

# Migrating historical AEA supplements - DRAFT



**AMERICAN  
ECONOMIC  
ASSOCIATION**

**OPEN ICPSR**

Since July 16, 2019, the American Economic Association has used the [AEA Data and Code Repository](#) at [openICPSR](#) as the default archive for its supplements. This archive serves a dual purpose: to share data with the AEA Data Editor prior to being published, and as a publication outlet for supplements to articles in AEA journals.

At the time, the AEA also announced that it would migrate the historical supplements, hitherto stored as ZIP files on the [AEA website](#), into the AEA Data and Code Repository.

On Oct 1, 2019, openICPSR had 867 deposits, which covered 94 deposits in the [DataLumos](#) archive, 46 in the [AERA archive](#), and 13 in the [PSID](#) archive. The **AEA Data and Code Repository** contained at the time 93 deposits, of which 5 were public, the others awaiting publication of the associated article.

Between Oct 11 and Oct 13, 2019, the staff at openICPSR ingested 2,552 historical supplements, increasing the size of the openICPSR repository **by a factor of 3**, to 3,461. This was only the first part of the migration, as there are about 1,000 more archives that need to be migrated.

## Characteristics of AEA supplement data

We can describe this subset of the historical supplements in a variety of ways.

### Number of files per supplement

doi	size	count
10.1257/pandp.20181045	32782817227	795
10.1257/pol.20150168	27251086584	691
10.1257/app.20170080	19924915397	236
10.1257/aer.20131496	15688843069	186
10.1257/app.20160510	14419727617	465
10.1257/app.6.3.206	12190878093	45
10.1257/aer.102.2.994	10789428265	36
10.1257/aer.20121662	10334181612	622
10.1257/mic.20130164	8042445763	20
10.1257/aer.20141374	6804364827	140

The 2,552 supplements contain a total of 94,465 files - programs, documents, datasets. The largest supplement within this group in terms of file count has 795 files, summing to 30.5 Gb ([Armour, Button, and Hollands, 2018](#)). Note however that among the remaining non-migrated supplements are very large packages: the largest we have identified has 201,972 files.

### Distribution by journal

First, we can look at the size of the supplements globally by journal. The following table shows cumulative and median size and number of files.

<b>Journal</b>	<b>Median Size (Mb)</b>	<b>Cumulative Size (Mb)</b>	<b>Median files</b>	<b>Total files</b>
American Economic Review	1.3	165,547.8	13	50,163
American Economic Journal: Applied Economics	3.1	101,969.8	10	9,979
American Economic Journal: Economic Policy	4.2	98,574.0	12	12,585
AEA Papers and Proceedings	0.9	44,497.3	5	2,589
American Economic Journal: Macroeconomics	1.4	14,943.2	21	10,231
American Economic Journal: Microeconomics	1.1	12,898.8	10	5,423
Journal of Economic Perspectives	0.8	12,755.7	7	2,833
Journal of Economic Literature	17.8	431.8	67	559
American Economic Review: Insights	8.2	74.7	5	103

### Distribution across JEL codes

The top 10 JEL codes associated with supplements are:

<b>Number of packages</b>	<b>Pct</b>	<b>JEL</b>	<b>Description</b>
263	10.31	E32	Business Fluctuations; Cycles
245	9.60	J24	Human Capital; Skills; Occupational Choice; Labor Productivity
217	8.50	O15	Economic Development: Human Resources; Human Development; Income Distribution; Migration
214	8.39	D12	Consumer Economics: Empirical Analysis
207	8.11	J31	Wage Level and Structure; Wage Differentials

*Note:*

\*A supplement can be associated with multiple JEL codes.\*

Number of packages	Pct	JEL	Description
191	7.48	J16	Economics of Gender; Non-labor Discrimination
183	7.17	J13	Fertility; Family Planning; Child Care; Children; Youth
176	6.90	I21	Analysis of Education
162	6.35	D72	Political Processes: Rent-seeking, Lobbying, Elections, Legislatures, and Voting Behavior
157	6.15	D83	Search; Learning; Information and Knowledge; Communication; Belief

*Note:*

\*A supplement can be associated with multiple JEL codes.\*

## Metadata

When planning the migration, the preservation of existing metadata - the information about the data and code - was important. The AEA Data Editor worked with the openICPSR staff to enhance the data infrastructure, adding the capability to store and display JEL codes in addition to subject terms. Going forward, in addition to adding the JEL codes that also describe the linked article, authors can add metadata such as *geographic coverage*, *funding sources*, *time periods*, *geographic units* as well as *units of observation*, greatly enhancing the ability of researchers to find data through the openICPSR search interface.

Two important caveats apply, however. First, none of the additional metadata exists for the historical archives. Second, the openICPSR search interface does not currently expose the ability to search by those fields. An enhancement is planned for implementation before the end of 2019.

## References