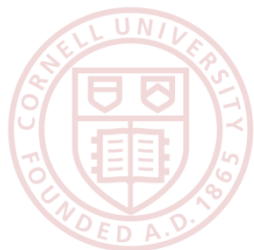




Consuming Transparency: View from the Journal Side

Lars Vilhuber
Cornell University

The opinions expressed in this presentation are solely the authors, and do not represent the views of the U.S. Census Bureau, the American Economic Association, the National Academies of Sciences, or any of the funding agencies.



Consuming Transparency:

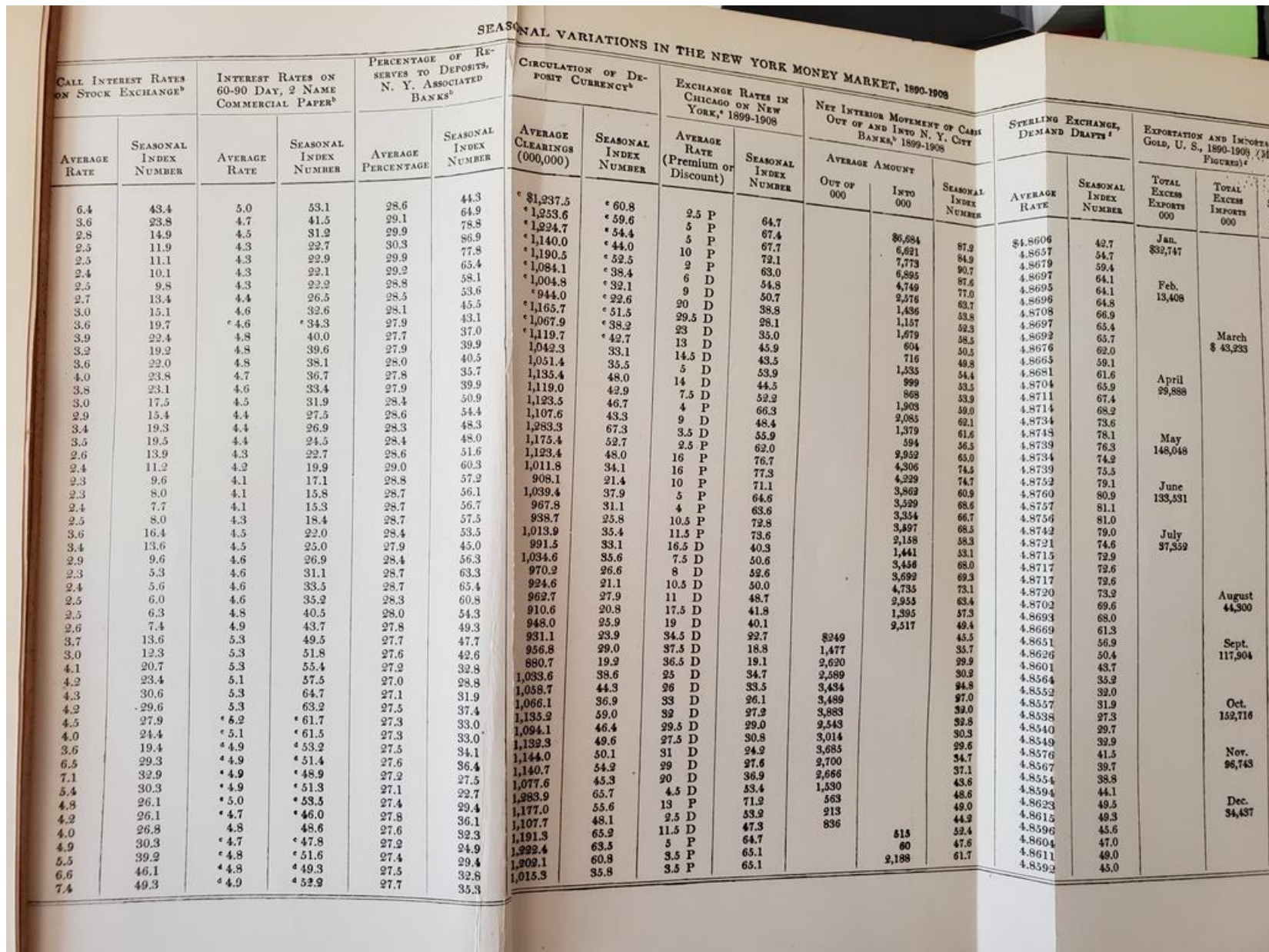
But they do rely on my experience as
Data Editor for the American Economic Association

Lars Vilhuber
Cornell University



Efficiency of scholarly discourse?

- Early publications (20th century) contained **tables of data**, and the **math** was simple (maybe)
 - **Data** became electronic, was no longer **included** or **cited**
 - **Math** was transcribed to **code**, and was no longer **included**



From @sdellavi
AER 1911



Efficiency of scholarly discourse!

**Modern publications need
the same transparency and completeness
as in the “old days”
to facilitate replicability**

Issues



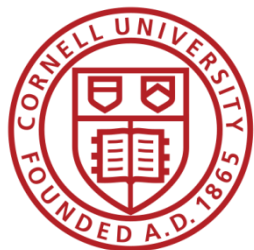
Economics makes wide use of public-use data

- **Macrodata:**

“We use data downloaded from the Bureau of Economic Analysis...”

- **Microdata:**

“... this paper uses data from the Current Population Survey...”



This should be easy!



Problems with Data Citations

Many datasets

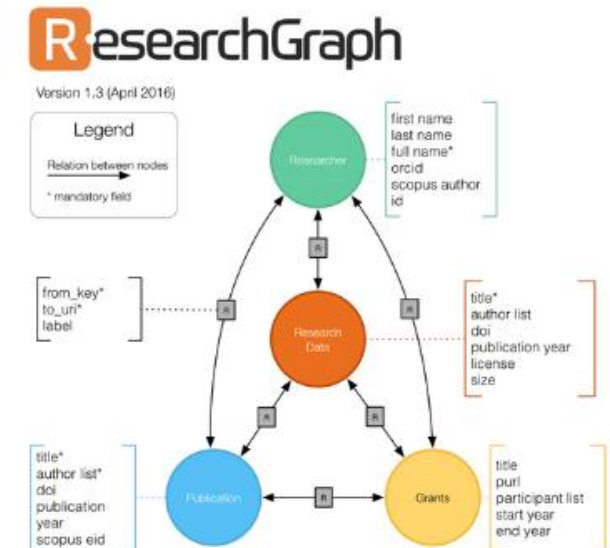
- Are imperfectly described
 - Very few data citations
- Are badly documented
- Have no (permanent) location defined
 - Even for data from high-profile organizations!
- All of the above

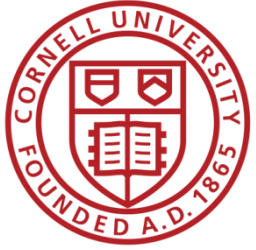
Evolving Journal and Data Infrastructure



Emerging citation infrastructure

- Currently:
 - Manual linking of data and publications
 - Few citations of data
- True for most journals, and most data archives
 - ICPSR (manual linking to articles)
 - RePEc (no linkage possible)
- Infrastructure starting to emerge
 - If article cites data (DOI!)
 - If archive and/or journal leverages infrastructure





Why do journals like “supplemental ZIP files” and affiliated repositories?

- They can ensure **longevity/ persistence**
- They can ensure **access**
- They can ensure **availability**



What are the characteristics of prominent data archives?

- They DO ensure **longevity/ persistence**
- They DO ensure **access**
- They DO ensure **availability**



Evolving Journal and Data Infrastructure

- More self-deposit repositories in the social sciences
 - Dataverse
 - Figshare
 - openICPSR
 - Zenodo
 - Qualitative Data Repository (QDR)
 - Others...



Evolving Journal and Data Infrastructure

- More ~~self-deposit~~ repositories in the social sciences

- Dataverse
- Figshare
- openICPSR
- Zenodo

- **CASD (France)**
- **IAB (Germany)**
- **Norway**
- **US Federal Statistical RDC**
-

- Qualitative Data Repository (QDR)
- Others...



Evolving Journal and Data Infrastructure

About 40% of data in economics journals is from restricted-access data sources



Evolving Journal and Data Infrastructure

**Journals ask Authors:
Describe your Sources!**

(cite them!)



The Future of Research Communications and e-Scholarship

Search

English

ABOUT ▾ COMMUNITY ▾ CODE OF CONDUCT GROUPS RESOURCES ▾ NEWS + BLOGS ▾ EVENTS ▾ PUBLIC

perceived criteria of importance.

1. Importance

Data should be considered legitimate, citable products of research. Data should be accorded the same importance in the scholarly record as citable research objects, such as publications[1].

DC¹
Data Citation Principles

2. Credit and Attribution

Data citations should facilitate giving scholars credit and recognition and

1 Bureau of Labor Statistics. 2000–2010. “Current Employment Statistics: Colorado, Total Nonfarm, Seasonally adjusted - SMS080000000000000001.” United States Department of Labor. <http://data.bls.gov/cgi-bin/surveymost?sm+08> (accessed February 9, 2011).

in scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited[3].

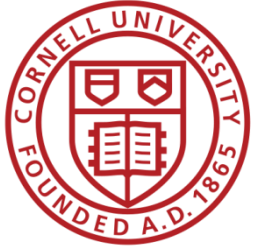
4. Unique Identification

A data citation should include a persistent method for identification that is actionable, globally unique, and widely used by a community[4].

5. Access

Data citations should facilitate access to the data themselves and to such related data, documentation, code, and other materials as are necessary for

Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014
[\[https://www.force11.org/group/joint-declaration-data-citation-principles-final\]](https://www.force11.org/group/joint-declaration-data-citation-principles-final).



Evolving Journal and Data Infrastructure

Data Citations are not
enough!



Why are data citations not enough?

- They tell you “where”
- But most do not
 - “who can access”
 - “for how long”
 - “under what conditions”
- (Though in theory, these are covered by the Data Citation Principles)



Verifying Reliability of Data Sources

- Not every restricted-access repository is created equal
 - The Second Bank of Third City credit card data is not a data/code repository
 - Is the School Board of Third City a reliable repository?
 - Is the JPMC Institute a reliable repository?
 - Is the US Census Bureau a reliable repository?
 - **Are any restricted-access repositories reliable archives?**



Data Availability Statements (DAS)

- A statement about **where data** supporting the results reported in a published article can be found
 - including unique identifiers linking to publicly archived datasets analyzed or generated during the study.
- DASs can **increase transparency** by providing a reason why data cannot be made (immediately) available
 - **need for registration**, ethical or legal restrictions, or because of an embargo period



Data Availability Statements

- A statement about **how long** data will be **available** (policy)
 - DOI assignments implies **long-term curation**
 - But long-term curation **does not require DOI!**
- A statement about **usage rights**
 - Not every dataset is in the public domain
 - Not everybody knows that U.S. Government data are usually in the public domain



Current efforts at the AEA

- **Pre-emptively improve code archives**
 - By conducting reproducibility checks when we can
 - By working with groups that conduct reproducibility checks when we cannot
- **Better archives**
 - Greater transparency of the code and data archives
- **Better provenance tracking**
 - Leave code where it is when appropriate
 - Leave data where it is almost always
 - Display that information



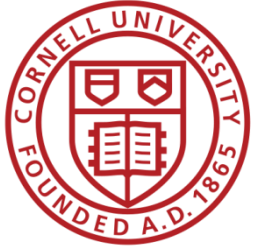
AEA “Data Availability Policy” (2019)

- It is the policy of the American Economic Association to publish papers only if the data used in the analysis are **clearly and precisely documented and access to the data and code is clearly and precisely documented and is non-exclusive to the authors.**
- Authors of accepted papers that contain empirical work, simulations, or experimental work must **provide, prior to acceptance**, the data, programs, and other details of the computations **sufficient to permit replication**, as well as **information about access to data and programs.**



Evolving Journal and Data Infrastructure

Authors struggle with
this!



Evolving Journal and Data Infrastructure

Data Publishers struggle
with this!

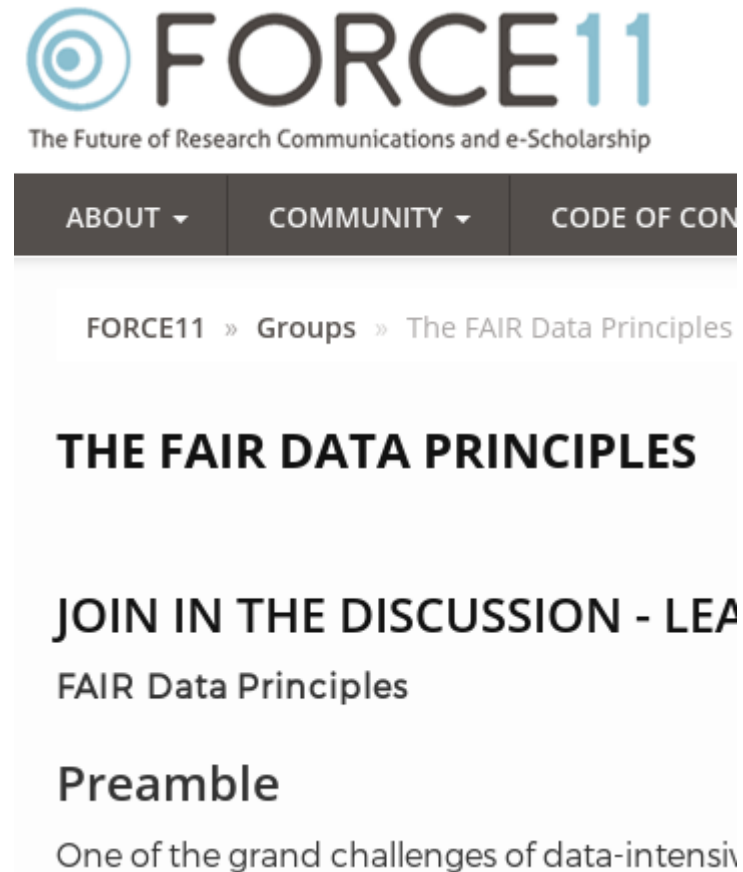
Solutions



Action: Data citations and metadata

What is **FAIR**?

- **F**indable,
- **A**ccessible,
- **I**nteroperable, and
- **R**e-usable





FAIR data principles rely on metadata

— Scope of Project

Subject Terms ?

Do not copy/paste multiple terms into this field. Terms must be entered individually.

× Russia × Industry × Factories × Russian Empire × Corporations

JEL Classification ?

× L20 General × N63 Europe: Pre-1913 × O43 Institutions and Growth

Manuscript Number ?

AER-2015-1656.R3 [✎ edit](#) [✕ remove](#)

Geographic Coverage ? [+ add value](#)

European Russia (Russian Empire) [✎ edit](#) [✕ remove](#)

Time Period(s) ? [+ add value](#)

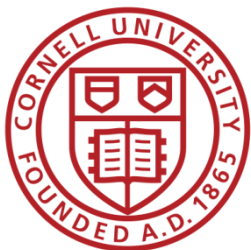
1894 – 1908 (Three years: 1894, 1900, and 1908) [✎ edit](#) [✕ remove](#)

Collection Date(s) ? [+ add value](#)

Universe ?

Manufacturing establishments in the European part of the Russian Empire. [✎ edit](#) [✕ remove](#)

Data Type(s) ?



FAIR data principles rely on metadata

— Scope of Project

Subject Terms ?

Do not copy/paste multiple terms into this field. Terms must be entered individually.

× Russia × Industry × Factories × Russian Empire × Corporations

```
<meta name="DC.identifier" content="10.3886/E110681V1" />
<meta name="DC.title" content="Imperial Russian Factory Database, 1894-1908" />

<meta name="DC.creator" content="Amanda Gregg, Middlebury College" />

<meta name="DC.publisher" content="Inter-university Consortium for Political and Social Research (ICPSR)" />
<meta name="DC.date" content="2020-01-29" />
<meta name="DC.type" content="Dataset" />
```

Geographic Coverage ? + add value

European Russia (Russian Empire) ✎ edit ✕ remove

Time Period(s) ? + add value

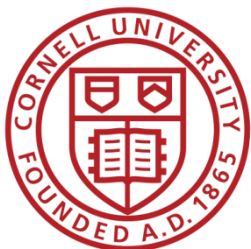
1894 – 1908 (Three years: 1894, 1900, and 1908) ✎ edit ✕ remove

Collection Date(s) ? + add value

Universe ?

Manufacturing establishments in the European part of the Russian Empire. ✎ edit ✕ remove

Data Type(s) ?



FAIR data principles rely on metadata

— Scope of Project

Subject Terms ?

Do not copy/paste multiple terms into this field. Terms must be entered individually.

```
<script type="application/ld+json">
```

```
{ "name": "Imperial Russian Factory Database, 1894-1908", "identifier": "http://doi.org/10.3886/E110681V1", "description": "This database dig  
manufacturing censuses. For each factory, the database includes industry, province, enterprise form, total workers, total revenue, and identifiers tha  
. 1908 years also include information on the factory's total machine power. The dataset was constructed to study why some Russian firms chose to become c  
consuming concession system. Note that the final analysis files exclude factories located outside of European Russia and, in the main data files, facto  
ax. &nbsp;", "url": "http://doi.org/10.3886/E110681V1", "version": "V1", "keywords": ["Russia", "Industry", "Factories", "Russian Empire", "Corporations"], "spati  
mpire)", "temporalCoverage": ["1894-01-01--1908-12-31 (Three years: 1894, 1900, and 1908)", "creator": [{"name": "Amanda Gregg", "affiliation": ["Middlebur  
{"name": "openICPSR Self-Deposit Archive", "url": "http://www.openicpsr.org/", "@type": "DataCatalog"}, {"funder": [{"name": "Economic History Association", "@ty  
Directorate for Social, Behavioral and Economic Sciences", "@type": "Organization"}, {"name": "Yale Economic Growth Center", "@type": "Organization"}, {"name  
Fund", "@type": "Organization"}, {"name": "Yale Program in Economic History", "@type": "Organization"}, {"name": "Yale MacMillan Center", "@type": "Organization"}  
{"fileFormat": "stata", "contentURL": "https://www.openicpsr.org/openicpsr/project/110681/version/V1/download/terms?path=/openicpsr/110681/fcr:versions/V  
stata", "encodingFormat": "application/zip"}, {"fileFormat": "stata", "contentURL": "https://www.openicpsr.org/openicpsr/project/110681/version/V1/download/t  
V1/AG_Corp_Prod_Database.dta&type=application/x-stata", "encodingFormat": "application/zip"}, {"fileFormat": "stata", "contentURL": "https://www.openicpsr.c  
terms?path=/openicpsr/110681/fcr:versions/V1/AG_Corp_RuscorpMasterFile_Cleaned.dta&type=application/x-stata", "encodingFormat": "application/zip"}, {"fil  
openicpsr/project/110681/version/V1/download/terms?path=/openicpsr/110681/fcr:versions/V1/AG_Corp_Prod_Database_withAksiz.dta&type=application/x-sta  
{"fileFormat": "stata", "contentURL": "https://www.openicpsr.org/openicpsr/project/110681/version/V1/download/terms?path=/openicpsr/110681/fcr:versions/V1  
stata", "encodingFormat": "application/zip"}], "license": "https://creativecommons.org/licenses/by/4.0", "@context": "http://schema.org", "@type": "Dataset" }
```

```
</script>
```

1894 – 1908 (Three years: 1894, 1900, and 1908) [✎ edit](#) [✖ remove](#)

Collection Date(s) ? [+ add value](#)

Universe ?

Manufacturing establishments in the European part of the Russian Empire. [✎ edit](#) [✖ remove](#)

Data Type(s) ?



... and findability relies on metadata

Google

Imperial Russian Factory



Updated Date

Download Format

Usage Rights

Free

2 datasets found



Imperial Russian Factory
Database, 1894-1908

www.openicpsr.org
www.da-ra.de

Updated Jan 29, 2020



Data from: Антиосманские
выступления болгар и русско-
турецкие войны второй...

explore.openaire.eu

Updated 24 нояб. 2015 г.



Not seeing a result you expected?
[Learn](#) how you can add new
datasets to our index.



Middlebury

Imperial Russian Factory Database, 1894-1908

[Explore at openICPSR Self-Deposit Archive](#)

[Explore at www.da-ra.de](http://www.da-ra.de)

Unique identifier

<https://doi.org/10.3886/E110681V1>

Dataset updated Jan 29, 2020

Dataset provided by

[Middlebury College](#)

Authors

Amanda Gregg

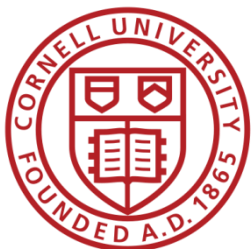
License

[Attribution 4.0 \(CC BY 4.0\)](#)

License information was derived automatically

Area covered

European Russia (Russian Empire)

[Find Data](#) / [Imperial Russian Factory Database, 1894-1908](#)

Imperial Russian Factory Database, 1894-1908

Principal Investigator(s): Amanda Gregg, Middlebury College

Version: V1



AMERICAN
ECONOMIC
ASSOCIATION

Name	File Type		Last Modified
1894MicroData.xlsx	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	4.5 MB	08/08/2019 11:01:AM

Project Citation:

Gregg, Amanda. Imperial Russian Factory Database, 1894-1908. Nashville, TN: American Economic Association [publisher], 2020. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-01-29. <https://doi.org/10.3886/E110681V1>

AG_Corp_CleaningandDatabaseCompiler.do	text/x-stata-syntax	23.4 KB	08/08/2019 11:02:AM
--	---------------------	---------	---------------------

Related Publications

The following publications are supplemented by the data in this project.

- Gregg, Amanda. "Factory Productivity and the Concession System of Incorporation in Late Imperial Russia, 1894-1908." *American Economic Review* 110, no. 2 (February 2020): 401-27. <https://doi.org/10.1257/aer.20151656>.



Behind this is metadata

- DataCite (DOI registrar for data)
- For instance,
 - <https://api.datacite.org/doi/10.3886/E110681V1>
- Screenshot of DataCite
- Screenshot of DataCite XML
- Screenshot of DataCite XML extract of rel link



Behind this is metadata

- DataCite (DOI)
- For instance

▼ 0:

relationType: "IsSupplementedBy"

relatedIdentifier: "10.1257/aer.20151656"

relatedIdentifierType: "DOI"

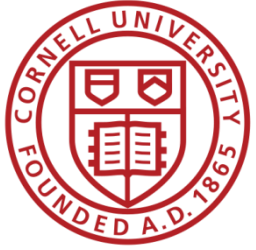
```
▼ data:
  id: "10.3886/e110681v1"
  type: "dois"
  ▼ attributes:
    doi: "10.3886/e110681v1"
    prefix: "10.3886"
    suffix: "e110681v1"
  ▼ identifiers:
    ▼ 0:
      identifier: "https://doi.org/10.3886/e110681v1"
```

```
0: "Middlebury College"
  nameIdentifiers: []
  ▼ titles:
    ▼ 0:
      lang: "en"
      title: "Imperial Russian Factory Database, 1894-1908"
  ▼ publisher: "ICPSR - Interuniversity Consortium for Political and Social Research"
    container: {}
    publicationYear: 2020
    subjects: []
    contributors: []
```



Metadata encodes linkages (Dublin Core)

- ****cites this deposit****
is cited by this deposit
- ****is supplemented by this deposit****
is a supplement to this deposit
- is referenced by this deposit
references this deposit
- is previous version of this deposit
is new version of this deposit
- documents this deposit
is documented by this deposit
- has this deposit as part
is part of this deposit



Linkage allows to demonstrate value

- Who uses your data?
- Who relies on your data
 - Directly
(original manuscript)
 - Indirectly
(citations of that manuscript)



How to leverage this?

- **Requires authors to cite** the data in the first place
- **Works easily with DOI** (registries need to know about the data object)
- **Might work** (in the future?) with Google Scholar, Web of Science

Examples


(good and bad)







Example 1: OECD

URL does not always change!

← → ↺ 🏠 <https://stats.oecd.org/Index.aspx>

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT  **OECD.Stat**

1. Gross domestic product (GDP) ⓘ

 Customise  Export  Draw chart  My Queries

Please refer to the dataset [Gross domestic product \(GDP\), 2019 archive](#) to access longer time series based

Please note that OECD reference year from 2010 to 2015 changed on Tuesday 3rd of December, 2019.

→ **Transaction**

■ B1_GA: Gross domestic product (output approach)

B1_GA: Gross domestic product (output approach)

■ B1G_P119: Gross value added at basic prices, excluding FISIM

B1G_P119: Gross value added at basic prices, excluding FISIM

■ B1G: Gross value added at basic prices, total activity

B1G: Gross value added at basic prices, total activity

■ B1GVA: Agriculture, forestry and fishing (ISIC rev4)

B1GVA: Agriculture, forestry and fishing (ISIC rev4)

■ B1GVB_E: Industry, including energy (ISIC rev4)

B1GVB_E: Industry, including energy (ISIC rev4)

B1GVC: of which energy (ISIC rev4)

B1GVF: Construction (ISIC rev4)

B1GVG_I: Distributive trade, repairs, transport, (ISIC rev4)

B1GVJ: Information and communication (ISIC rev4)

B1GVK: Financial and insurance activities (ISIC rev4)

B1GVL: Real estate activities (ISIC rev4)

B1GVM_N: Prof., scientific, techn., admin., sup (ISIC rev4)



Example 1: OECD

URL does not always
change!
(but sometimes it does)

← → ↺ 🏠 <https://stats.oecd.org/Index.aspx?DatasetCode=STLABOUR>

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT **OECD.Stat**

Data by theme Popular queries

in Themes >> Reset

our

labour

Earnings

Employment Protection

Labour Force Statistics

- Annual Labour Force Statistics
- LFS by sex and age

Short-Term Statistics

- Registered Unemployed and Job Vacancies

Short-Term Labour Market Statistics

- Short-Term Labour Market Statistics - Employment Rates
- Active Population
- Activity Rates
- Employed Population
- Employment - by economic

Short-Term Labour Market Statistics

Customise Export Draw chart My Queries

→ Subject	Employment rate, Aged 15-64, All persons
→ Measure	Level, rate or quantity series, s.a.
→ Frequency	Quarterly
Unit	
→ Time	Q4-2016 Q1-2017 Q2-2017
→ Country	
Australia	72.3 72.4 72.8
Austria	71.7 71.8 72.2
Belgium	63.3 62.4 62.9
Canada	72.9 73.3 73.4
Chile	62.4 62.4 62.6
Czech Republic	72.7 73.1 73.3
Denmark	rev4) 72.0 72.0 72.0



Example 1: OECD

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT **OECD.Stat**

← → ↺ 🏠 <https://stats.oecd.org/Index.aspx?DatasetCode=STLABOUR>

Data by theme Popular queries

Find in Themes >> Reset

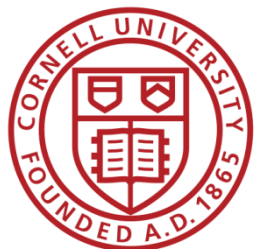
our
labour
Earnings
Employment Protection
Labour Force Statistics
+ Annual Labour Force Statistics
+ LFS by sex and age
■ Short-Term Statistics
+ Registered Unemployed and Job Vacancies
■ Short-Term Labour Market Statistics
■ Short-Term Labour Market Statistics - Employment Rates
■ Active Population
■ Activity Rates
■ Employed Population
■ Employment - by economic activity
■ Employment Rates by Age Group

Short-Term Labour Market Statistics : Unemployed Population

Customise Export Draw chart My Queries

	→ Subject	Unemployed population, Aged 15 and over, All person			
	→ Measure	Level, rate or quantity series, s.a.			
	→ Frequency	Quarterly			
	Unit				
	→ Time	Q2-2017	Q3-2017	Q4-2017	Q
→ Country					
Australia		721	718	717	
Austria		249	245	242	
Belgium		363	355	319	
Canada		1 275	1 215	1 185	
Chile		595	581	609	
Czech Republic		164	147	132	
Denmark		174	177	160	
Estonia		48	38	38	
Finland		rev4)	227	221	222

URL does not always change!
(and then it doesn't...)




CITATION AND USE OF THE IPUMS-CPS

Publications and research reports based on the IPUMS-CPS database must cite it appropriately. The citation should include the following:

Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles and J. Robert Warren. Integrated Public Use Microdata Series, Current Population Survey: Version 6.0 [dataset]. Minneapolis, MN: IPUMS, 2018. <https://doi.org/10.18128/D030.V6.0>

For policy briefs or articles in the popular press that use the IPUMS-CPS database, we recommend that you cite the use of IPUMS data as follows:

IPUMS-CPS, University of Minnesota, www.ipums.org.

The licensing agreement for use of IPUMS-CPS data requires that users supply us with the title and full citation for any publications, research reports, or educational materials making use of the data or documentation. Please add your citation to the [IPUMS bibliography](#) .



Example 3: FRED (St. Louis Fed)

FRED
ECONOMIC DATA | ST. LOUIS FED

ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

MY ACCOUNT

Search FRED

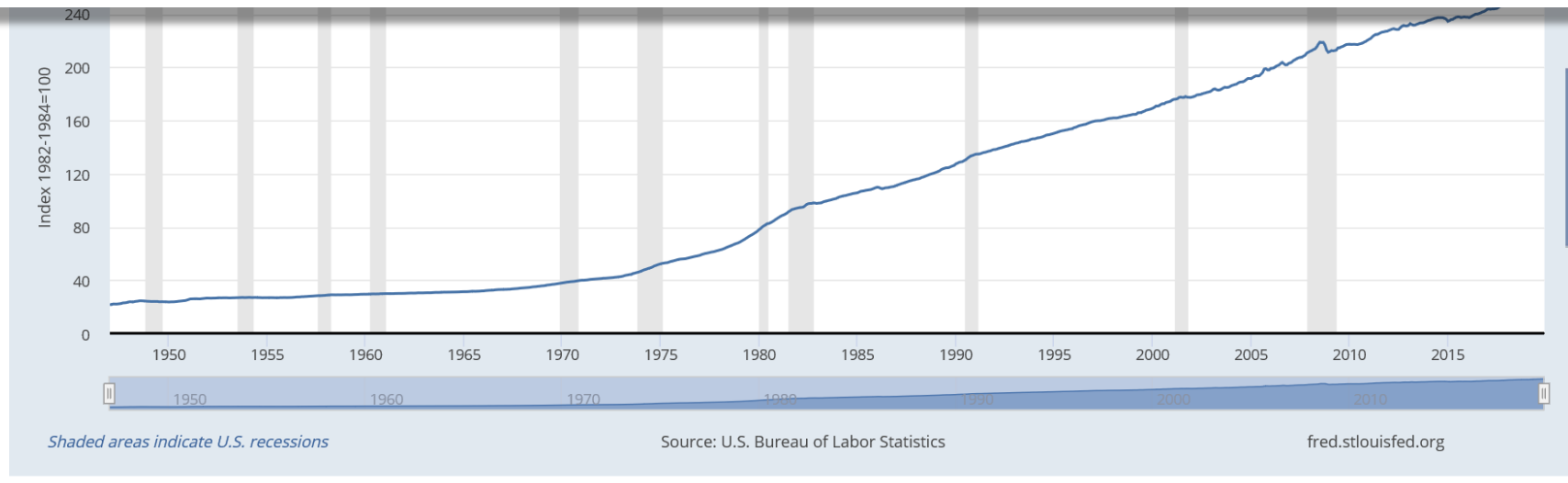
FRED® Economic Data Information Services Publications Working Papers Economists About

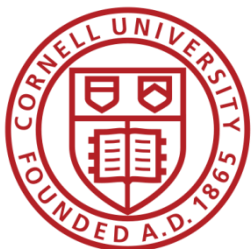
St. Louis Fed Home

Categories > Prices > Consumer Price Indexes (CPI and PCE)

Suggested Citation:

U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCSL], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/CPIAUCSL>, February 3, 2020.





Example 4: German Restricted-access



RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

[Home](#) | [Newsletter](#) | [Jobs](#) | [Contact](#) | [Data Privacy](#) | [Imprint](#)



Data Version	DOI (Link to Description of Data Version)	Availability (yyyy-mm-dd)
BHP 7518 v1 (current)	10.5164/IAB.BHP7518.de.en.v1	2020-01-13
BHP 7517 v1	10.5164/IAB.BHP7517.de.en.v1	2018-12-12
BHP 7516 v1	10.5164/IAB.BHP7516.de.en.v1	2018-04-11

External data

[Data Archive](#)

[Data Access](#)

[Campus Files](#)

[Publications](#)

[Events](#)

[Projects of FDZ users](#)

[FDZ Projects](#)

[Complaint point of the
RatSWD](#)

[Figures of the FDZ](#)

employees, both in total and broken down by gender, age, occupational status, qualification and nationality. Means and medians of wages for full-time employees are given, too. Additional datasets providing information about (gross) worker flows and about foundations and closures of establishments are available on request.

Data Versions

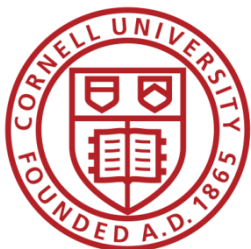
Old versions are only available for replication studies and only in justified exceptional cases for new Projects.

Data Version	DOI (Link to Description of Data Version)	Availability (yyyy-mm-dd)
--------------	---	---------------------------

BHP 7518 v1 (current)

[10.5164/IAB.BHP7518.de.en.v1](#)

2020-01-13



Example 4: German Restricted-access



RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

[Home](#) | [Newsletter](#) | [Jobs](#) | [Contact](#) | [Data Privacy](#) | [Imprint](#)



Data Version	DOI (Link to Description of Data Version)	Availability (yyyy-mm-dd)
BHP 7518 v1 (current)	10.5164/IAB.BHP7518.de.en.v1	2020-01-13
BHP 7517 v1	10.5164/IAB.BHP7517.de.en.v1	2018-12-12
BHP 7516 v1	10.5164/IAB.BHP7516.de.en.v1	2018-04-11

External data

[Data Archive](#)

[Data Access](#)

[Campus Files](#)

[Publications](#)

[Events](#)

[Projects of FDZ users](#)

[FDZ Projects](#)

[Complaint point of the
RatSWD](#)

[Figures of the FDZ](#)

employees, both in total and broken down by gender, age, occupational status, qualification and nationality. Means and medians of wages for full-time employees are given, too. Additional datasets providing information about (gross) worker flows and about foundations and closures of establishments are available on request.

Data Versions

Old versions are only available for replication studies and only in justified exceptional cases for new Projects.

Data Version	DOI (Link to Description of Data Version)	Availability (yyyy-mm-dd)
--------------	---	---------------------------

BHP 7518 v1 (current)

[10.5164/IAB.BHP7518.de.en.v1](#)

2020-01-13



Example 4: German Restricted-access

Establishment History Panel (BHP) – Version 7518 v1

DOI: 10.5164/IAB.BHP7518.de.en.v1

Summary

Data source:

Data Access



The IAB Establishment Panel is available via the following ways of access:

- On-site use at the FDZ. Further information on Applying for [on-site use](#).
- Remote data Access. Further information on Applying for [remote data access](#).

nationality. Means and medians of wages for full-time employees are given, too. Additional datasets providing information about (gross) worker flows and about foundations and closures of establishments are available on request.

Dataset Descriptions and Frequencies

German

- DOI: [10.5164/IAB.FDZD.2001.de.v1](#)
-  [FDZ-Datenreport 01/2020](#)
-  [Fallzahlen und Labels](#)

English

- DOI: [10.5164/IAB.FDZD.2001.en.v1](#)

Back to journals



American Economic Review 2020, 110(2): 401–427
<https://doi.org/10.1257/aer.20151656>

Factory Productivity and the Concession System of Incorporation in Late Imperial Russia, 1894–1908[†]

By AMANDA G. GREGG*

poration Reform, 1890–1905. *Economic History Review* 10 (3): 919–43.
Gregg, Amanda G. 2019. “Imperial Russian Factory Database, 1894–1908.” <http://doi.org/10.2886/E110681V1>.

owned by corporations were larger, more productive, and more mechanized than unincorporated factories. Higher productivity factories were more likely to incorporate and, after incorporating, added machine power and became even more labor productive. Russian firms sought the corporate form's full set of advantages, not just stock markets access, to obtain scarce long-term financing. (JEL D24, G31, G32, L60, N43, N63)

The corporation's role in industrial growth remains controversial. Though an older literature maintained that corporations mobilized the capital critical for large-scale industrialization, recent evidence suggests that this effect may be overstated.¹ Because the corporation is ultimately a legal institution, this controversy crystallizes a larger debate about whether institutions influence economic development. Measuring the effect of institutions on economic performance is challenging,

Factory Productivity and the Concession System of Incorporation in Late Imperial Russia, 1894–1908

Amanda G. Gregg

AMERICAN ECONOMIC REVIEW
VOL. 110, NO. 2, FEBRUARY 2020
(pp. 401–27)

Download Full Text PDF

mechanized than unincorporated factories. Higher productivity factories were more likely to incorporate and, after incorporating, added machine power and became even more labor productive. Russian firms sought the corporate form's full set of advantages, not just stock markets access, to obtain scarce long-term financing.

Citation

Gregg, Amanda G. 2020. "Factory Productivity and the Concession System of Incorporation in Late Imperial Russia, 1894–1908." *American Economic Review*, 110 (2): 401–27.

DOI: 10.1257/aer.20151656

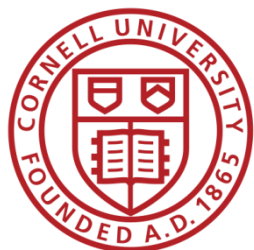
Choose Format:



Additional Materials

Data Set

Online Appendix (1.27 MB)



Predation, Protection, and Productivity: A Firm-Level Perspective.



Abstract



References



Online appendix



Supplementary
materials



Notes

Supplementary materials

- Code and Data

Besley, Timothy, and Hannes Mueller. 2018. "Replication data for: Predation, Protection, and Productivity: A Firm-Level Perspective." American Economic Association [publisher] DOI: [10.1257/mac.20160120.data](https://doi.org/10.1257/mac.20160120.data)

- ☒ Data is freely accessible under CC BY-NC 4.0 at [10.1257/mac.20160120.data](https://doi.org/10.1257/mac.20160120.data).

- Data

Statistics Norway. 2015. "Firm-level statistics 1975-2013 [dataset]" Norwegian Data Archive [curator], v2. DOI: [10.7654/nda::7643A::34](https://doi.org/10.7654/nda::7643A::34)

- ☐ Data restricted-access, under Norwegian Data Access license (has residency requirement, has citizenship requirement), accessible at Norwegian Data Archive in Oslo, Norway

Recommendation



Recommendation

FAIR Principles

- **Findable,**
- **Accessible,**
- **Interoperable, and**
- **Re-usable**

Start here:

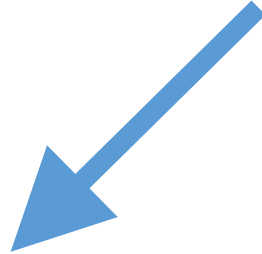
- **Create bibliographic metadata**
(DC, Schema.org, data.gov, DDI)
- **Present it in a human-readable form**
(Suggested citations)
- **Present it in a machine-readable form**
(citation managers)



Recommendation

FAIR Principles

- Findable,
- **Accessible,**
- Interoperable, and
- Re-usable



Continue here:

- **Describe access**
(license, data use agreement, application procedure)
- **Describe persistence (future access)**
(Record schedule, archive policy)
- **Present it in a machine-readable form**
(re3data.org)



Recommendation

FAIR Principles

- **F**indable,
- **A**ccessible,
- **I**nteroperable, and
- **R**e-usable



Follow up with:

- **Data documentation (what)**
(codebooks, summary stats)
- **Process documentation (how)**
(process flow, software code)
- **Present it in a machine-readable form**
(DDI, SDMX, schema.org)



Recommendation

FAIR Principles

- **F**indable,
- **A**ccessible,
- **I**nteroperable, and
- **R**e-usable

Improve with:

- **Use of common taxonomies**
(variable values, standards)
- **Implement APIs**
(live query, code submission)
- **Present it in a machine-readable form**
(data.gov, others)

Thank you!

DOI: [10.5281/zenodo.3652080](https://doi.org/10.5281/zenodo.3652080)



Social science “guild”



Data and Code Guidance by Data Editors

Guidance for authors wishing to create data and code supplements, and for replicators.

Authors: Lars Vilhuber

This project is maintained by social-science-data-editors

Disclaimer

Unofficial guidance on various topics by Social Science Data Editors

Guidance on creating replicable data and program archives

This guidance is for the author wanting to create a replication archive.

See [Requested information](#) for the information the Data Editor may request from you, prior to the acceptance of your paper for publication.

Guidance on testing replicability of code

This guidance has two audiences:

- the author wanting to verify whether her code passes muster as a replicable archive
- the replicator wanting to verify the replicability of such an archive

See [Verification guidance](#)

FAQ

See our growing [FAQ](#). If you have questions or answers to add, please notify us by creating a [new issue](#).

[https://](https://social-science-data-editors.github.io/guidance/)
[social-science](https://social-science-data-editors.github.io/guidance/)
[-data-editors.](https://social-science-data-editors.github.io/guidance/)
[github.io/](https://social-science-data-editors.github.io/guidance/)
[guidance/](https://social-science-data-editors.github.io/guidance/)