

A manual to RePEcPHP*

Ben Greiner[†]

Version 1.0: January 2, 2004

Abstract

RePEcPHP is a free PHP and MySQL based web interface to maintain a RePEc archive. The RePEc project provides a volunteer-driven public-access database of more than 100.000 working papers plus other items. With RePEcPHP providers of working paper series are able to easily do their work via a comfortable web interface. A customizable series homepage is automatically created, including a search form. RePEcPHP can create the documents in ReDIF format parsed by the RePEc robot either on the fly or via an one-button click. All data fields used in the standard ReDIF templates for working paper series are supported. An import function is provided to read in existing archives.

Keywords: RePEc, PHP, MySQL, Working Paper Series

JEL Classification: Z00

*I thank the RePEc team for providing their services to the scientific community, Carsten Schmidt for indicating the need of such a software, and David K. Levine for the wjecon.php3 script, which I used before developing this software.

[†]University of Cologne, Faculty of Economics and Social Sciences, Albertus-Magnus-Platz, D-50932 Köln, Germany. Tel.: +49/221/470-6116, Fax: +49/221/470-5068, e-mail: bgreiner@uni-koeln.de

1 Introduction

RePEc (Research Papers in Economics)¹ is a volunteer-driven project to provide economic research material to the scientific community. This includes working papers, journal articles, software components, books and chapters, and author and institution information. By now, the database holds more than 200.000 items, of which more than 100.000 are available online. The largest part of the database consists of working paper information.

Technically, RePEc works as follows: An institution, let's say an university, decides to include it's working paper series on economics in RePEc's database. It contacts the RePEc team, which returns a so called 'Handle' for the new archive to create. On a publicly and anonymously accessible ftp- or http-(web)-server, the institution builds an archive consisting of data files in the ReDIF² format. These files describe the archive, the series contained in the archive, and the papers published in these series, including author and download information.

Every night, the robot program of RePEc comes along, downloads all files in the archive, parses the information and includes it in the RePEc database. From now on, the working papers are listed at services of the RePEc family, as IDEAS³, EconPapers⁴, WoPEc⁵ and so on.⁶

Institutions can register at EDIRC⁷, and an author may register at the RePEc author service⁸, note her affiliation with an institution and associate her papers listed in RePEc with her profile. Detailed access statistics for papers, series and authors are provided by LogEc⁹.

However, providing an RePEc archive can be sometimes laborious. One has to learn about the ReDIF format, to edit the ReDIF files either directly on the server or to upload them via ftp, to keep an eye on the format, to care about file and entry order, and last but not least to maintain the data twice: once in the RePEc archive and once at the own working paper series homepage. This is where RePEcPHP comes into play, the software this manual is about.

With RePEcPHP you are able to maintain your RePEc archive via a comfortable web interface. The software should run on every apache+php+mysql

¹<http://www.repec.org>

²The ReDIF format is described under ftp://netec.mcc.ac.uk/pub/NetEc/RePEc/all/root/docu/redif_1.html

³<http://ideas.repec.org/>

⁴<http://econpapers.hhs.se/>

⁵<http://netec.mcc.ac.uk/WoPEc/>

⁶For more information on the services provided by RePEc see their homepage.

⁷<http://edirc.repec.org/>

⁸<http://authors.repec.org/>

⁹<http://logec.hhs.se/>

server. Installation instructions are provided in Section 2. You may configure and customize the look of your public accessible working paper series homepage and the administration area, as described in Section 2.2. If you are already running a RePEc archive, you may import it into RePEcPHP using the import function (see Section 2.4). RePEcPHP may create the ReDIF files either on the fly or via a one-button-click. You can find a guide to these options in Section 2.5.

The administration of the RePEc archive using RepEcPHP is shortly described in the second part of the paper, in Section 3. An online help provides useful hints directly in the web interface.

RePEcPHP is distributed under the GNU General Public Licence, i.e. you may use it for free.

2 Installation

To install and use RePEcPHP, you need:

- A webserver like Apache.
- PHP running under this webserver.
- The MySQL database.

In the following, we will assume that your webserver's document path is located in `/var/www/html`, that you want to provide the working series homepage under `http://yourwebserv.er/wpseries/`, and that the database you use for RePEcPHP is called `wparchive`.

2.1 Software Installation

The installation of RePEcPHP is straightforward. First copy your source file to your webserver path, then unpack and rename it.

```
cp repecphp.tgz /var/www/html/  
cd /var/www/html  
tar xvfz repecphp.tgz  
mv repecphp wpseries  
rm repecphp.tgz  
cd wpseries/
```

Now you have the software installed in `wpseries`. In the following we install the database. You may create a new database or use an existing one. Note,

that all tables created for RePEcPHP are starting with `wp_`, and that only these tables are used.

If you want to install a new database, type

```
mysql -u mysqlusername -p
mysql> create database wparchive;
mysql> quit
```

To import the default tables and data into your database, do

```
mysql -u mysqlusername -p < install/repecphp.sql
```

That's it with the installation. Now proceed to configuration.

2.2 Main Configuration

The main configuration file is located in `funcs/conf.php`. Before starting with RePEcPHP, you have to edit this file. Variables in this file are set in PHP style in the form

```
$variable_name=variable_value;
```

The most important configurations are the database settings. State here your database name, the username and the password. For example:

```
$db_host = "localhost";
$db_name = "wparchive";
$db_user = "mysqlusername"
$db_pass = "mysqlpassword";
```

The next three options can be set to the values `true` and `false`. `$show_help` indicates, whether the online help for the RePEc data fields in the administration area should be shown or not. At the beginning, you should enable this feature, and after you are familiar with RePEc and this software, you may switch it off.

If `$enable_import` is set to `true`, a menu point for importing an existing archive is shown in the administration area (see Section 2.4). If `$enable_export` is set to `true`, you will find a menu option to generate the ReDIF (`.rdf`) files by a one-button-click. The state of the variable `$export_create_index` indicates, whether index files should be created in each folder when creating the archive. For both variables, see Section 2.5 for details.

The next variables describe the pattern how the URLs at IDEAS, EconPapers and LogEc are organized. You only have to change this if one of these

services changes their URL organization, but this seems not to be due in the long run.

The last PHP function call sets the error reporting level for RePEcPHP overriding the servers default configuration. This is done because we want to know if some files cannot be found or are not readable. If everything is configured smoothly, no errors should be shown.

After these settings, you may sign on in the administration area. Browse to the URL

```
http://yourwebserv.er/wpseries/admin/
```

and log in with the username `repecadmin` and the password `repecphp`. The first thing you should do is to create a new user (with superuser rights) and a good password, and then to delete the default user `repecadmin`. The procedure is described in Section 3.1.

2.3 Customizing

This section elaborates on the ways you can change the look of the pages served by RePEcPHP, especially your working paper series homepage. All files regarding this topic are located in the folder `style/`. The URL of the working paper series homepage is `http://yourwebserv.er/wpseries/`.

The files `series_header.php`, `series_footer.php` and `series_style.css` describe the design of your series homepage.

Although ending with `.php`, `series_header.php` and `series_footer.php` are simple HTML files. The series homepage delivered by RePEcPHP consists essentially of three parts in this order:

- `series_header.php` including the starting `<HTML>` tag, the `<HEAD>` definitions, the starting `<BODY>` tag, and everything that the series page should show before the list of papers. For example, you may add here the head graphics and the menu as in your other pages.
- the working paper list generated by RePEcPHP and formatted as described below, eventually including a search form and year links.
- `series_footer.php` including everything what should appear after the working paper list and finally the ending `</BODY>` and `</HTML>` tag. You may include footer links or a disclaimer etc.

The file `series_style.css` is simply the style sheet referred to in the `<HEAD>` section of the file `series_header.php`. In the stylesheet there should be defined

a class called `.paper`, which is automatically applied to the working paper list, the query box and the links generated by RePEcPHP.

The default files delivered with the source code build a very simple page header with a simple menu. As an example we deliver the files we use for our working paper series at the University of Cologne as well. They simply end with `.cln` instead of `.php`. If you overwrite the default files with these files, you see the new design.

The files `admin_header.php`, `admin_footer.php` and `admin_style.css` are the respective style files for the administration area. They can be treated as the series files. Note, that the menu generated by RePEcPHP is included in the `admin_header.php` file with the command

```
<?php include ("menu.php"); ?>
```

Regarding the appearance of the working paper list, two kinds of configuration have to be done: the paper template and the settings in the administration area.

The paper template is as well located in the `style` folder, namely in the file `series_paper_template.php`. There, the PHP variable `$paper_template` is set, describing the HTML code for each paper in the list. The variables (shortcuts) enclosed by `"#"`, which are listed in the header of the file, will be replaced by the according values for the paper.

RePEcPHP goes through the template line by line. If there is no variable found in the line, the full line will simply be printed out. If there is a known variable, it will be checked if this variable is set for the actual paper. If the variable is non-empty, the variable name will be replaced by the variable value. If it is empty, the full line will be excluded.

To see how this works, here are two examples from the default file:

```
<strong><font color="black">No. #Number#</font></strong>
```

If a number for the paper is given, the full line will be printed. If no number is found, not only the number but also the `` and `` directives will not be included in the HTML output.

```
(<a class="paper" href="#Download_Link#">#Download_Text#</a>)
```

If both the download link and the download text are present, the line will be shown. Otherwise, the whole line will be suppressed. (However, note that in this example the download text is only non-empty if the download link is non-empty.)

To output all HTML stuff independent of the fact if the variables are present or not, simply write all variables in separate lines.

The other settings are made in the administration area at the series creating/editing page. The following options can be set:

- **Series homepage linking:** You can decide if you want to link the downloadable working papers directly, or if the download link should point to the download page at EconPapers or IDEAS. The advantage of the latter is that LogEc counts the accesses to the papers for you, but with the disadvantage of another page delivered before the download starts.
- **Sort papers by:** You may sort the working papers by Handle, Number, Creation-Date or Revision-Date, all ascending or descending.
- **Series HP offset:** You can determine the maximum number of working papers you want to list per page. If there are more working papers found for the current query, the papers are separated on the necessary number of pages, and links to these pages are presented at the top and the bottom of the paper list.
- **Show creation year links:** If this feature is enabled, RePEcPHP will lookup if the papers found for the current query stem from different years as given in the Creation-Date. If this is true, at the top of the page some links to the years will be presented. After a click on a year the current query will be restricted to that year.
- **Show query box:** If this feature is enabled, the visitor of your series homepage is presented a small query form. Search is done through Title, Authors, Abstract, Keywords and JEL-Classifications. The default query (the query used when you enter an empty search term or access the series homepage URL directly) delivers all working papers.

When the working paper series homepage is accessed, and only one working paper series is defined in RePEcPHP, the papers of this series are presented to the visitor. If there are more than one series defined, first a series selection page is presented, where the existing working paper series are listed, each linked to the respective working paper list. You may copy these links to link the different working paper series directly from your other websites.

Having customized the look of your pages, you are ready to start with data input. If you are already running an archive, proceed with the next section.

If you are just starting an archive, consider Section 3 on how to fill in your data. Anyway you should read Section 2.5 on how to produce ReDIF files with RePEcPHP.

2.4 Archive Import

The use of the ReDIF archive import function is quite simple. However, the preconditions for importing your archive are:

- The variable `$enable_import` in the configuration file `conf/conf.php` is set to `true`.
- The RePEc archive tree is located in the same file system as your web server running RePEcPHP. If this is not the case, copy the tree from it's origin to your file system.
- The RePEc archive is readable for the user the web server (and thereby PHP and RePEcPHP) is running under.

Browse to the administration area, log in and choose "import" from the menu. In the first field fill in the fully qualified path name to your RePEc archive, i.e. it should start with "/" for root and end with the archive handle RePEc has assigned to you. In the second field provide again the archive handle. After a click on the button, RePEcPHP will search through your archive and parse the archive, series and paper ReDIF files. It will include the data in the database. Note, that existing archive information in RePEcPHP will be overwritten, while series and papers will be created as new items.

The parser is written to understand files in the ReDIF format. The format is described in the ReDIF documentation noted in the introduction. RePEcPHP will only import data which it supports. This includes all fields used in the "Archives", "Series of Papers" and "Papers" templates provided at the RePEc homepage.

Essentially, the parser reads in a template file line by line. Recognition pattern are case-insignificant. If it finds a line containing the phrase "Template-Type", it assumes that a new item starts from here.

If a line starts by a none-space character, and contains a colon (:), it is assumed that the line part before the colon is a field name, and the part after the colon is (the beginning of) the value of this field. If a line is non-empty, but doesn't fit this pattern, it is assumed that the full line is continuing the value of the field found just before. An exception is made for the paper field "Abstract", where an empty line indicates a paragraph break.

If your existing archive passes the error checker provided by RePEc, there should be no problems to import the files.

One further exception has to be noted: the way the parser treats the fields "Creation-Date" and "Revision-Date". The ReDIF-standard requires the pattern "yyyy[-mm[-dd]]". That is, the date has to either 2004-01-01, or 2004-01, or 2004. However, RePEcPHP uses fully qualified dates with day, month and year. The parser looks for the date, and if it does not find a day, it assumes the 15th (15), and if it does not find a month, it assumes June (06). It takes dates like 2004-1-1 as well. If the date does not comply to this standard, it makes another attempt: if it finds a string with a space and four numbers at the end, it assumes that these numbers represent the year (as in January 1, 2004). If none of these pattern yields a result, the date field is left empty.

For programmers: The parsing of files is done by the function `parse_archive()`. The return value of this function is described in Appendix B.

After importing your archive, you should go over the RePEc specific options at the archive and series edit pages in the administration area.

2.5 ReDIF Export

The aim of RePEcPHP is to comfortably maintain your working paper series and to easily generate the ReDIF files to be parsed by the RePEc robot. There are two ways to create the ReDIF files: on the fly or by a button click, both with specific preconditions. We will discuss both in this section.

2.5.1 By button click

To enable this option, the variable `$enable_import` in the configuration file `conf/conf.php` should be set to `true`. Then in the administration menu the option "files" appears.

To configure this generation method, first create an archive path. If you provide the RePEc archive via a ftp server, the path should be located in your anonymously accessible ftp tree. If you use a (http) webserver for the archive, the path should indeed be located in your webserver's document tree.

The name of your archive folder has to be identically to your archive handle the RePEc team provides to you. Two examples for the archive path on your computer (with an archive handle "xxx") are:

```
/var/www/html/RePEc/xxx
```

```
/var/ftp/pub/RePEc/xxx
```

Make this path writable to the user the webserver (and PHP and RePEcPHP) is running under. This enables RePEcPHP to write the files into this folder. If you don't have root access on the webserver, create the folder and change it's user permission rights to world-writable:

```
cd /var/www/html/RePEc
chmod 777 xxx
```

If you have root access, switch to the root user, and set the archives owner to the webserver user:

```
cd /var/www/html/RePEc
su root
chown webserveruser:webserverusergroup xxx
exit
```

Next, browse to the archive edit page. In the "Archive-Path" field put in the fully qualified path to the archive folder you created.

From now on, if you click on the "Generate them!"-button at the files page, RePEcPHP will create the ReDIF files and folders in your archive path. Then they are accessible for the RePEc robot. You can view and check the generated files by clicking on the file names listed at the files page.

It is clear that RePEc will only include changes you made to your working paper archive when you updated the ReDIF files by regenerating them.

When you provide your data via a http webserver, the server has to allow directory listing, i.e. the listing of all files in a folder if no index file is found. However, to circumvent problems with a server which is not allowing for directory listing, you can set the variable `$export_create_index` in the main configuration file to `true`. Then RePEcPHP will create files named `index.php` in each directory created. When the robot requests the directory, `index.php` will be served by the webserver, which does nothing else than listing all files and directories found in the folder. Indeed, if you use a ftp server, you don't need this option.

2.5.2 On the fly

With this method you don't use the generation button. Thus you can set the variable `$enable_import` in the configuration file to `false`.

When generating the ReDIF files on the fly, the `.rdf` files the RePEc robot is asking for are processed by the PHP engine of the webserver before delivery. Thus, two preconditions have to be met: first you have to provide your RePEc archive via the same (http) webserver where RePEcPHP is running, and second you have to have write access to the webserver's configuration file.

If both conditions are true, proceed as follows:

Edit the webserver's configuration file to treat files ending with `.rdf` as PHP files. In the Apache2 configuration, this looks like

```
<Files *.rdf>
  SetOutputFilter PHP
  SetInputFilter PHP
</Files>
```

Now copy all files provided with the source in the folder `onthe-fly/` to your archive path. In the following we assume your archive handle to be `kls`.

```
cp -r onthe-fly/* /var/www/html/RePEc/kls/
```

Rename the `.rdf` files in the folder:

```
cd /var/www/html/RePEc/kls/
mv xxxarch.rdf klsarch.rdf
mv xxxseri.rdf klsseri.rdf
```

For every working paper series you maintain, create a copy of the `series/` folder with the series handle as the name. So, if you have two series with the handles `ecopap` and `socpap`, do

```
mkdir ecopap
cp series/* ecopap/
```

```
mkdir socpap
cp series/* socpap/
```

You may delete the `series/` folder afterwards.

Last, edit the file `settings.php` and set the variable `$funcdir` to the path to the folder where the RePEcPHP function files are located. For the examples we used in the installation and in this section, this path would be either relative `../../wpseries/funcs` or absolute `/var/www/html/wpseries/funcs`.

From now on, when the RePEc robot comes along, it will ask for the files `klsarch.rdf` and `klsseri.rdf`. Before delivery, the files will be processed by

PHP and filled with the appropriate data from your working paper archive. Next the robot will browse the series folders to get the paper data.

When you make changes to your archive, they will immediately be reflected in the files created on the fly, and thus included in the RePEc database the next time when the RePEc robot surfs your archive.

3 Usage

You may log in into the administration area by browsing to the URL

```
http://yourwebserv.er/yourseriesfolder/admin/
```

and using the username `repecadmin` and the password `repecphp`. To log out simply click on the link "logout" in the menu.

From the main page, you may access the different pages via the menu provided at top left. For nearly all pages and fields an online help is provided, if the variable `$show_help` in the configuration file `conf/conf.php` is set to `true`.

RePEcPHP supports all ReDIF fields contained in the "Archive", "Series of Papers" and "Papers" template files provided by RePEc.¹⁰ This should be sufficient for nearly all working paper series. Input fields with a red star are mandatory according to the ReDIF standard.

3.1 User

In this section, you can add, edit and delete users who are allowed to maintain the archive. There are two user levels: normal and superuser. The only difference is that superusers are allowed to add, edit and delete other users, while normal users are not. Thus, normal users will not see the user menu option.

3.2 Archive, Series and Papers

On the archive page you fill in the options for the archive. First, all ReDIF archive fields supported are listed and can be filled in. The last input field you need to fill in if you use the export method by button-click. State here the path where the archive should be created. Click the button to save the data.

On the series page you may create, edit or delete working paper series. On the creation/editing page, again first the ReDIF fields are listed for input. The RePEcPHP specific fields follow. In the field "Default download url" you

¹⁰See <http://ideas.repec.org/templates.html>.

may fill in the web address of the directory where your downloadable papers normally reside. If on the paper creation page you fill in a file URL which does not start with `http://` or `ftp://`, the file URL will be appended to the default download address to build the download link. This may save you a lot of keyboard input. An example: all your downloadable papers reside in the web path `http://yourwebserv.er/wpseries/papers`. Then insert this as your default download URL, and on the paper creation page you just have to fill in the file names of the papers.

The next RePEcPHP specific fields are described in full detail in Section 2.3. In the field "Admin paper offset" state how much papers per page should be listed in the papers list in the administration area.

If you click on the papers menu option and only one working paper series is defined, the papers of this series are immediately listed. Otherwise, you have to choose a series first.

On the paper creation/editing page, you can fill in the working paper fields defined by the ReDIF standard. At the very top of the form you may enable or disable a paper. A disabled paper will not appear in the working paper series homepage and in the generated ReDIF files. In the paper list in the administration area, it will be highlighted with a light red background.

The author and file fields are so called "cluster" fields. That means that you can add more than one author/file with their respective corresponding fields. The "key" for authors is the author name, the "key" for files their URL. If the key is present, the author/file data will be saved, if it is empty, the data entry will be deleted. The other fields in the author/file lines are optional. Instead of using the other author fields we recommend authors to register at the RePEc author service and to assign their papers to their profile.

If you create a paper, only two author/file lines are presented in the form. After saving you can always add another author/file. Thus, if you have four authors, add the first two, save, add the third, save, add the last, save.

3.3 Files and Import

These options are only shown if the appropriate configuration variables in `conf/conf.php` are set to `true`. For the "files" page please read Section 2.5, and for the "import" page Section 2.4, respectively.

A Technical notes and license

The program was developed on Debian GNU/Linux Kernel 2.4.21-pre6 and an Apache 1.3.29 webserver, with PHP 4.3.3 and MySQL 4.0.15.

The source distribution contains the file `index.php` serving the publicly accessible working paper series homepage, and the directories `admin`, `funcs`, `install`, `onthe-fly` and `style`.

`admin` contains the administration page with header, footer and menu as well as another folder `mod` with the administration modules corresponding to the menu options.

In `funcs` the main configuration file and all functions used by RePEcPHP are located.

`install` houses a short installation note, the licence and the mysql import file.

In `onthe-fly` you find the default files for ReDIF archive generation on the fly. See section 2.5.2 for details on these files.

Last the `style` folder collects (nearly) all files relevant for the look and feel of the pages served by RePEcPHP.

The software RePEcPHP is distributed under the GNU General Public License. You find a copyright note and the GPL license in the `install` folder.

B The function `parse_archive()`

The function `parse_archive()` is located in the file `funcs/import_funcs.php`. It uses the helper functions `parse_text()` and `strip_nl()`. As arguments `parse_archive()` takes the path to the archive and the archive handle. It returns an array with all the parsed information organized as follows:

```
array(2) {
  ["archive"]=>
  array(Number of fields found) {
    ["Template-Type"]=>
    string(17) "ReDIF-Archive 1.0"
    ["Handle"]=>
    string(3) "xxx"
    ["Full-Handle"]=>
    string(9) "RePEc:kls"
    ... other archive fields ...
  }
  ["series"]=>
  array(Number of series found) {
    [0]=>
    array(Number of fields found) {
      ["Template-type"]=>
      string(16) "ReDIF-Series 1.0"
      ["provider"]=>
      array(Number of provider clusters) {
        [0]=>
        array(3) {
          ["Provider-Name"]=>
          string(46) "University of Cologne, Department of Economics"
          ... other provider fields ...
        }
        ... other providers ...
      }
    }
    ["papers"]=>
    array(Number of papers found) {
      [0]=>
      array(Number of fields found) {
```

```

["Template-type"]=>
string(15) "ReDIF-Paper 1.0"
["Number"]=>
string(1) "4"
["Handle"]=>
string(4) "0004"
["author"]=>
array(Number of authors found) {
    [0]=>
    array(Number of author fields found) {
        ["Author-Name"]=>
        string(11) "Ben Greiner"
        ... other author fields ...
    }
    ... other authors ...
}
["file"]=>
array(Number of files found) {
    [0]=>
    array(Number of file fields found) {
        ["File-URL"]=>
        string(55) "http://www.my.org/wp0004.pdf"
        ["File-Format"]=>
        string(15) "application/pdf"
    }
    ... other files ...
}
... other paper fields ...
}
... other papers ...
}
... other series fields ...
}
... other series ...
}
}
}

```