

RePEc and S-WoPEc:
Internet access to electronic preprints in Economics*

March 15, 1999

Sune Karlsson
Department of Economic Statistics
Stockholm School of Economics
Box 6501
113 83 Stockholm
Sweden
Sune.Karlsson@hhs.se
<http://www.hhs.se/personal/SuneK/>
RePEc:per:1960-09-13:sune_karlsson

Thomas Krichel
Department of Economics
University of Surrey
Stag Hill
Guildford GU2 5XH
United Kingdom
T.Krichel@surrey.ac.uk
<http://gretel.econ.surrey.ac.uk>
RePEc:per:1965-06-05:thomas_krichel

Abstract

The first electronic Economics preprint appeared in 1993. Since then the growth has been dramatic as the use of the World Wide Web has exploded. RePEc has been instrumental in facilitating access to Economics preprints and in bringing order to the chaos that the WWW frequently represents. In a related effort S-WoPEc provides user friendly tools for adding data to the RePEc system. While this is significant in itself it has also been instrumental in fulfilling S-WoPEc's second goal—to provide increased exposure to Swedish Economics research.

*The work discussed here has received financial support by the Joint Information Systems Committee of the UK Higher Education Funding Councils through its Electronic Library Programme (RePEc/WoPEc) and the Swedish Royal Library through its BIBSAM office (S-WoPEc).

1 Introduction

Electronic dissemination of Economics working papers can be traced back to April 1993 and the start of the Working Papers in Economics (WoPEc) project. By March 1999 this single archive has grown into an interconnected network of over 60 archives holding over 13,000 downloadable papers and over 50,000 descriptions of offline papers from close to 1,000 series, as well as data about over 4,000 academic Economics department and research institutes. We call this network of archives Research Papers in Economics or RePEc for short. All archives operate independently. There are several end user services in the form of databases and current awareness e-mail reports. RePEc will soon register authors and work towards a relational database that describes the output of the discipline.

RePEc archives range in size and scope from small departmental archives providing a single working paper series to national initiatives providing data about working papers from several universities and research organisations. Swedish Working Papers in Economics, S-WoPEc at <http://swopec.hhs.se/>, is an example of such a national initiative. The aim of S-WoPEc is to give exposure to Swedish Economics research by acting as a facilitator for organisations' provision of electronic working papers. S-WoPEc operates a web site giving access to Swedish Economics working papers, provides tools for entering and maintaining the bibliographic data about working papers and optionally hosts the full text files. This approach to promoting electronic publication has proven quite successful and Sweden has one of the highest participation rates in RePEc.

The remainder of the paper is organised as follows. In Section 2 we describe the historical background before the foundation of RePEc. In Section 3 we discuss some important aspects of RePEc. Section 4 introduces S-WoPEc, and Section 5 concludes the paper.

2 Historical Background

RePEc is essentially an initiative of the NetEc project. The NetEc project was founded in February 1993, by Thomas Krichel. At the outset NetEc was conceived as a collection of projects to help the distribution of research results in Economics via the Internet. Economists had been exchanging preprints or "working papers" as they are more commonly known for a long time in printed form. There were no electronic working papers at the time when NetEc was founded. However a donation of Fethy Mili of Université de Montréal made bibliographic data about 250 working paper series available. The data for some series went back as far as 1988. The first gopher servers allowed to store data in a way that an uninitiated user could easily retrieve it. Manchester Computing Centre donated disk space and CPU time. Geoff Lane, their system administrator, managed to build a facility for WAIS database queries on the gopher system soon after the start of the project. This was a state of the art system at the time. However this medium was not much used. The main usage of the data came through mailing current awareness data to the corryfee@hasaravm1.earn (yes, a bitnet address) mailing list for academic economists. The project dealing with printed papers was called BibEc. BibEc was NetEc's first constituent project.

In April 1993, "MatClass: a Matrix Class for C++" by Chris Birchenhall became the world's first electronic Economics working paper. This single paper would have been lost in the over 5,000 items strong dataset about the printed papers. A separate name WoPEc was invented for the collection of downloadable papers. On the gopher screen the presentation of WoPEc was quite primitive. Each paper was represented by two lines. The first line showed the name of the author(s). Selecting that line lead to the bibliographic

information. That would be the author, title, sometimes abstract, Journal of Economics Literature (JEL) classification, etc. The second line showed the title of the paper and when the user hit this line the full text of the paper was delivered to her machine.

The text of the paper, usually in PostScript format, was not necessarily held on the Manchester site. Instead it could have been on any other machine. Although WoPEc started off as collection of papers—simply because it was the first collection of its kind to open—it has focused on the collection of metadata about papers rather than the papers themselves. This shift of activity became more prominent since September 1993 when the “Economics Working Paper Archive” opened at <http://econwpa.wustl.edu> sponsored by Washington University in St. Louis. As suggested by the name, this site claimed to be a central archive for working papers in Economics. This vision was imported from High Energy Physics, where the XXX Archive at <http://xxx.lanl.gov> collects all papers on a central site. To establish the same degree of coordination among economists seems to be more difficult, there are nowadays (early 1999) several hundreds of web and ftp servers with online working papers.¹ Our theory is that economists have a built-in distrust of monopolies. In their book of tales, there are numerous accounts of the welfare losses caused by monopoly supply. They may also be afraid of the power accumulated by a person who controls a hard disk where the complete output of the discipline is stored. That is why a centralised system has problems to reach broad acceptance. That does not only hold for an archive that stores papers, it also affects the collection of metadata about holdings on local sites, i.e., the work WoPEc was primarily concerned with. Right from the start Thomas Krichel believed that a distributed database protocol was the way of the future in order to allow a many institutions to contribute to WoPEc without need for extensive coordination. In that way each institution would keep its own papers on its own server, and a user would be able to query many servers. Such a protocol started to appear in 1994 with the whois++ protocol. In Summer 1995, José Manuel Barrueco Cruz and Thomas Krichel converted the dataset to the format proposed by the Internet Anonymous Ftp Archive (IAFA) working group of the Internet Engineering Task Force. These were meant to be compatible with the whois++ protocol. The IAFA data was converted to static web pages. These web page were indexed with WAIS.

In the spring of 1995 the US mirror of BibEc and WoPEc appeared at Washington University of St. Louis, and a Japanese mirror at Hitotsubashi University Tokyo was added in the autumn of that year. In 1996, WoPEc was awarded funding by the Joint Information Systems Committee (JISC) of the UK Higher Education Funding Councils. José Manuel Barrueco Cruz, who was working on the project as a volunteer before, took the post of project manager. The objectives of the JISC funding were two-fold, first to raise the total stock of papers in WoPEc above the 1,000 papers held before the start of the funding period, and second to increase the participation of UK universities. For the first objective the project was very successful, by the middle of the following year, the stock had increased to 4,500. But the response from UK institutions was sluggish. In late 1996, Thomas Krichel initiated contacts between the Royal Library of Sweden and Sune Karlsson that lead to the initial funding of the S-WoPEc project. In early 1997, the Dutch DEGREE working paper publishing consortium agreed to participate in the exchange of working paper data. With three institutions working together, it was felt that a protocol should be established that would allow the participants to exchange data without need for extensive coordination. Another aim was to separate the collection of

¹Some are homepages of authors, others are organised by academic departments or research centres. In some rare cases we find common catalogues of participating institutions, for example the US Federal Reserve have the Fed in Print catalogue of publications issued by the regional banks.

data and the implementation of the data in user services. A set of archives would provide data on papers. The full text of these papers would live in the archives or on other publicly accessible computer systems, or simply be offline. Robot software was to be written to mirror these archives. Thomas Krichel prepared draft protocols. These were adopted at a meeting on between DEGREE, S-WoPEc and WoPEc in Guildford on 12 May 1997. These protocols—with very minor modifications—are still used by RePEc today.

3 RePEc

The nature of RePEc is not precisely defined, but RePEc is most commonly understood as referring to three things. First it is a collection of archives that provide data about Economics. Second it is the data that is found on these archives. Third, it is often also understood to represent the persons and institutions building archives and providing the data. RePEc has no formal management structure.

RePEc has two aims. The “library aim” is to provide a complete description of the Economics discipline that is available on the Internet. The “publishing aim” is to provide *free* access to Economics resources on the Internet.² By “free” we mean that the provider, rather than the user pays. RePEc has no peer-reviewing aim, however it can be used to support peer review.

RePEc is based on the following principle

$$\text{Many archives} \implies \text{One dataset} \implies \text{Many services}$$

The collection of material within RePEc is decentralised. Each participating provider provides an archive on a http or ftp server that will allow the storage of structural data about objects relevant to Economics, and possibly the storage of some of the objects themselves. We will examine the structure of an archive in Subsection 3.1.

Robot software is used to unite the various archives. The key concept here is the “site”, which is a unification—on a single machine—of a local archive and one or more remote archives that are mirrored. If a site mirrors all archives known to RePEc then it has access to the complete RePEc dataset. The contents of the RePEc archives forms a single dataset, where each component is identified. Therefore a relational structure can be put in place. We will detail the basic logical structure of the relational model in Subsection 3.2. Some details of that structure have yet to be implemented.

RePEc sites are used to build user services. There is no official user service for RePEc. We will look at some user services in Subsection 3.3.

3.1 The structure of an archive

RePEc stand on two pillars. First, a template format inspired by the IAFA templates called ReDIF. This acronym stands for something like *Research Documentation Information Format* but it is best understood as a literal. ReDIF defines a number of templates. Each templates has a set of allowable fields, some are mandatory, and some may be repeated. The second pillar is the Guildford protocol. It fixes rules how to store ReDIF in an archive. It is possible to deploy ReDIF without using the Guildford protocol. But in the following

²These aims are sometimes conflicting. For example, assume that a certain amount of money is available for cataloguing purposes. Then the library objective might be best served by using these funds to gather information about a high-quality toll-gated journal resource, whereas the publishing objective would be better served by considering a collection that is on the Internet and may not be of the same quality since it has not yet been extensively peer-reviewed.

we will ignore this conceptual distinction, because it is easiest to understand the structure and contents of an archive through an example.

RePEc identifies each archive by a simple identifier or handle. Here we look at the archive RePEc:sur which lives at <ftp://www.econ.surrey.ac.uk/pub/RePEc/sur>. On the root directory of the archive, two files are mandatory. The file *surarch.rdf* contains a single ReDIF archive template.

```
Template-type: ReDIF-Archive 1.0
Name: University of Surrey Economics Department
Maintainer-Email: T.Krichel@surrey.ac.uk
Description: This archive provides research papers from the
             Department of Economics of the University of Surrey, in the U.K.
URL: ftp://www.econ.surrey.ac.uk/pub/RePEc/sur
Homepage: http://www.econ.surrey.ac.uk
Handle: RePEc:sur
```

In this file we find the basic information about the archive. The other mandatory file is *surseri.rdf*. This must contain one or more series templates.

```
Template-Type: ReDIF-Series 1.0
Name: Surrey Economics Online Papers
Publisher-Name: University of Surrey, Department of Economics
Publisher-Homepage: http://www.econ.surrey.ac.uk
Maintainer-Name: Thomas Krichel
Maintainer-Email: T.Krichel@surrey.ac.uk
Handle: RePEc:sur:surrec
```

The papers for the series RePEc:sur:surrec are confined to a directory *surrec*. It may contain files of any type. Any file ending in *.rdf* is considered to contain ReDIF templates. Let us consider one of them, *surrec/surrec9601.rdf*³

```
Template-Type: ReDIF-Paper 1.0
Title: Dynamic Aspect of Growth and Fiscal Policy
Author-Name: Thomas Krichel
Author-Email: T.Krichel@surrey.ac.uk
Author-Name: Paul Levine
Author-Email: P.Levine@surrey.ac.uk
Author-WorkPlace-Name: University of Surrey
Classification-JEL: C61; E21; E23; E62; O41
File-URL: ftp://www.econ.surrey.ac.uk/pub/RePEc/sur/surrec/surrec9601.pdf
File-Format: application/pdf
Creation-Date: 199603
Revision-Date: 199711
Handle: RePEc:sur:surrec:9601
```

The pair (File-URL:, File-Format:) could be repeated any number of times. An additional File-Function: field may be added to indicate the status of the file within the paper. Note that in this example the full text of the paper is located within the archive directory structure itself. Therefore the full text is mirrored together with the bibliographic data. Services can then link to a mirrored copy of the full text. If the URL of the paper would

³We suppress the Abstract: field to conserve space.

point to a place outside the archive structure, the link to the full text would always point to that location.

There is a central archive RePEc:all, that mirrors all the *???arch.rdf* and *???seri.rdf* files from all archives. It also contains the software that allows sites to mirror archives. RePEc:all also provides reading and checking software for templates as well as general RePEc documentation. This archive lives at <ftp://netec.mcc.ac.uk/pub/RePEc/all>.

3.2 The relational structure

RePEc views the discipline as a set of four basic elements that are in relation to each other. The basic model can be represented in the following table

paper	collection
person	institution

Here the “paper” can be a preprint or a “published” article, a book, piece of software, a dataset etc. The types most frequently found now are preprints and articles, but there are already some software components in the RePEc dataset. The “collection” can be any set of papers grouped together. At the moment the collection definition as used in RePEc comprises preprint series and journals of “published” articles. Note that each of the papers is initially part of a single series. In principle the collection concept could be used to allow for peer review by placing the works that are reviewed in a separate collection. However it would also be possible to add a review type in the relational framework.

At the time of writing the personal information is built into the templates of the papers. There are however efforts being made to allow for the separation of personal information. Soon we will see things like

```
Template-Type: ReDIF-Person 1.0
Name: Thomas Krichel
Email: T.Krichel@surrey.ac.uk
Author-Paper: RePEc:sur:surrec:9404
Author-Paper: RePEc:sur:surrec:9601
Homepage: http://gretel.econ.surrey.ac.uk
Handle: RePEc:per:1965-06-05:thomas_krichel
```

and we can replace the author information for the first author in the paper template for RePEc:sur:surrec:9601 by

```
Author-Name: Thomas Krichel
Author-Person: RePEc:per:1965-06-05:thomas_krichel
```

The benefit of this system are clear. There is a much reduced load on administration of the system. When the phone number of an author changes, this change has to be registered at only one point in the system. A pervasive use of these relational features will allow the resolution of current author information through the current person template of the author. The user of a RePEc service would therefore find the author of the paper, even though the contact information on the papers’ title may no longer be current.

Finally the institution can be thought of as a set of persons just like the collection is a set of papers. Author registration services will use the institution data that RePEc has already collected.

3.3 User services

The implicit ability and explicit intention to allow for many user services at a time is a key feature of RePEc. The disadvantage of this approach is that RePEc is a more abstract concept to understand than say an archive like XXX where storage and user service are linked together. However this provides an important selling point once a potential provider understands that submitting data to RePEc means submitting the data to all the user services at once. Here we list the most important user services in order of historical origin.

BibEc at <http://netec.mcc.ac.uk/BibEc.html> &

WoPEc at <http://netec.mcc.ac.uk/WoPEc.html>

provide static html pages for all working papers that are only available in print (BibEc) and all papers that are available electronically (WoPEc). Both datasets use the same search engines. There are three search engines, a full text WAIS engine, a fielded search engine based on mSQL and a ROADS fielded search engine. Both are mirrored in the United States and Japan as part of the NetEc project.

EDIRC at <http://ideas.uqam.ca/EDIRC>

provides a browsable index of the institution data in RePEc. This service is also mirrored by NetEc.

IDEAS at <http://ideas.uqam.ca>

provides an Excite index of static html pages that represent all Paper, Article and Software templates. This is by far the most popular RePEc user interface.

NEP: New Economics Papers at <http://netec.wustl.edu/NEP>

is set of reports on new additions of papers to RePEc. Each report is edited by subject specialists who receive information on all new additions and then filter out the papers that are relevant to the subject of the report. These subject specialists are PhD students and young researchers. They work as volunteers.

INOMICS at <http://www.inomics.com/query/search>

provides an index of RePEc data, but also allows a simultaneous search in indexes of other web pages related to Economics.

Finally note that a Z39.50 server for all downloadable papers in RePEc is available at dbiref.kub.nl:9997. The database name is "repref". The attribute set is Bib-1, and the record syntax supported are USmarc, SUTRS, GRS-1 (only string tags, tag type 3).

4 S-WoPEc

S-WoPEc and its sister site S-WoBA have two basic and complimentary aims. The first is to give exposure to and increase the awareness of Swedish research in Economics and Business Administration. The second is to facilitate and encourage participation in RePEc.

By combining these two features participation in S-WoPEc becomes very attractive. A university department or research institute gets assistance in putting its research output online and at the same time gets enhanced exposure. As a result all major and many minor Swedish Economics research organisations participate in S-WoPEc less than two years after the launch of the site. S-WoBA has recently been launched and the participation is considerably lower. Similarly to S-WoPEc we expect this to change rapidly.

4.1 Making research visible

Towards the first aim RePEc end user services featuring the Swedish contribution to RePEc are provided at <http://swopec.hhs.se/> and <http://swoba.hhs.se/>. Compared to the old paper based distribution system for preprints these sites are extremely successful,

averaging between two and three thousand unique downloads of full text files a month. The synergy effects from participating in S-WoPEc and S-WoBA are obvious. Very few organisations can provide the richness of material available at S-WoPEc on their own. By participating an organisation adds to the attractiveness of the site *and* benefits from the large number of visitors attracted by the critical mass of the site.

In addition, the data about the papers are automatically included in the S-WoPEc or S-WoBA RePEc archives and are thus locatable through RePEc services such as WoPEc, IDEAS and NEP.

At S-WoPEc and S-WoBA the abstract of each paper is presented as a static html page generated from metadata stored as a ReDIF template. Where available, the abstract page links to full text files stored locally on the S-WoPEc server or a server maintained by the participating organisation. A fielded search facility is implemented using meta tags in the html pages.

4.1.1 Dublin Core

S-WoPEc and S-WoBA make heavy use of Dublin Core. Where applicable the meta tags used for the fielded search are based on Dublin Core. Most of the metadata is, in fact, stored in the html documents using Dublin Core meta tags. This allows Dublin Core aware search engines to extract structured data and provides for more precise searches.

As of yet there are relatively few Dublin Core aware search engines but we expect this to change as the use of Dublin Core increases. A notable example of a Dublin Core aware search engine is SAFARI (<http://safari.hsv.se/>) which indexes research related html pages on the webservers of Swedish universities including the S-WoPEc and S-WoBA servers. The research of participating organisations is thus included in SAFARI without any extra effort on the organisations' part.

4.1.2 Library systems

In addition to being searchable in various RePEc services and Internet search engines, the papers contributed to S-WoPEc and S-WoBA are catalogued in LIBRIS, the union catalogue of Swedish research libraries. The LIBRIS WWW interface is available at <http://www.libris.kb.se>. When using the WWW interface the user can directly access the abstracts and downloadable files at S-WoPEc and S-WoBA through links.

4.2 Facilitating participation in RePEc

Running a RePEc archive can, depending on the resources and technical know-how available, be a daunting proposition. To facilitate the contribution of data to RePEc, S-WoPEc and S-WoBA provides a streamlined process for entering or modifying data about preprints and uploading full text files to the servers.

The user interface is implemented using html forms and Perl CGI scripts and is thus platform independent. Data is entered in simple forms and is validated to ensure that syntactically correct ReDIF templates are written. In this process we distinguish between two types of users, authors and working paper coordinators (WP-coordinator). Authors as well WP-coordinators can submit data about a paper to S-WoPEc or S-WoBA but the author submitted papers must be approved by the participating organisation's WP-coordinator before the data about the paper is made available. This is to safeguard against papers being added to the wrong series by mistake. Some organisations choose to disable author submission of data and delegate this work to a few persons within the organisation.

In the description of the process we will take S-WoPEc as an example. The procedure is identical for S-WoBA.

4.2.1 Collecting working paper metadata

This script lives at <http://swopec.hhs.se/scripts/MetatData.pl?s=series> (where *series* is the RePEc series identifier). When MetaData.pl (as well as Update.pl and Upload.pl) is called without the *s=series* argument the user is presented with a form where the series can be selected.

The script reads the configuration file for the series and presents a html form for collecting the basic metadata about the paper. At this stage a WP-coordinator can enter his/her password and the number of the working paper. On submit, the information is checked and the submitter can correct erroneous data and add missing required information. When there are no remaining errors, the information is pretty printed back and the user is given the opportunity to correct any mistakes or commit the data. If a WP-coordinator is submitting the information there are buttons for providing data for additional papers and uploading files for the paper. The latter takes the WP-coordinator directly to the upload form.

When the information is committed the script takes the following actions:

1. Assigns the author access code to the paper.
2. Writes a ReDIF paper template for the paper
 - (a) If the submitter is not a WP-coordinator:
 - i. Writes the collected information to a temporary ReDIF paper template with a name based on the author access code for the paper.
 - ii. E-mails the collected information to the WP-coordinator and asks for verification of the paper. This verification is provided by replying to the e-mail with the text of the original e-mail quoted and adding the WP-number to the appropriate line of the quoted text. If the paper does not belong to the series the WP-coordinator indicates this by entering 'delete' on the Number line.
 - (b) If the submitter is a WP-coordinator:
 - i. Writes the collected information to a permanent ReDIF paper template file.
 - ii. E-mails the WP-number and password for the paper together with instructions for uploading the paper and updating the information about the paper to the submitter and optionally the authors for which we have an e-mail address.
 - iii. The writing of an abstract html file and update of the appropriate index page is scheduled by sending an e-mail which is later read by the scheduled script MailCheck.pl.

The ReDIF paper template A few non-standard fields are added to the ReDIF paper template by MetaData.pl

X-MyEmail: e-mail of submitter
X-PW: author access code assigned to WP
X-Series-Code: the RePEc series identifier

If a WP-coordinator submits the data all fields for which data is available are written to the template file. (The scripts doesn't allow committing data until there is data for all required fields.)

When an author submits data, the number of the working paper isn't available and the Number: and Handle: fields are not written to the temporary template file. The temporary template file is promoted to a permanent template and the paper is added to the archive when the Working Paper Coordinator assigns a number to the paper by replying to the e-mail in 2.a.ii.

4.2.2 Scheduled processing

The script MailCheck.pl runs at regular intervals, currently on the hour from 08:00 to 22:00, and reads incoming mail. This incoming mail signals actions that MailCheck.pl should take. There are currently three actions that MailCheck.pl recognises.

1. Promote a temporary template to a permanent one.

This is signalled by the Working Paper Coordinator replying to an e-mail sent from the script account (2.a.ii above).

The information needed to locate the temporary template file and the number assigned to the paper is extracted from the e-mail and verified. If the data is valid a permanent template file is written, html files are generated for the paper and a notification mail to the submitter and authors is sent. Otherwise a new mail is sent to the WP Coordinator asking for correct data or (depending on the type of the error) the e-mail is forwarded to an S-WoPEc administrator for human intervention.

2. Reject a paper

This occurs when the word delete is added on the Number: line instead of a number for the working paper. The temporary paper template is deleted and the submitter is notified by e-mail.

3. Generate html files for a working paper.

A html for the abstract of the paper is written and the index html file for the year of the working paper is updated.

4.2.3 File upload

This script provides the forms interface for uploading full text and other files for the working paper. In the notification e-mail sent to authors upload instructions and a http URL providing transparent validation of the uploader is included. There is a similar URL for the WP coordinator giving access to the series, after access is granted the WP-coordinator can specify the number of the paper to work with.

Files can be uploaded using http file upload or the url for files stored on the organisations own server can be supplied. In the latter case S-WoPEc links to the organisations server and does not keep a copy of the file on the S-WoPEc server.

When working with file uploads we divide the files in two groups: "full text" files which provide content corresponding to the printed version of the working paper. This may be a single Word document, a Word document together with a file containing Excel charts, a LaTeX file and so on. The second group is "additional files" containing supplemental material such as datasets and source code for computational routines used in the paper.

This is a distinction not present in ReDIF itself and the purpose is mainly to provide more information to the users of S-WoPEc.

Apart from the name of the file(s) the upload script gives the user the opportunity to provide some additional information about the files; a description of the content and the type (format) of file (e.g. a Word document or a SPSS dataset).

When this information is submitted it is added to the paper template as the non-standard fields

File-X-Desc: description of content supplied by uploader

File-X-FileFormat: type/format of the file

File-X-FileType: ''full text'' or ''additional file''

In addition we ask for the e-mail address of the uploader and this will replace the previous content of the X-MyEmail: field in the paper template.

Linking to other servers When the user provides a url for a file on a non S-WoPEc server the url is first verified by the script. If it is correct the url is added to the paper template and a new abstract html page with a link to the downloadable file is written.

HTTP file upload With HTTP file upload it is not practical to enforce that the description of the content of the file and the type of file is supplied. When no e-mail address is supplied this is replaced with the IP number of the uploader, thus providing some means of identification. On the other hand we will know when the file is uploaded and we know the names of the files. In addition, the browser will in most cases provide us with the MIME type of the uploaded file. This give us essentially the same information as when the uploader provides information about the type/format of the file.

With http file upload we will thus add three more fields to the paper template:

File-X-OrgName: name supplied by uploader

File-X-LocalName: Name of file as stored on the server

File-X-MIME: MIME type as supplied by the browser

On a successful upload the uploaded files and their file sizes (providing a minimal check on the success of the upload) are printed back to the uploader and a summary of the available files is given. If a failure is detected any uploaded files are discarded, an error message is printed and the upload form is presented again.

Finally an e-mail notification about the uploaded files is sent to the archive administrator. S-WoPEc staff then convert the uploaded full text files to postscript and pdf as appropriate and package the additional files into a ZIP archive. The files are moved to the series online file directory on the web server. When this is completed File-URL:, File-Format: and possibly File-Function: fields are written to the template and the publicly available template in S-WoPEc's RePEc archive is updated.

Removing online files In order to comply with copyright restrictions when a working paper is published, the author or WP-coordinator must be able to remove the full text online files if this is needed.

Upload.pl and Update.pl list the available online files when the script is first called and offer the option to remove files. Files are removed in groups, i.e. all the full text files can be removed or all the additional files can be removed (this should never be necessary). When removing the files the corresponding fields in the paper template are removed and files on the web site are deleted. A notification e-mail is then sent to the authors and the submitter, informing them that online files have been removed.

4.2.4 Revising the metadata and adding publication information

The script Update.pl provides means for updating the working paper metadata and adding information about published versions of the paper. As for Upload.pl there is a transparent http url in the e-mails sent to authors and submitters of data and the WP-coordinator has access to the script using the coordinator password for the series.

After the user has been validated the scripts presents the available information about the paper; the metadata about the working paper, any uploaded files and publication information. The user is given the options to revise available information, upload new files, remove current on line files and to add or revise publication information.

There are currently three types of publications; Article for journal articles, Book and Chapter for publications in proceedings volumes, reprint volumes and other collections of papers. The processing follows the same lines as for the original submission or revision of metadata. The data are saved to Article, Book and Chapter templates.

4.2.5 Feedback to authors

As the success of S-WoPEc and RePEc in general depends on the willingness of authors and organisations to contribute data it is important that they are made fully aware of the exposure and value added provided by S-WoPEc and RePEc. To this end we gather access statistics on a per paper basis from the http logs and send monthly e-mails to the authors and WP-coordinators summarising the accesses for each paper or organisation. In our experience this feedback is appreciated and has a tendency to arouse the authors' competitive spirits.

5 Conclusions

The decentralised RePEc model for providing data about research related activities has proved to be extremely successful in Economics. In the short time span of 2 years it has gained wide acceptance within the discipline and is now the premier source of information about recent research in Economics. This is in contrast with the monolithic xxx.lanl.gov model of Internet-based preprint distribution which has failed to catch on in Economics. The xxx model, represented by EconWPA, still plays a useful role in Economics as it provides one way of adding data to the RePEc system. But our experience suggests that different models suit different disciplines, depending on the traditions of the discipline.

RePEc is broader in scope than other initiatives we are aware of. Distribution of preprints is just one—though important—aspect of the research process that RePEc collects data about. In addition, RePEc promotes innovation in the area of end user services. Any participating archive can use the data to provide its own brand of end user service with unique capabilities, user interface or criteria for which material to include.

Participating in RePEc is not without effort although it should be clear that the benefits more than compensate for the effort. Still, the efforts required for the initial set up and maintenance of a RePEc archive may be viewed as too large by some organisations. S-WoPEc demonstrates one successful way of lowering the threshold for participating in RePEc by providing a streamlined process for entering metadata and relieving organisations of the requirement to run their own archive.