

Enhancing Information Flow in Economics via Linked Metadata Archives

Christopher F Baum

Department of Economics, Boston College

October 1999

Prepared for AUBER Conference, Little Rock

<http://fmwww.bc.edu/RePEc/docs/AUBERf99.pdf>

Linked Metadata Archives

The RePEc (Research Papers in Economics) project is an international effort to enhance the information freely available to economics researchers through the development of linked metadata archives.

Metadata are the bibliographic details, similar to the information content of a library catalog, which describe a particular archive component and permit it to be located by its title, author(s), keywords, or words in its description (abstract). It extends the library card catalog model to hyperlinks, which might allow you to read a working paper, download a journal article, or install a software component on your desktop computer.

RePEc archives and services

A RePEc archive is a set of templates—ASCII files, which may be produced manually or automatically from other information sources—which are automatically processed on a daily basis by the software underlying RePEc services.

A RePEc service provides a user interface to these metadata. There are many archives (over 100 at this time), but only one “virtual database” containing the compendium of all archives’ contents. The model:

- Many archives
- One database
- Many services

Open access for non-commercial use

The data contained in RePEc are freely accessible to any who wish to make use of it, or repackage it via a RePEc “service” for any non-commercial purpose. The inclusion of meta-data in RePEc does not preclude a charge being made for the components themselves: e.g. a journal article may be downloadable only by subscribers, or a NBER working paper or an institute’s research report may require payment. It does imply that the summary information is available without charge.

Contrast this with existing services such as *EconLit*, which are not freely available over the Web. You must access them through a library which has subscribed to web access, and deal with firewall and authentication issues.

RePEc versus individual Web servers

Each research institution, or individual, could make their work freely available via the Web. General-purpose search engines could in principle find it by reference to its title, keywords, etc. Why do we need metadata archives?

- Search engines are looking for the “needle in the haystack”
- A metadata archive is searchable without reference to the entire WWW
- Items classified by e.g. JEL code can be grouped

RePEc versus individual Web servers

The individual archives of materials—e.g. working papers or research reports from a single department or research center—are useful as a method of disseminating those materials. If a user knows that Dr. John Doe is in the BC Economics Department, she can find his papers on the BC Economics web site.

But what if that affiliation is not known? Or if she is a Ph.D. candidate searching for recent working papers on a particular subject? Then the existence of hundreds or thousands of web servers without a unifying framework becomes a serious detriment.

RePEc versus monolithic archives

For more specialized materials—e.g. statistical software tools—a generalized search is almost worthless. One approach involves the creation and maintenance of a ‘monolithic’ archive. But that requires substantial hardware and software resources, and funding to maintain the archive, which for a collection of thousands of items will be nontrivial.

Furthermore, the monolithic archives may not be unique. If competing archives emerge, contributors will question whether they should participate in several to maximize visibility of their materials. This is cumbersome to say the least.

The RePEc concept

RePEc addresses these concerns via its design concepts. It is not a monolithic archive, but a decentralized network of archives. Each archive is maintained by local information providers or volunteers who serve to ‘catalog’ materials of one or more institutions. The set of archives can be scaled to any size, with minimal central administration; the only coordination required is the allocation of archive names.

Via decentralization, each institution can choose to provide information about those research materials which they are interested in sharing via RePEc. This information generally overlaps with the information compiled for local web pages, and RePEc templates may be generated mechanically (or manually, with minimal effort) from the details already ‘published’ on the Web.

A ReDIF working paper template

Template-Type: ReDIF-Paper 1.0

Title: Is the Proportion of College Workers in 'Non-College' Jobs Increasing?

Author-Name: Peter Gottschalk

Author-Email: peter.gottschalk@bc.edu

Author-WorkPlace-Name: Boston College

Author-WorkPlace-Postal: Dept. of Economics, Boston College, Chestnut Hill,
MA 02467 USA

Author-Name: Michael Hansen

Author-WorkPlace-Name: Center for Naval Analysis

Abstract: This paper explores the claim that college educated workers are increasingly likely to be in "non-college" occupations. We provide a conceptual framework which gives analytical content to the previously vague distinction between college and non-college jobs. This framework is used to show that skill bias technological change will lead to a decline in the proportion of college workers in non-college jobs. This prediction is supported by the data.

Series: Boston College Working Papers in Economics

Number: 429

Creation-Date: 19990910

Length: 32 pages

File-URL: <http://fmwww.bc.edu/EC-P/WP429.pdf>

File-Format: application/pdf

File-Size: 209 Kb

File-Function: main text

Handle: RePEc:boc:bocoec:429

8 bis

RePEc for software distribution

The RePEc framework was designed to provide metadata for ‘papers’ (working papers, or preprints), ‘articles’ (published articles), and other print materials such as ‘chapters’ or ‘books.’ But the framework may be used for the storage of metadata of any sort, and potentially to provide access to an item stored in virtually any format (HTML, PDF, ZIP file, tar.gz archive, etc.)

These characteristics make the RePEc framework very useful for the dissemination of information about **software components**, and for the delivery of those components themselves, as well as sample programs, data, and documentation.

A ReDIF software template

Template-Type: ReDIF-Software 1.0
Title: MDENSITY: Stata module for univariate kernel density estimation,
for variables or groups
Author-Name: Nicholas J. Cox
Author-Email: N.J.Cox@durham.ac.uk
Author-WorkPlace-Name: University of Durham
Programming-Language: Stata
Abstract: mdensity produces kernel density estimates for one or more variables
or groups and graphs the result. mdensity is a wrapper for kdensity, which is
called in turn for each variable or group specified. This is version 1.5.1 of
the software.
Keywords: density estimation, distributions, graphics
Requires: Stata version 6.0
Note: This module may be installed from within 'net-aware' Stata by typing
"net cd <http://fmwww.bc.edu/repec/bocode/m>" followed by "net install mdensity"
Series: Statistical Software Components
Number: S391601
Creation-Date: 19990930
Revision-Date: 19991012
File-URL: <http://fmwww.bc.edu/repec/bocode/m/mdensity.ado>
File-Format: text/plain
File-Size: 225 lines
File-Function: program code
File-URL: <http://fmwww.bc.edu/repec/bocode/m/mdensity.hlp>
File-Format: text/plain
File-Size: 210 lines
File-Function: help file
Handle: RePEc:boc:bocode:S391601

9 bis

What are **software components**?

Many statistical packages (e.g. Stata, SAS, RATS, Ox) or specialized programming languages (Mathematica, GAUSS, MATLAB, S+) support the use of 'components': specialized functions, or procedures, or modules that add functionality to the package. These 'components' are not programs per se, but rather components of programs: i.e. functions that may be called within a user program, or additional commands that may be invoked by the user of that package or language. Some degree of generality is implied, in that useful components are not completely specific to their author's task at hand, but have the capability to perform some function based on their argument list.

RePEc and software components

Two years ago, the RePEc standard was extended to include a ReDIF-Software template, and the first RePEc series containing software templates was established at the Boston College Department of Economics. This series, the **SSC** (Statistical Software Components) archive, was designed to provide users of statistical packages and specialized programming languages with a way of making their public-domain contributions accessible via the Web.

Many users of RePEc metadata employ the IDEAS service. IDEAS, maintained by Christian Zimmermann at UQAM (Montreal), is accessible at <http://ideas.uqam.ca>.

SSC-IDEAS

The SSC archive, when accessed in IDEAS, provides access to almost 400 software components, for Stata, MATLAB, Mathematica, GAUSS, and Ox. Contributions in any language (including code in standard languages such as FORTRAN and C) are welcomed. The vast majority of the components are Stata code, most having appeared in the Statalist listserv within the last two years. These components have been contributed by a wide range of authors from the US, the UK, Europe and Australia. Some components have been written or coauthored by StataCorp staff, including Vince Wiggins, and represent 'preview' versions of software that will eventually become available in 'official' Stata.

This file is part of [IDEAS](#), which uses [RePEc](#) data

[[Series](#) | [JEL Classification](#) | [Search](#) | [Viewers](#) | [Help!](#)]

If a paper is available in full text, download information is at the bottom of its page between two lines. If not, see [Help!](#)

Boston College Department of Economics

Statistical Software Components

Publisher-Postal: Boston College, Carney Hall, Chestnut Hill MA 02467 USA

Publisher-Fax: +1-617-552-2308

Publisher-Homepage: <http://fmwww.bc.edu/EC/>

For information about this series, please contact baum@bc.edu.

[RePEc:boc:bocode:S375501](#) **ARCHUTIL:** Stata modules to list, type, copy contents of SSC-IDEAS Archive *by* Christopher F Baum & Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:T961401](#) **ECONOMETRICS:** MATLAB toolbox of econometrics functions *by* James P. LeSage [[Downloadable!](#)]

[RePEc:boc:bocode:S392801](#) **CFLPOIS:** Stata module to calculate confidence limits for rates based on Poisson outcome *by* Jens M. Lauritsen [[Downloadable!](#)]

[RePEc:boc:bocode:S391701](#) **MATSELRC:** Stata module for selection of rows and/or columns from matrix *by* Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S392602](#) **BYVAR:** Stata module to repeat a command by variable *by* Patrick Royston [[Downloadable!](#)]

[RePEc:boc:bocode:S392601](#) **STQUANT:** Stata module to estimate quantiles for survival time *by* Enzo Coviello [[Downloadable!](#)]

[RePEc:boc:bocode:S391801](#) **KDMANY:** Stata module to perform kernel density estimation for several variables *by* Stanislav Kolenikov [[Downloadable!](#)]

[RePEc:boc:bocode:S391601](#) **MDENSITY:** Stata module for univariate kernel density estimation, for variables or groups *by*

Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S391301](#) **LISTUTIL: Stata modules to manipulate lists of words** by Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S390602](#) **BPAGAN: Stata module to perform Breusch-Pagan test for heteroskedasticity** by Christopher F Baum & Vince Wiggins [[Downloadable!](#)]

[RePEc:boc:bocode:S390601](#) **WHITETST: Stata module to perform White's general test for heteroskedasticity** by Christopher F Baum & Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S390502](#) **LFSUM: Stata module to summarize variables located by lookfor** by Fred Wolfe [[Downloadable!](#)]

[RePEc:boc:bocode:S390501](#) **LOOKFOR2: Stata module to locate variables (enhanced lookfor)** by Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S389602](#) **TOLOWER: Stata module to rename variables with lower case names** by Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S389601](#) **XPOSE2: Stata module to transpose a dataset (version 5 only)** by William W. Gould [[Downloadable!](#)]

[RePEc:boc:bocode:S389501](#) **FELDTI: Stata module to compare two independent coefficients of reliability** by Herve M. Caci [[Downloadable!](#)]

[RePEc:boc:bocode:S388902](#) **DOUB2FLT: Stata module to change storage type of real variable** by Fred Wolfe [[Downloadable!](#)]

[RePEc:boc:bocode:S388901](#) **RECAST2: Stata module to change storage type of variable** by Fred Wolfe [[Downloadable!](#)]

[RePEc:boc:bocode:S388801](#) **STCSTAT: Stata module to generate evaluation of fit for Cox regression model** by William W. Gould [[Downloadable!](#)]

[RePEc:boc:bocode:S388202](#) **CNSRSIG: Stata module to evaluate validity of restrictions on a regression** by Christopher F Baum & Vince Wiggins [[Downloadable!](#)]

[RePEc:boc:bocode:S388201](#) **SKEWPLOT: Stata module to draw skewness plots** by Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S388102](#) **RENAMES: Stata module to rename variables** by Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S388101](#) **GPHUDAK: Stata module to estimate long memory in a timeseries** by Christopher F Baum & Vince Wiggins [[Downloadable!](#)]

[RePEc:boc:bocode:S388001](#) **ARCHLM: Stata module to calculate LM test for ARCH effects** by Christopher F Baum & Vince Wiggins [[Downloadable!](#)]

[RePEc:boc:bocode:S387901](#) **MNTHPLOT: Stata module for scatter plot for monthly data with repetition of data** by Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S387302](#) **BGTEST: Stata module to calculate Breusch-Godfrey test for serial correlation** by Christopher F Baum & Vince Wiggins [[Downloadable!](#)]

[RePEc:boc:bocode:S387301](#) **DURBINH: Stata module to calculate Durbin's h test for serial correlation** by Christopher F Baum & Vince Wiggins [[Downloadable!](#)]

[RePEc:boc:bocode:S386701](#) **TSMKTIM: Stata module to generate time-series calendar variable** by Christopher F Baum & Vince Wiggins [[Downloadable!](#)]

[RePEc:boc:bocode:S386602](#) **MOREOBS: Stata module to add observations to dataset** by Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S386601](#) **ARIMAFIT: Stata module to calculate AIC, SIC for ARIMA model** by Christopher F Baum [[Downloadable!](#)]

[RePEc:boc:bocode:S386404](#) **MARGFX: Stata module to compute mean marginal effects with nonlinear models** by Jonah B. Gelbach [[Downloadable!](#)]

[RePEc:boc:bocode:S386403](#) **AREG2: Stata module to estimate 2SLS regression absorbing one categorical factor** by Jonah B. Gelbach [[Downloadable!](#)]

[RePEc:boc:bocode:S386402](#) **DESMAT: Stata modules to generate a design matrix of dummy variables** by John Hendricks [[Downloadable!](#)]

[RePEc:boc:bocode:S386401](#) **_GNMISS: Stata module to count the number of missing values** by Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S386301](#) **_GRANKT-_GRANKF: Stata modules to generate ranks by "track" and "field" methods** by Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S384802](#) **EBB: Stata module to estimate the beta-binomial model** by Bradley Palmquist [[Downloadable!](#)]

[RePEc:boc:bocode:S384801](#) **MFILEGR: Stata module to view and save multiple graphs with filenames based on a numeric identifier** by Philip Ryan [[Downloadable!](#)]

[RePEc:boc:bocode:S382501](#) **_GPROD: Stata module to extend egen for product of observations** by Philip Ryan [[Downloadable!](#)]

[RePEc:boc:bocode:S382802](#) **STRPARSE: Stata module to parse string variables** by Michael Blasnik & Nicholas J. Cox [[Downloadable!](#)]

[RePEc:boc:bocode:S382801](#) **COMMA: Stata module to comma-format numbers** by Thomas Steichen [[Downloadable!](#)]

'Net-aware' Stata

With the advent of Version 6, Stata is 'net-aware': that is, the program may enquire via HTTP to determine whether there are updated elements of official Stata available. Likewise, Stata components associated with the Stata Technical Bulletin (STB) may be accessed via 'net' commands from the StataCorp web site.

'Net-awareness' also allows for any Stata user to share his or her Stata code with other users. The Stata documentation contains detailed instructions on creating your own web site of materials; all that is needed is access to a Web server, and the ability to place text files on that server.

'Net-aware' Stata

What is missing from this model?

Just as each department or individual can establish a web page for publications dissemination, they may establish a Stata web site from which Stata components may be downloaded. Stata will dutifully retrieve those materials, as long as the site's design follows a few modest requirements.

But how are you to find the materials if you don't know where to look?

SSC-IDEAS and 'net-aware' Stata

If a Stata component accessible to 'net-aware' Stata is described in a ReDIF-Software template and included in SSC-IDEAS, that component may be located via the IDEAS eXcite search facility by:

- author
- word in title
- keywords
- 'abstract': text description of functionality

SSC-IDEAS and 'net-aware' Stata

The Stata code itself (the '.ado') may be viewed in the web browser, as may the help file ('.hlp'). But if you're using 'net-aware' Stata, the best way to install the code in your copy of Stata is via the 'net install' command.

Additional utilities have been recently added to SSC-IDEAS to facilitate this process of:

- locating the component you need
- accessing the SSC-IDEAS archive
- installing the component

SSC-IDEAS 'archutil' package

'archutil' contains utilities for SSC-IDEAS archive access. The 'archlist' command produces a list of all Stata components on the archive, organized by first letter of package name, with a short title (that appearing on the web page listing).

'archlist *letter*' will provide a listing of those packages with names beginning with *letter*. Furthermore, the 'net install *package*' command may be given immediately following, to access the archive and install a specific package, or component, with only two simple commands.

SSC-IDEAS 'archutil' package

'archtype *file*' allows you to see the text of an .ado or .hlp file on the SSC-IDEAS archive without actually installing it. This mirrors the capability of viewing these files from the web browser accessing the SSC-IDEAS archive, but works within 'net-aware' Stata.

'archcopy *file*' will copy that file to the appropriate directory on your computer. This should not be used as an alternative to 'net install', but enhances the functionality of Stata's native 'copy' command by prespecifying the web site from which the file is to be copied.

By combining the metadata archive of Stata components produced by dozens of authors with the ‘net-aware’ facilities built in to the application, the whole is more than the sum of the parts.

The metadata archive allows the Stata user to do a ‘fuzzy search’ for available components without having to know their names, who wrote them, or when they were announced or updated on Statalist. Simply specifying keywords that appear in the package title or description in a SSC-IDEAS search will suffice.

The ‘net-aware’ facilities of Stata ensure that an entire SSC-IDEAS-accessible package will be properly installed, if that is possible, so that the application’s functionality is protected.

How can you participate?

You may make use of SSC-IDEAS to access the latest developments in user-written components for Stata and several other programming languages.

If you use Stata in your work and develop materials that you want to share with other Stata users, contribute them to the SSC-IDEAS archive, and note their availability on Statalist. (The same goes for users of other packages and languages).

Please contact me for any information about SSC-IDEAS at baum@bc.edu.

How can your institution participate?

You may establish a RePEc archive for your institution, which may contain one or more 'series' of materials: e.g. working papers, research reports, or software components. You need only make templates—ASCII text files describing your archive, series, and templates for each item included—and store them on a server accessible for anonymous FTP or web (HTTP) access. Your templates will be automatically checked for consistency with the ReDIF standard and included in the RePEc collection. Each series will be displayed within IDEAS, associated with your institution, and its contents will be searchable via the IDEAS search engine (as well as those of other RePEc services).

The IDEAS home page contains links to all the information needed to accomplish these tasks. If you already have these materials on the web and described on a web page, it is even simpler to construct the metadata templates describing each item.

Current RePEc information providers

In establishing a RePEc archive, you will join some of the world's most prestigious research institutions, including the National Bureau of Economic Research (NBER), the U.S. Federal Reserve System, CEPR, the Bank of England, Economics Working Paper Archive and many leading university departments of economics and research centers. A list of the nearly 1,000 working paper series in RePEc—containing metadata on over 58,000 working papers—is available from IDEAS. RePEc also contains metadata describing over 14,000 journal articles and nearly 500 software components. Over 17,000 of these items are currently downloadable in full-text form.

You are welcome to join the RePEc effort!

the RePEc team