

# Data Description

*Lars Vilhuber*

*2/21/2020*

## Contents

1000plus_file_count.txt . . . . .	1
aea_icpsr_all_files.csv . . . . .	2
aea_article_data.csv . . . . .	2
aea_issue_data.csv . . . . .	3
aea_jel_data.csv . . . . .	5
aea_icpsr_mapping.20191014.txt . . . . .	5
Further information . . . . .	6

The files in this directory constitute metadata from the AEA's migration of supplements from the AEA website (<https://www.aeaweb.org>) to the new AEA Data and Code Repository at openICPSR.

To migrate, a metadata file (`metadata.rdf`) and a default license (`LICENSE.txt`) was injected into the ZIP files, and a citation generated to link back to the original manuscript.

The resulting list of files are cataloged in `aea_icpsr_all_files.csv`, indexed by DOI.

Once the ingest was finalized, the technical team at openICPSR provided a mapping from the new DOI at openICPSR to the name of the datafile ingested (`aea_icpsr_mapping.20191014.txt`).

Certain supplements were too big, because they contained more than 1000 component files. A full list is not part of this metadata, but their file counts and overall sizes are listed in `1000plus_file_count.txt`.

To map back to the original articles, additional metadata regarding the articles was extracted from the AEA's internal database. Most of this data can be recovered from the CrossRef DOI registry, and all of this data is embedded in the AEA's article landing pages.

This document simply documents the file formats used by these files.

## 1000plus\_file\_count.txt

Supplements that were too big to be migrated in the first batch

- Tab-separated with column headers

```
##      filecount \tuncompressed(bytes) \tcompressed(bytes)
## 1:      201972      19998872239      7471194261
## 2:       41689      17890493201      9158004104
## 3:       21273      7355341527      2132558676
## 4:       16681      1170769060      264990707
## 5:       13257      22317865136      10120553308
## 6:        10136       39670512      5158729
```

- With alternate column names:

```
##      filecount      pkgsize      compressed
## Min.   : 1010      Min.   : 12460417      Min.   : 5158729
## 1st Qu.: 1188      1st Qu.: 87556930      1st Qu.: 41520555
## Median : 1634      Median : 470801893      Median : 109014215
## Mean   : 9788      Mean   : 3364826479      Mean   : 1111437669
```

```
## 3rd Qu.: 3451    3rd Qu.: 2757059250    3rd Qu.: 455990736
## Max.      :201972    Max.      :22317865136    Max.      :10120553308
```

The file has **39** rows.

## aea\_icpsr\_all\_files.csv

List of all files in the packages provided to openICPSR:

```
##      doi          file_path      file_size
## Length:102151    Length:102151    Min.      :0.000e+00
## Class :character  Class :character  1st Qu.:2.000e+03
## Mode  :character  Mode  :character  Median   :8.626e+03
##                                     Mean      :4.637e+06
##                                     3rd Qu.:4.967e+04
##                                     Max.      :2.130e+09
```

The file has **102151** rows.

- **doi**: DOI of the published article (links to other files, CrossRef). There are 2562 unique DOIs.
- **file\_path**: full file path and name of the file within the ZIP file
- **file\_size**: file size in bytes of the file

The administrative files occur in every package:

```
## .
## LICENSE.txt metadata.rdf
##      2562      2562
```

## aea\_article\_data.csv

Database of the articles and the associated packages sent to ICSPR.

```
##      issue_id          doi start_page end_page
## 1:      48 10.1257/0002828042002516      1169      1182
## 2:      49 10.1257/0002828043052196      1705      1716
## 3:      49 10.1257/0002828043052358      1693      1704
## 4:      50 10.1257/0002828053828446       425       436
## 5:      50 10.1257/0002828053828455       395       406
## 6:      50 10.1257/0002828053828509       450       460
```

```
##
## 1:                                     Saving, Risk Sharing, and Preference
## 2:                                     Progressive Taxation and
## 3:                                     The Effect of Health Risk on Housing Values: Evidence from a
## 4: The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macro
## 5:      Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the H
## 6:                                     Patent Citations and the Geography of Knowledge Spillovers:
```

```
##      icpsr_package_name
## 1: 0002828042002516_data.zip
## 2: 0002828043052196_data.zip
## 3: 0002828043052358_data.zip
## 4: 0002828053828446_data.zip
## 5: 0002828053828455_data.zip
## 6: 0002828053828509_data.zip
##                                     aea_url
## 1: https://www.aeaweb.org/aer/data/sept04_data_mazzocco.zip
```

```
## 2:      https://www.aeaweb.org/aer/data/dec04_data_sarte.zip
## 3:      https://www.aeaweb.org/aer/data/dec04_data_davis.zip
## 4: https://www.aeaweb.org/aer/data/mar05_data_gurkaynak.zip
## 5: https://www.aeaweb.org/aer/data/mar05_data_schlenker.zip
## 6: https://www.aeaweb.org/aer/data/mar05_data_thompson.zip

##      issue_id      doi      start_page      end_page
## Min.   : 48.0    Length:2552    Min.    :   1.0    Min.    :   4.0
## 1st Qu.:243.0    Class :character  1st Qu.: 118.0    1st Qu.:  47.0
## Median :354.0    Mode  :character  Median : 247.0    Median :  71.0
## Mean   :342.5                    Mean   : 745.1    Mean   : 263.9
## 3rd Qu.:452.0                    3rd Qu.:1046.2    3rd Qu.:  98.0
## Max.   :550.0                    Max.    :4231.0    Max.    :4204.0

##      title      icpsr_package_name      aea_url
## Length:2552    Length:2552      Length:2552
## Class :character Class :character  Class :character
## Mode  :character Mode  :character  Mode  :character
##
##
##
```

The file has **2552** rows.

- **issue\_id**: An internal sequential number for the issue of the article (links to **aea\_issue\_data**).
- **doi**: DOI of the published article (links to other files, CrossRef). There are 2552 unique DOIs.
- **start\_page**: Start page in the issue in which the article was printed.
- **end\_page**: End page in the issue in which the article was printed.
- **title**: Title of the article
- **aea\_url**: URL of the package on the AEA website. Can be used in combination with other files in this package to map pre-2019 URLs in references to the new openICPSR DOI.
- **icpsr\_package\_name**: name of the ZIP file as sent to ICPSR (not the same as originally presented on the AEA website). Links to **aea\_icpsr\_mapping.20191014.txt**.

## aea\_issue\_data.csv

Issue information

```
##      issue_id      journal journal_abbreviation volume
## 1:      1 Journal of Economic Perspectives      jep      16
## 2:      2 Journal of Economic Perspectives      jep      16
## 3:      3 Journal of Economic Perspectives      jep      16
## 4:      4 Journal of Economic Perspectives      jep      16
## 5:      5 Journal of Economic Perspectives      jep      17
## 6:      6 Journal of Economic Perspectives      jep      17

##      issue year month  cover_date
## 1:      1 2002      3 Winter 2002
## 2:      2 2002      6 Spring 2002
## 3:      3 2002      9 Summer 2002
## 4:      4 2002     12  Fall 2002
## 5:      1 2003      3 Winter 2003
## 6:      2 2003      6 Spring 2003

##      issue_id      journal      journal_abbreviation      volume
## Min.   :   1.0    Length:534    Length:534      Min.    :   1.00
## 1st Qu.:134.2    Class :character  Class :character  1st Qu.:   7.25
## Median :267.5    Mode  :character  Mode  :character  Median :  25.00
```

```
## Mean      :272.6                      Mean      : 42.14
## 3rd Qu.:400.8                      3rd Qu.: 92.00
## Max.      :565.0                      Max.      :109.00
##
##      issue          year          month          cover_date
## Length:534      Min.      :1987      Min.      : 1.000      Length:534
## Class :character 1st Qu.:2005      1st Qu.: 4.000      Class :character
## Mode  :character Median :2012      Median : 6.000      Mode  :character
##                      Mean      :2010      Mean      : 6.625
##                      3rd Qu.:2016      3rd Qu.: 9.000
##                      Max.      :2019      Max.      :12.000
##                      NA's      :1
```

The file has **534** rows and 14 unique issues.

- **issue\_id**: An internal sequential number for the issue of the article (links to **aea\_issue\_data**).
- **journal**: Journal name
- **journal\_abbreviation**: Abbreviation used for the journal.
- **volume**: Volume of the particular issue
- **issue**: Issue number of the particular issue
- **year**: Year of publication of the issue
- **month**: Month of publication of the issue
- **cover\_date**: The date on the cover of the issue

```
##
##                      AEA Papers and Proceedings
##                      2
## American Economic Journal: Applied Economics
##                      44
## American Economic Journal: Economic Policy
##                      41
## American Economic Journal: Macroeconomics
##                      42
## American Economic Journal: Microeconomics
##                      41
##                      American Economic Review
##                      150
## American Economic Review: Insights
##                      2
##                      Journal of Economic Literature
##                      83
##                      Journal of Economic Perspectives
##                      129
##
## aer  aeri  app  jel  jep  mac  mic  pandp  pol
## 150   2   44  83  129  42  41   2   41
```

### Note: missing issue information

Some articles do not have issues in the database. This affects 10 articles. The relevant information can be obtained from CrossRef:

```
# not executed
aea.complete.B <- rcrossref::cr_works(dois=aea.not_complete$doi)
```

## aea\_jel\_data.csv

JEL codes associated with articles

```
##                               doi jel_code
## 1: 10.1257/0002828042002516      D14
## 2: 10.1257/0002828042002516      D81
## 3: 10.1257/0002828042002516      D91
## 4: 10.1257/0002828043052196      E62
## 5: 10.1257/0002828043052196      H21
## 6: 10.1257/0002828043052196      O41
##                               jel_description
## 1:                               Household Saving; Personal Finance
## 2:                               Criteria for Decision-Making under Risk and Uncertainty
## 3: Intertemporal Household Choice; Life Cycle Models and Saving
## 4:                               Fiscal Policy
## 5:                               Taxation and Subsidies: Efficiency; Optimal Taxation
## 6:                               One, Two, and Multisector Growth Models

##      doi      jel_code      jel_description
## Length:11395      Length:11395      Length:11395
## Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character
```

The file has **11395** rows and 2552 unique DOIs.

- doi: DOI of the published article (links to other files, CrossRef). There are 2552 unique DOIs.
- jel\_code: 3-character JEL code, see <https://www.aeaweb.org/econlit/jelCodes.php?view=jel>. There are 564 unique JEL codes on this file.
- jel\_description: Verbose description of the JEL code.

## aea\_icpsr\_mapping.20191014.txt

Mapping of ICPSR packages to ICPSR-based DOI

```
##      http://doi.org/10.3886/E112305V1
## 1: http://doi.org/10.3886/E112306V1
## 2: http://doi.org/10.3886/E112307V1
## 3: http://doi.org/10.3886/E112308V1
## 4: http://doi.org/10.3886/E112309V1
## 5: http://doi.org/10.3886/E112310V1
## 6: http://doi.org/10.3886/E112311V1
##      Original Zip: 0002828042002516_data.zip
## 1: Original Zip: 0002828043052196_data.zip
## 2: Original Zip: 0002828043052358_data.zip
## 3: Original Zip: 0002828053828446_data.zip
## 4: Original Zip: 0002828053828455_data.zip
## 5: Original Zip: 0002828053828509_data.zip
## 6: Original Zip: 0002828053828635_data.zip
```

The file needs a bit of cleaning:

```
aea.mapping <- fread::fread(file.path(acquired,"aea_icpsr_mapping.20191014.txt"),
                             col.names = c("DOI","Zipfile")) %>%
  mutate(icpsr_package_name = str_remove(Zipfile,"Original Zip: " ),
```

```
icpsr_doi = str_remove(DOI,"http://doi.org/") %>%  
select(-Zipfile,-DOI)
```

The file has **534** rows and 14 unique issues.

- `icpsr_package_name`: name of the ZIP file as sent to ICPSR. Links to `aea_icpsr_all_files_all`.
- `icpsr_doi`: newly assigned DOI of the ingested supplement.

## Further information

These files are provided without any warranty.