security Statement](#)

2. Privacy and Security: Additional Issues

2a. IP Logging

The HRS provides the file distribution system as a public service. As part of the management process, certain user information (in addition to registration information) is automatically logged. Although the user's IP address is may be logged to help diagnose problems with the Apache Web server, MySQL database server, for statistical purposes and for other site administration purposes, we do not link IP addresses to anything personally identifiable. This means that a user's session will be tracked, but the user will be anonymous. When we do save personal information about a user (e.g., when registering for access to public data files) we do not save any information about which IP address they used. Except for authorized law enforcement investigations, no other attempts are made to identify individual users or their usage habits. Raw data logs are used for no other purposes and are scheduled for regular destruction.

¹ <http://hrsonline.isr.umich.edu/index.php?p=regcou>

² <http://hrsonline.isr.umich.edu/index.php?p=regpriv>

2b. Cookies

Our site uses cookies for authentication purposes during the login process and while during the file access process. The cookie contains a hash that uniquely identifies the user session. Thus the user must accept cookies from the HRS file download site and keep them for the duration of the session. The cookie is created by `\\linux-src\www\htdocs\ssl\maintain2\cookietokenauth2.php`

2c. Registration Form

Our site's registration form requires users to give us contact information (such as their name, e-mail address, and telephone number) and demographic information (such as their state/province and institutional affiliation). Users are also requested to provide information on research interests and environment as well as educational background). *Users must provide a valid email address; if they do not, then they will not receive username and password information and will not be able to login to the site.*

| | |
|--|--|
| HRS | HEALTH AND RETIREMENT STUDY at the University of Michigan |
| HRS Data Download System : User Registration | |
| ← [Log in] | |
| Note: Fields marked with an asterisk (*) must be entered. | |
| *Username: | <input type="text"/> (5 to 8 letters or numbers; first character must be a letter) Your username will be used for all current and future transactions. |
| *E-mail address: | <input type="text"/> A valid, current email address is required for password processing and account renewal. |
| *First Name: | <input type="text"/> |
| Middle: | <input type="text"/> |
| *Last Name: | <input type="text"/> |
| *Telephone: | <input type="text"/> |
| *Organization: | <input type="text"/> |

2d. Security

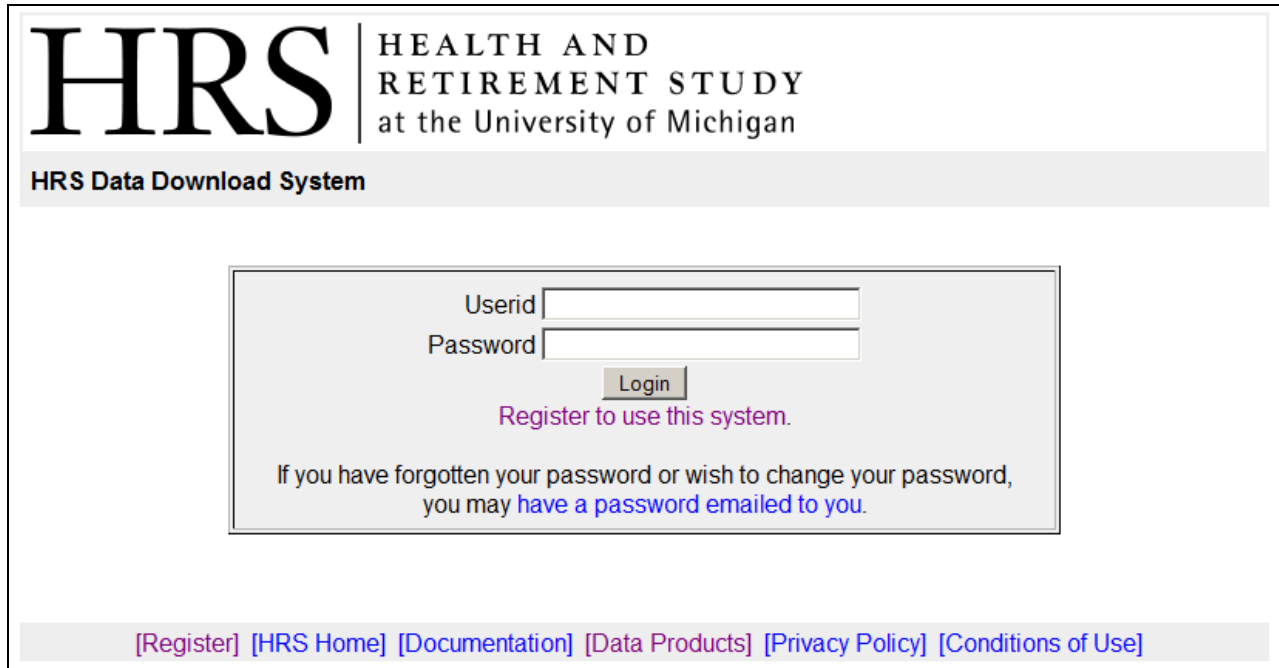
We rely on the security measures put in place by CMT to protect against the loss, misuse and alteration of the information under our control. However it is our responsibility to make sure that the security model embedded in the registration process is operating correctly, that file and directory permissions are properly set, that user behavior is monitored, and that the site in general operates correctly.

3. User Actions

3a. Login Process

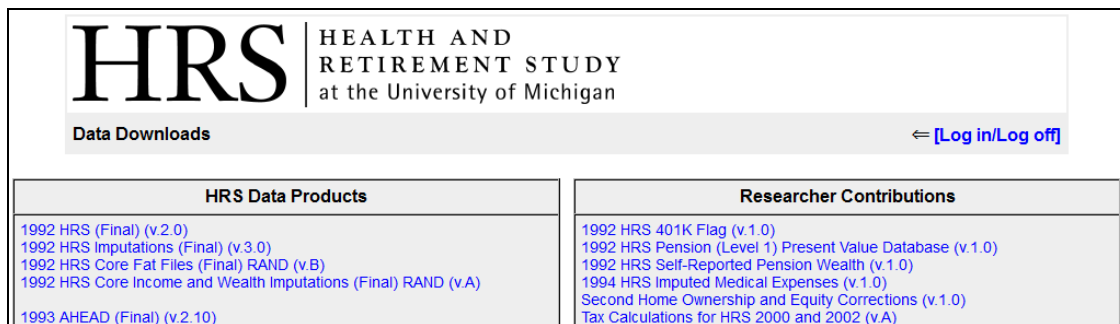
From the user's perspective, the first step is to visit the site and register. Once the registration process is complete, the new user will receive an email message containing a username (unique and chosen by the user) and password. The password contains random numbers and characters and generated by the registration program. Users are not allowed to create their own passwords. The password is not stored in

the MySQL database. Instead a SHA-256 hash of the password (plus a salt value) is stored in the **user** table. Unique salt values are created for each user record from an SHA-1 hash of a randomly determined value.



3b. Product Selection

Following a successful login, the user is presented with a welcome page that outlines the various products that can be downloaded from the site. Users then follow a link to the *Data Downloads* page.



3c. File Downloads

When the user selects a data product from the Datasets and Files page, s/he is presented with a list of files for that product. The file list page may also contain special information about the data product.

HRS HEALTH AND RETIREMENT STUDY at the University of Michigan

Researcher Contribution: 1992 HRS 401K Flag (v.1.0) ← [Data Downloads]

Note: This file is an HRS Researcher Contribution file. Researcher contribution files are provided by fellow researchers interested in sharing their work. Interested researchers are encouraged to contribute their own files by sending electronic mail to HRS User Support at hrsquest@isr.umich.edu. HRS did not produce this file and thus can not support it, nor be responsible for its content or use. It is provided only as a service to the research community.

Please view the File Merge Cross-Reference Table at the HRS Restricted Data Web site (<http://hrsonline.isr.umich.edu/rda/xref4.html>) for information on merging this file with restricted data files.

| | | | |
|------------|--|----------|------------|
| a_401k.zip | HRS Wave 1 Respondent 401K information | 31.81 KB | 07/15/1999 |
|------------|--|----------|------------|

Instructions:
Download: To download .zip or .exe files left-click on the file icon or the file name beside it and choose the "Save File" option. Alternate method (.zip, .exe, .pdf, .txt): Right-click on the file icon/name and then select "Save link as", "Save target as", "Download linked file"... depending on your browser and operating system.
View: To view a .pdf or .txt file, click on the file icon or the file beside it. Use the browser's back arrow icon (←) to return to the file download page.

[Log in/Log off] [Data Downloads] [Documentation] [Data Products]

The user selects one or more files from the download page (instructions provided); download should begin automatically.

HRS HEALTH AND RETIREMENT STUDY at the University of Michigan

Researcher Contribution: 1992 HRS 401K Flag (v.1.0) ← [Data Downloads]

Note: This file is an HRS Researcher Contribution file. Researcher contribution files are provided by fellow researchers interested in sharing their work. Interested researchers are encouraged to contribute their own files by sending electronic mail to HRS User Support at hrsquest@isr.umich.edu. HRS did not produce this file and thus can not support it, nor be responsible for its content or use. It is provided only as a service to the research community.

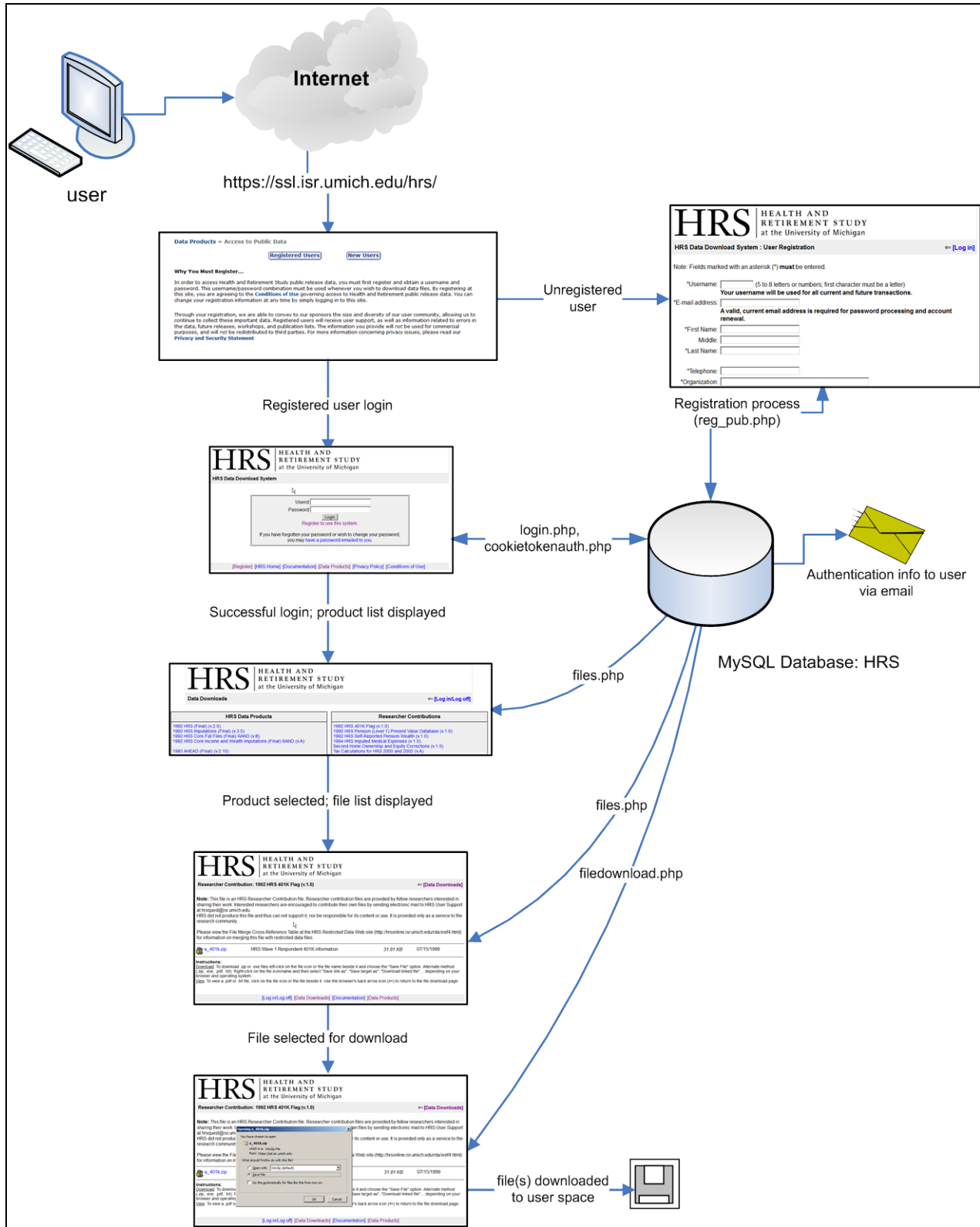
Please view the File Merge Cross-Reference Table at the HRS Restricted Data Web site (<http://hrsonline.isr.umich.edu/rda/xref4.html>) for information on merging this file with restricted data files.

| | | | |
|------------|--|----------|------------|
| a_401k.zip | HRS Wave 1 Respondent 401K information | 31.81 KB | 07/15/1999 |
|------------|--|----------|------------|

Instructions:
Download: To download .zip or .exe files left-click on the file icon or the file name beside it and choose the "Save File" option. Alternate method (.zip, .exe, .pdf, .txt): Right-click on the file icon/name and then select "Save link as", "Save target as", "Download linked file"... depending on your browser and operating system.
View: To view a .pdf or .txt file, click on the file icon or the file beside it. Use the browser's back arrow icon (←) to return to the file download page.

[Log in/Log off] [Data Downloads] [Documentation] [Data Products]

4. Registration/Login to Download – Step by step



II. HRS File Distribution System: Management Perspective

1. Overview

Distribution system administrators will probably want to use the Microsoft Access front end database (\\isr-hrs\hrs\UserReg\userreg_linux.mdb) to add new products and/or files, change product and/or file information, manage users, and generate reports. The tables in the MySQL database are linked to Access via ODBC. See Appendix B for an overview of table relationships.³

| Table Name | Purpose |
|--------------|---|
| accgroup | Define access groups |
| accgroupuser | Define which users can see which access groups |
| file | Information about each file: name, description, date, type flag; availability (accgroupid) |
| filetype | File types: name, icon, mime type |
| product | Name of the product; user notes; availability (accgroupid) |
| user | Information for each user: username, userid, password, email, research interests, institutional affiliation |
| userauth | Session information |
| userlog | Information on user activities |
| userold | Old version of user table records |
| version | Used in conjunction with the Product table |

³ An alternative method for maintaining the file download database is to use the MySQL Workbench, an open-source application provided by Oracle. The Workbench provides tools for SQL development, Data Modeling, and Server Administration. The SQL Development option allows the user to connect to existing databases and run SQL Queries and SQL Scripts, to edit data, and to manage database objects. The Server Administration option can be used to configure the database server, to setup user accounts, to browse status variables and server logs, and to import/export database content to/from the project folder used for backup. Note: Certain administrative functions that are theoretically available through the MySQL Workbench actually require special permissions from the MySQL system administrator.

2. User Management

As noted above, users need only provide a unique username and a valid email address to obtain HRS public data, although we ask them to provide much more information on their research interests and institutional affiliation. In general, users are responsible for updating their information. There may be times, however, when HRS staff may need to change information in the **user** table. This can be done directly or through the *user_maintenance* form. (See below for special issues related to limited access files.)

The screenshot shows a web browser window titled "user_maintenance". The form contains the following fields and values:

- userid: 2
- username: manolte
- email alerts?: 1
- email events?: 1
- fname: Michael
- middle: A
- lname: Nolte
- Org Name: ISR-HRS
- Organization Type: 1
- Address 1: 426 Thompson St
- Primary role: 3
- Address 2: Room 3100
- Highest degree: 3
- City/State: Ann Arbor, MI
- HRS Pubs?: 1
- Postal Code: 48104
- Country: USA
- email: manolte@isr.umich.edu
- telephone: 734.936.1903

Research Interests (checkboxes):

- Anthropology:
- Epidemiology:
- Health Services:
- Public Policy:
- Demography:
- Gerontology:
- Political Science:
- Sociology:
- Economics:
- Health Sciences:
- Psychology:
- Other: Information Systems

Co-workers?: 2 Janet Keller

| accgroupuserid | accgroupid | userid |
|----------------|------------|--------|
| 2 | 1 | 2 |
| 7318 | 4 | 2 |
| 1373 | 2 | 2 |

Record: 1 of 14

Record: 2 of 8762

3. Adding a Product and Files

3a. New Products

Information for new products should be added to the **product** and **version** tables via the *product* form:

product

productid: 125 productname: HRS 2008 Production Data Modules Delete Product

seqnum: 9036 accgroupid: 15

productnotes: <h4>Conditions of Use</h4><p>Pre-release production data sets are made available to HRS investigators, project staff members, and associates for quality control and external review purposes. Although great care is taken to remove identifying information, these early releases must be treated as restricted data. They are provided solely for internal development and improvement of the HRS instrument.</p>Publication and/or presentation of analyses based on pre-release production data is not allowed. Pre-release production data must not be merged with any other restricted data sets. Merging with prior waves is permissible where necessary for evaluating the data in the current wave.</p><p>Upon completion of quality control work or upon request, user must agree to destroy the original data set and all derived data sets. User must also agree to make no attempt to identify members of the data set and not to distribute or make the data accessible to third parties.</p><p>User also agrees to provide additional protection for these data. At a minimum this means:</p>

product table input

product-version

| versionid | productid | versionname | versiondate |
|-----------|-----------|-------------|-------------|
| 125 | 125 | 4.0 | 05/08/2008 |
| * | 125 | | |

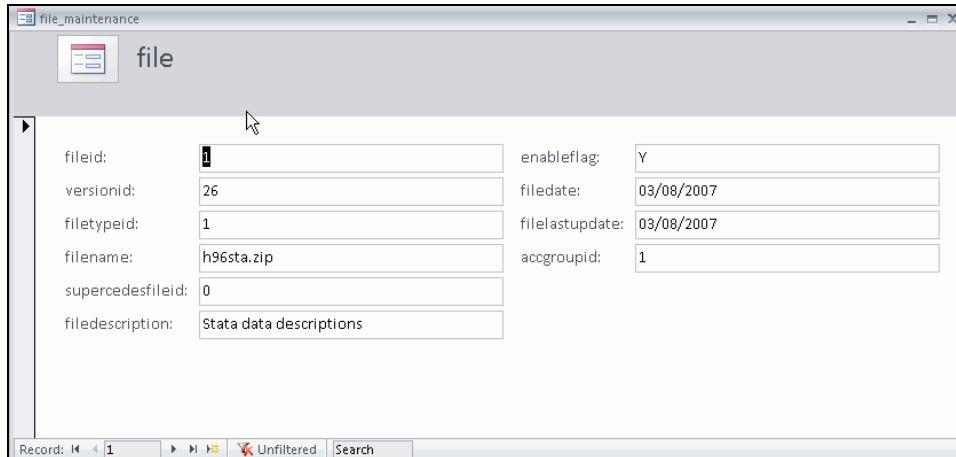
version table input

Save Record

Record: 88 of 88 Unfiltered Search

3b. Product Files

Information in the `file` table can be added directly to the table or via the `file_maintenance` form:



4. Installing Product Files

Once information about a product and its files has been added to the MySQL database, the next step is to add the files that make up the data set to the distribution file system. There are two methods for file transfer:

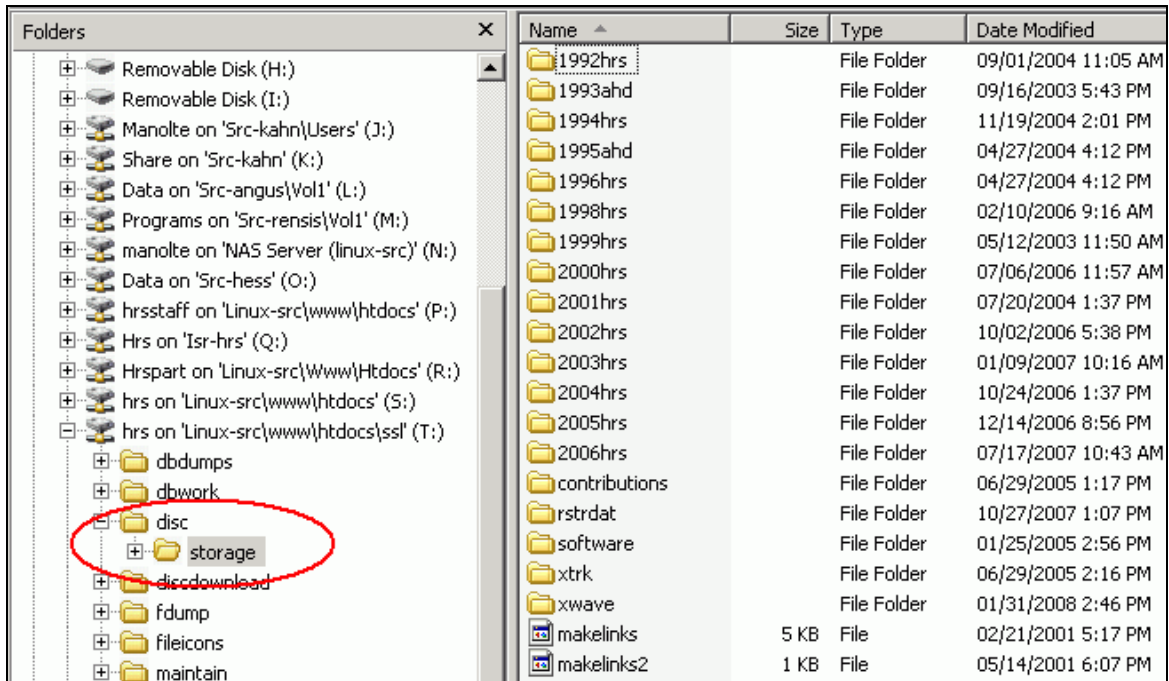
- Map the distribution system `disc` folder to a local letter on your workstation
- Connect to `linux-src` using your SSH client

The following set of instructions assumes an SSH connection;

1. Log in and then move to the File Distribution System Storage area

```
manolte@galatia:~> cd hrsSSL/
manolte@galatia:~/hrsSSL> ls
dbdumps      filedownload.php      hrslogo1.gif          reg_cou.html
dbwork       fileicons             index.html            reg_infocollect.html
disc         files_20070526._php  index_tmp._html      reg_priv.html
discdownload files_20080314._php  login.php             reg_pub.php
errfunc.php  files.php             maintain              reg_update.php
favicon.ico  forgotten.php         old                   source
fdump       hrs001.css           pics                  test_php
manolte@galatia:~/hrsSSL>
```

- Copy the file(s) to the appropriate area under `./hrsSSL/disc/storage`



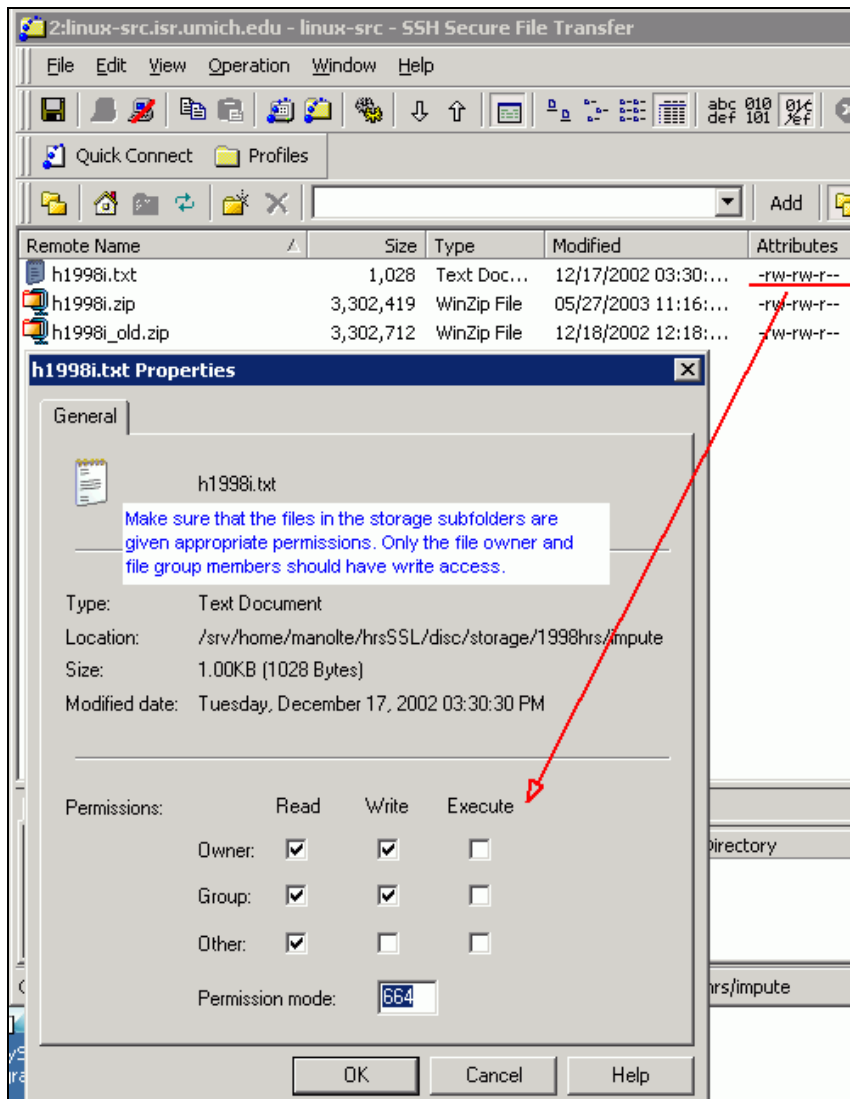
- The next step is to associate a file with its counterpart row in the file table. To establish the association between file and database, create a symbolic link.

```
manolte@galatia:~/hrsSSL> cd disc
manolte@galatia:~/hrsSSL/disc> ls -la | more
```

| Permissions | Link Count | Owner | Group | Size | Timestamp | Name | Description |
|-------------|------------|---------|-------|------|------------------|---------|--|
| drwxrwxr-x | 3 | manolte | hrs | 8192 | 2008-04-04 18:43 | . | Symbolic links in folder <code>disc</code> map the value of <code>fileid</code> in table <code>file</code> to the actual file located in the sub-folder <code>storage</code> |
| drwxrwx--- | 13 | root | hrs | 4096 | 2008-04-07 15:48 | . | |
| lrwxrwxrwx | 1 | manolte | 23 | 35 | 2006-03-15 13:02 | 100.txt | Example: the <code>fileid</code> value for the row in table <code>file</code> that contains information on <code>h1998i.txt</code> equals 100 |
| lrwxrwxrwx | 1 | manolte | 23 | 35 | 2006-03-15 13:02 | 101.zip | |
| lrwxrwxrwx | 1 | manolte | 23 | 27 | 2006-03-15 13:02 | 102.txt | |
| lrwxrwxrwx | 1 | manolte | 23 | 27 | 2006-03-15 13:02 | 103.zip | |

```
manolte@galatia:~/hrsSSL> cd disc
manolte@galatia:~/hrsSSL/disc> ln -s ./storage/1998hrs/impute/h1998i.txt 100.txt
```

- 4. As a final step, make sure that the new files have the correct permissions:



III. Special Procedures for Limited Access Files

The HRS File Distribution System is also used to distribute limited access files such as production review data and sensitive health data. Using the site for these purposes requires that both administrators and users follow special procedures.

1. User Registration Procedures

1a. Procedures for Sensitive Health Data Users

Users must register on the site and obtain a username and password. Applicants for sensitive health data products should:

- Identify the desired data set(s) under the Requesting HRS Special Access File heading.
- Download and complete the appropriate Data Use Agreement for each product selected (hardcopy format)
- Send a signed copy of the agreement via surface mail or fax to the HRS DUA Review Committee

An HRS representative (Mary Cheatham) reviews sensitive health requests, and based on the researcher's registration information verifies her/his identity and institutional affiliation. Once this authentication process has been completed an HRS system administrator (Janet Keller) will authorize access to the desired data set(s), communicating with the researcher at the email address in the user registration database via the HRS Data Requests mailbox. At this point the user can download the encrypted file.

1b. Procedures for Production Data Reviewers

Prior to each new wave of interviews, the HRS management team obtains a commitment from interested users (co-investigators and others) to review production data as it is received from the field. They are emailed a DUA which they return via email or fax, having signed both the ISR Policy on Safeguarding Respondent Confidentiality and the HRS Pre-release Production Data Conditions of Use agreement. Janet Keller creates a group in the file distribution system for the production data product. Users be registered on the site (e.g., have a username and password) and have returned their DUA before they can be added to the `accgroupuser` table (see below).

1c. Download Process for All Limited Access Files

- The user receives notification of access approval
- The user logs in at the HRS File Distribution System Web site.
- In the HRS Special Access Files box on the right hand side of the page the user should now see links to the data set(s) that were requested and approved

The approved encrypted files can now be downloaded in the standard fashion.

2. Administrative Procedures

Information for the limited access data process should be entered into the product, version, and file tables as described above for public data products. In the case of limited access data, two additional administrative steps are necessary: First, the administrator should add a record to the `accgroup` table

for the limited access data product in question.⁴ Then add a record to the **accgroupuser** table linking the researcher's userid to the limited access data product. This will make the product visible to the applicant.

2b. accgroup Table Modification

| accgroupid | accgroupname | permcode |
|------------|--------------|-------------|
| 1 | PUBLIC | PUB |
| 2 | RESTRICTED | REST |
| 3 | PRODUCTION | PROD |
| 4 | ADAMS | ADAM |
| 5 | DIABETES | DIAB |
| 6 | PRODUCTION2 | PROD2 |
| 7 | PDS | PDS |
| 8 | IMPUTATION | IMP |
| 9 | COGNITION | COG |
| 10 | PRODUCTION3 | PRODINT |
| 11 | PRODUCTION4 | PRODINT2 |
| 12 | BIO2006 | BIO2006 |
| 13 | PROD2008PT | PROD2008PT |
| 14 | PROD2008ALL | PROD2008ALL |
| 15 | PROD2008MOD | PROD2008MOD |
| * | | |

⁴ Be sure to check the effect of these changes on the operation of files.php

2b. User permission change

Give the user permission to see/download products tagged with a limited access identifier.

The screenshot shows a web application window titled "user_maintenance". It contains a form for user details and a table of access groups. The user details form includes fields for user ID, username, email alerts, email events, first name, middle name, last name, organization name, address, city/state, postal code, country, email, and telephone. There are also checkboxes for various fields of study and a section for co-workers. The table below the form has columns for accgroupuserid, accgroupid, and userid. The value "14" in the accgroupid column is circled in red. A red arrow points from a text box to this value. The text box contains the instruction: "Add accgroup value corresponding to the limited access dataset in question".

| accgroupuserid | accgroupid | userid |
|----------------|------------|--------|
| 10101 | 15 | 2 |
| 10106 | 14 | 2 |
| * | | 2 |

Appendix A: How the Server System Works

Starting Point: LAMP Software

The acronym LAMP refers to a solution stack of software, usually free and open source software, used to run dynamic Web sites or servers. The original expansion is as follows:

- Linux, referring to the operating system;
- Apache, the Web server;
- MySQL, the database management system (or database server);
- Perl, Python, and PHP, the programming languages.

The combination of these technologies is used primarily to define a web server infrastructure, define a programming paradigm of developing software, and establish a software distribution package.

Though the originators of these open source programs did not design them all to work specifically with each other, the combination has become popular because of its low acquisition cost and because of the ubiquity of its components (which come bundled with most current Linux distributions). When used in combination they represent a solution stack of technologies that support application servers. Source: Wikipedia, the free encyclopedia

[http://en.wikipedia.org/wiki/LAMP_\(software_bundle\)](http://en.wikipedia.org/wiki/LAMP_(software_bundle))

Linux

Linux is a Unix-like computer operating system kernel. For more information, start with the Wikipedia entry: <http://en.wikipedia.org/wiki/Linux>

Apache HTTP Server

Apache HTTP Server is a free software/open source web server, primarily used to serve both static content and dynamic Web pages on the World Wide Web. Many web applications are designed expecting the environment and features that Apache provides. Apache is the web server component of the popular LAMP web server application stack, alongside MySQL, and the PHP/Perl/Python programming languages. Go to Wikipedia, the free encyclopedia, for more information:

http://en.wikipedia.org/wiki/Apache_HTTP_Server

MySQL

MySQL is a multithreaded, multi-user, SQL Database Management System (DBMS) owned by Sun Microsystems. Libraries for accessing MySQL databases are available in all major programming languages with language-specific APIs. In addition, an ODBC interface called MyODBC allows additional programming languages that support the ODBC interface to communicate with a MySQL database, such as ASP or ColdFusion. The MySQL server and official libraries are mostly implemented in ANSI C/ANSI C++. MySQL is popular for web applications and acts as the database component of the LAMP platform (Linux-Apache-MySQL-PHP/Perl/Python) Go to Wikipedia, the free encyclopedia, for more information: <http://en.wikipedia.org/wiki/MySQL>

PHP: Hypertext Preprocessor

PHP is a reflective programming language designed for producing dynamic Web pages. PHP is used mainly in server-side application software (it generally runs on a web server), but can be used from a command line interface or in standalone graphical applications. PHP is especially suited for web development and can be embedded into HTML. Sometimes Perl and Python are used in place of PHP. Go to Wikipedia, the free encyclopedia, for more information: <http://en.wikipedia.org/wiki/PHP>

