

# Implications

Julia Lane

NYU

## Longitudinal Employer-Household Dynamics

[Main](#)
[Applications](#)
[Data](#)
[Learn More](#)
[Research](#)
[State Partners](#)
[LED in Action](#)

**Applications**

- JQJ Explorer (Beta)
- QWJ Explorer
- OnTheMap
- OnTheMap for Emergency Management
- LED Extraction Tool

**Useful Links**

- Center for Economic Studies
- QWJ Data
- LODES Data
- JQJ Data
- PSEO Data
- LED Workshop

**Education Outcomes Data Expanded and Updated Data Visualization Released**

The LEHD program announces the expansion of the Post-Secondary Employment Outcomes (PSEO) statistics to include the University of Michigan - Ann Arbor and the University of Wisconsin - Madison. The interactive PSEO visualization tool has also been updated to include a degree level selector and data from these new institutions.

[View PSEO Data page](#)  
[View PSEO Visualization Tool](#)

## Data Enclave

NORC is a recognized innovator in the field of secure data management, analysis, and dissemination. The NORC Data Enclave®—NORC's constantly evolving data storage and management solution—is an integral part of the **Advanced Data Solutions Center**. The Enclave's high performance computing environment includes cutting-edge statistical, analytical, visualization, and reporting tools.

Since 2006, state and federal agencies, research institutes, foundations, and universities have used the Enclave to securely house and provide remote access to confidential data. Enclave-based research informs a wide spectrum of public and private sector decision-making, as well as journal articles, books, position papers, conference presentations, dissertations, etc. At any given time, the Enclave supports over 1000 researchers via contracts and grants with a wide variety of government, academic, nonprofit, and commercial clients.

**COMPUTING**

Providing a secure platform where users connect to relevant datasets for policy research using confidential micro-data.

### ADRF

The **Administrative Data Research Facility (ADRF)** provides a secure platform to host confidential micro-data. The ADRF is designed to promote collaboration, facilitate documentation, and provide information about use to data stewards. See our [whitepapers](#) for a more thorough description of ADRF components.

The ADRF has hosted almost 50 confidential government datasets from 12 different agencies at all levels of government. The ADRF has received Authorization to Operate from the Census Bureau, has achieved a FedRAMP Moderate approval and is listed on the [FedRAMP Marketplace](#).

## Stats NZ

STATISTICS TOOLS SERVICES AND SUPPORT INTEGRATED DATA CENSUS ABOUT US

Home > Integrated data > Integrated Data Infrastructure

## Integrated Data Infrastructure

01 July 2018, 10:00pm

[f](#)
[t](#)
[in](#)

The Integrated Data Infrastructure (IDI) is a large research database. It holds microdata about people and households.

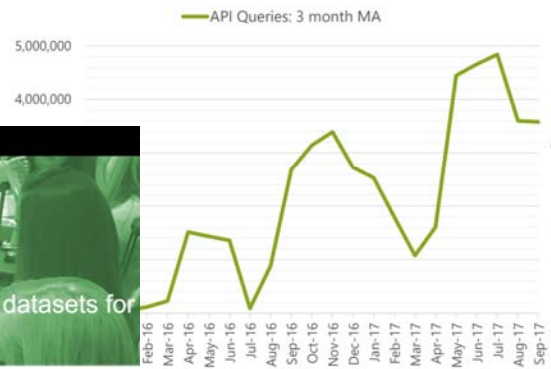
The data is about life events, like education, income, benefits, migration, justice, and health. It comes from government agencies, Stats NZ surveys, and non-government organisations (NGOs). The data is linked together, or integrated, to form the IDI.

The IDI complements the [Longitudinal Business Database](#) (LBD), which holds linked microdata about businesses. The two databases are linked through tax data.

Researchers use the IDI to gain insight into our society and economy. The research can help answer questions about complex issues that affect New Zealanders.

See:

## PatentsView API



**74%** year-to-date (Sep 2017) growth

**104K** average queries per day in 2017

**27.9M** total queries in 2017

**IRIS is...**

...a national consortium of research universities organized around an IRB-approved data repository. IRIS develops data for research and reporting to understand, explain and improve the public value of higher education.

UNIVERSITIES

RESEARCHERS

POLICY MAKERS


# Overview

- Three issues for statistical agencies
  - Remember core mission
  - Be realistic
  - Build ecosystem
- Three areas for future investment
  - Utility measures
  - Access
  - Output protection

# Overview

- Three issues for statistical agencies
  - Remember core mission
  - Be realistic
  - Build ecosystem
- Three areas for future investment
  - Utility measures
  - Access
  - Output protection

## Core Mission

 OFFICE OF BUREAU OF LABOR STATISTICS

**BJS**

Home | Correction | Courts | Crime Type | Criminal Justice Improvement | Employment and Expenditure | Federal | Indian Country Statistics | Law Enforcement | Victims

**BUREAU OF LABOR STATISTICS**

Home | Subjects | Data Tools | Publications

**JUN 06**

**Productivity rises 3.4% in first rates)**  
Productivity increased 3.4 percent unit labor costs decreased 1.6 percent productivity increased 0.4 percent  
[HTML](#) | [PDF](#) | [RSS](#) | [Charts](#)

06/05/2019 **County Employment and Wages (2018)**

05/29/2019 **April jobless rates down over the**

05/23/2019 **Productivity rises in 18 of 28 selected**

05/22/2019 **Tippecanoe, IN, has largest count**

**GEOGRAPHIC INFORMATION »**

**REGIONAL HOME**

**NEW YORK TIMES BESTSELLER**

"This is that rare book that not only informs but entertains. You'll never look at these four companies the same way again."  
—JONAH BERGER, author of Contagious and Invisible Influence

the  
**four**

THE HIDDEN DNA OF AMAZON, APPLE, FACEBOOK, AND GOOGLE

Scott Galloway

**WWW.OJP.USDOJ.GOV**

Home | About Us | Contact Us | Help | A-Z Topic List

Enter keywords **GO**

**Research**   **Print** **Text Size:** [-] [ + ]

**Tweets** by @BJSgov

**In Inmates,**

eral prisoners said  
g their offense.

ssed a gun during  
btain it from a  
ot obtain it from a

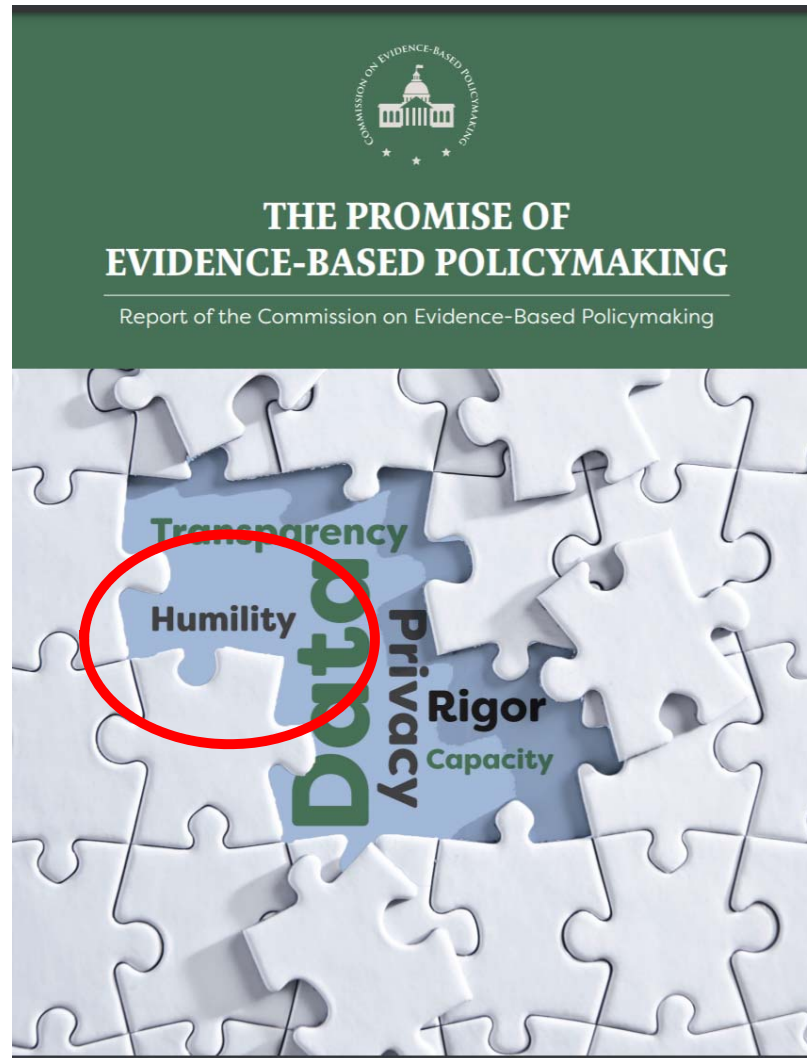
mmon type of  
ederal prisoners, as  
offense.

**BJS BJS STATS** @BJSgov

The National Crime Statistics Exchange (NCS-X) Initiative supports law enforcement agencies to transition to the FBI's National Incident-Based Reporting System (NIBRS). Learn more about NIBRS: [youtube.com/playlist?list=...](https://www.youtube.com/playlist?list=...) #NIBRS

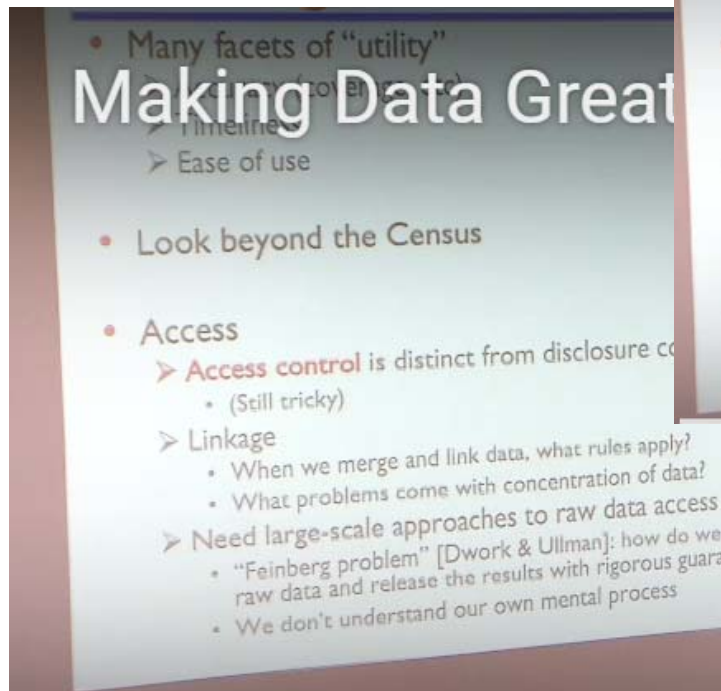
 **National In...**  
The FBI's N...  
[youtube.com](https://www.youtube.com)

Be realistic





# Realistic



• Many facets of “utility”

- Timeline
- Ease of use

• Look beyond the Census

• Access

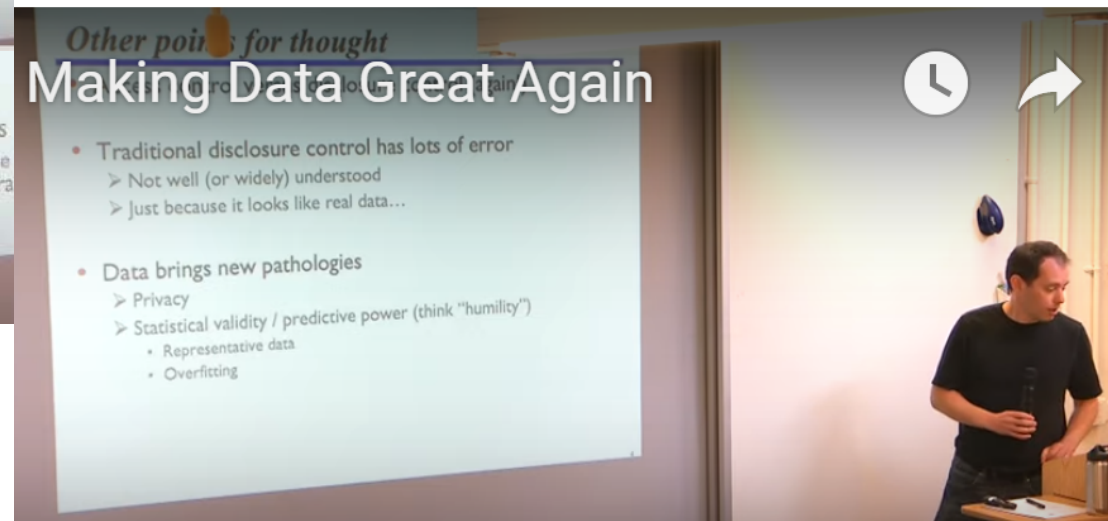
- Access control is distinct from disclosure control
  - (Still tricky)
- Linkage
  - When we merge and link data, what rules apply?
  - What problems come with concentration of data?
- Need large-scale approaches to raw data access
  - “Feinberg problem” [Dwork & Ullman]: how do we get raw data and release the results with rigorous guarantees?
  - We don’t understand our own mental process



• Dealing with small samples

## Making Data Great Again

- Outliers matter
  - Troublesome statistically, anecdotally crucial
- Refined notions of statistical utility
  - E.g., bias,
- Better tools to incorporate data uncertainty
  - Sampling theorems exist, require sampling distribution
  - Pufferfish/coupled-worlds approaches are hard to apply
- Participatory tradeoff management
  - E.g., how should Census decide what to release?



*Other points for thought*

## Making Data Great Again

- Traditional disclosure control has lots of error
  - Not well (or widely) understood
  - Just because it looks like real data...
- Data brings new pathologies
  - Privacy
  - Statistical validity / predictive power (think “humility”)
    - Representative data
    - Overfitting

# Areas for future research

## Understand the data

- Understand the structure of the core data – Census is a special case
- Consider sub samples
- Consider outliers

## Incorporate utility

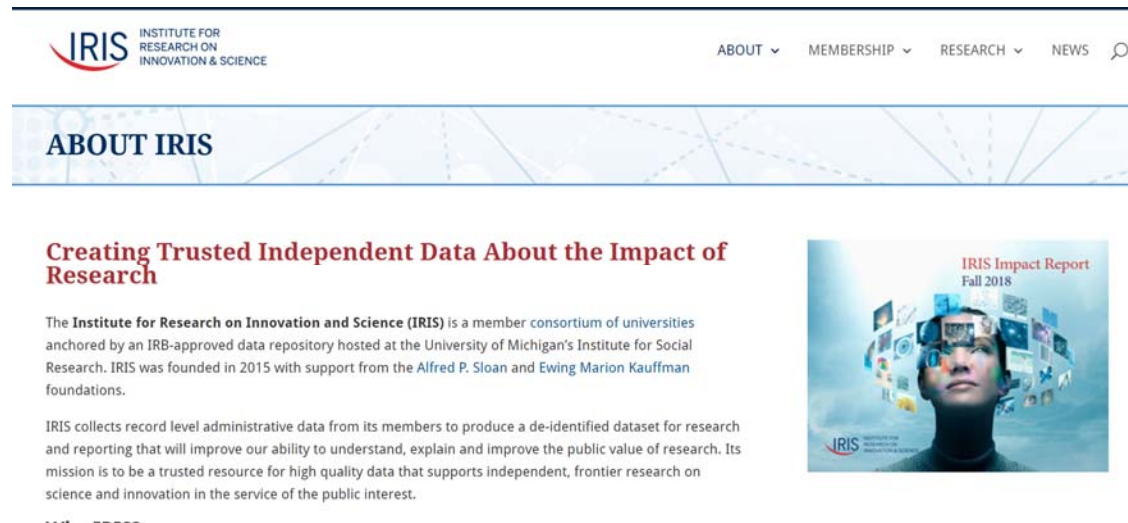
- Explicitly incorporate measures of utility in determining privacy budget
- Address more complicated variable structures in a way that preserves a sufficient amount of data utility

## General comments

- Explicitly incorporate size of sample into analysis
- Don't treat all fields equally



# Build ecosystem



# Overview

- Three issues for statistical agencies
  - Remember core mission
  - Be realistic
  - Build ecosystem
- Three areas for future investment
  - Utility measures
  - Access
  - Demonstrate value and move budget constraint

# Context



EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF MANAGEMENT AND BUDGET  
WASHINGTON, D.C. 20503

June 4, 2019

M-19-18

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Russell T. Vought  
Acting Director

A handwritten signature in dark ink, appearing to be "RV", written over the printed name of Russell T. Vought.

SUBJECT: Federal Data Strategy - A Framework for Consistency

- *Building a Culture that Values Data and Promotes Public Use (practices 1-10)*
- *Governing, Managing, and Protecting Data (practices 11-26)*
- *Promoting Efficient and Appropriate Data Use (practices 27-39)*

# Utility (and privacy) challenges

## The new world

- Datasets are not well defined entities
- Data are dirty and badly documented
- Missingness is common

## Access is feature not a bug

- Linkages are error-prone and often wrong
- Outliers are where the action is
- New products are created through trial and error
- Community engagement and use is the core business model

# Utility measures (examples)

- Completeness – data rich enough
- Timeliness – data deliveries adhere to schedules
- Relevance – data satisfy user needs
- Accessibility – access to data is user friendly
- Interpretability – documentation; meta-data
- Granularity – data detailed enough
- Value - Number of ways data are put to use
- Cost-effectiveness – value for money

# Access

- Safe People
  - Tiers of approved and trained researchers; tiers of legal controls
- Safe Projects
  - Tiers of approved projects, consistent with agency mission and utility
- Safe Settings
  - Tiers of secure environments
- Safe Data
  - Tiers of deidentified data
- Safe Outputs
  - Disclosure reviews and export controls

# Move bu

**IDEAS** Economic literature Authors Institutions Rankings Help/FAQ MyIDEAS (now with weekly digests)

follow this keyword search My keywords

Search econ literature

Advanced search

## IDEAS/RePEc search

All Microdatabase Bundesbank

In: Whole record Sort by: Relevance From: Any Year To: Any Year

Found 12 results for 'Microdatabase Bundesbank', showing 1-10

- Elena Biewen (2012): KombiFID: A First Glimpse of the Bundesbank Data**  
 This contribution gives a first overview of the data in KombiFID which refers to information on the balance sheets of firms (as collected in the Corporate Balance Sheet Statistics, USTAN) and also to information on foreign direct investment activities of German firms (as collected in Microdatabase Direct Investment, MiDI). Using descriptive statistics and a simple regression analysis we compare the results based on the original and the KombiFID data. Although many statistics based on the KombiFID sample are similar to the original, we observe partly large deviations, in particular regarding the MiDI sample. These differences may be attributable to the fact that large firms and firms from the manufacturing sector are over-represented in the KombiFID sample.  
*RePEc:aeq:aeqsbv132\_y2012\_i3\_q3\_p405-418*
- Christian Bellak & Markus Leibrecht (2010): Does Lowering Dividend Tax Rates Increase Dividends Repatriated? Evidence of Intrafirm Cross-Border Dividend Repatriation Policies by German Multinational Enterprises**  
 This paper explores the effect dividend taxes exert on dividends repatriated from foreign affiliates to their German parent companies. The empirical analysis based on firm-level data from the Microdatabase Direct Investment provided by the Deutsche Bundesbank first signals the validity of the original Lintner model for cross-border intrafirm dividend payments of German affiliates abroad. Second, results imply that high dividend taxes indeed have a statistically significant negative effect on dividends repatriated. Our calculations suggest that a one-percentage-point decrease in the dividend tax rate would increase dividends repatriated by about 3.75%. Evaluated at the mean of positive dividend payments, a semielasticity of -1.71 is derived.  
*RePEc:mhr:finarcurn:sic:0015-2218(201012)66:4\_350:dlttri\_2.0.tx\_2-s*
- Kesternich, Iris & Schnitzer, Monika (2007): Who is Afraid of Political Risk? Multinational Firms and their Choice of Capital Structure**  
 This paper investigates how multinational firms choose the capital structure of their foreign affiliates in response to political risk. We focus on two choice variables, the leverage and the ownership structure of the foreign affiliate, and we distinguish different types of political risk, such as expropriation, corruption and confiscatory taxation. In our theoretical analysis we find that, as political risk increases, the ownership share always decreases, whereas leverage can both increase or decrease, depending on the type of political risk. Using the Microdatabase Direct Investment of the Deutsche Bundesbank, we find supportive evidence for these different effects.  
*RePEc:prc:ceprdp:6468*

### Browse Econ Literature

- Working papers
- Journals
- Software components
- Books
- Book chapters
- JEL classification

### More features

- Subscribe to new research
- RePEc Biblio
- Author registration
- FRED data



IDEAS

Economic literature

Authors

Institutions

Rankings

Help/FAQ

MyIDEAS (now with weekly email digests)

KombiFID:

Author & abstract

Author

Listed:

- Elena Biewer

Abstract

This contribution g  
firms (as collected  
investment activiti  
statistics and a sim  
Although many sta  
deviations, in parti  
and firms from the

Suggested Cit

Elena Biewer

Applied Social S  
Berlin, vol. 132

Handle: *RePEc:aeq*

Download referen

More services and

MyIDEAS

Follow serials, authors, ke  
more

New papers by email

Subscribe to new addition

Author registration

Public profiles for Econom  
researchers

Rankings

Various rankings of resear  
Economics & related fields

RePEc Genealogy

Who was a student of who  
RePEc

RePEc Biblio

Curated articles & papers on various  
economics topics

DATA SCIENCE

Microdatabase direct investment

Mikrodatenbank Direktinvestitionen

Topics: Economics; Commerce, Management, Tourism and Services; Studies in Human Society

Join community

Access code books

Contribute annotations

Use dataset

Data Set context

3,453 Experts

156 Papers

51 Code books

1,077 Annotations

Overview

Metadata

Publications

83

Citations

19,759

Grants

21

Funding amount

USD 75.2 M

Patents

7

Citations

352

Clinical trials

9

Policy documents

7

Papers - 83

Experts - 127

Code books - 21

Annotations - 529

Related Data Sets

Sort by: Number of citations

The Effect of National Culture on the Choice of Entry Mode

Bruce Kogut, Harbir Singh

1988, Journal of International Business Studies - Article

Citations

2.9k

Altmetric

16

View PDF

Add to Library

Data

Splintering Urbanism

Stephen Graham, Simon Marvin

2002 - Monograph

Citations

1.6k

Altmetric

1

View PDF

Add to Library

Data

International R&D spillovers

David T. Coe, Elhanan Helpman

1995, European Economic Review - Article

Citations

1.6k

Altmetric

51

View PDF

Add to Library

Data

Understanding the rationale of strategic technology partnering: Nterorganizational modes of cooperation and sectoral differences

John Hagedoorn

1993, Strategic Management Journal - Article

Citations

1.1k

Altmetric

3

View PDF

Add to Library

Data

A transaction costs theory of equity joint ventures

Jean-Francois Hennart

1988, Strategic Management Journal - Article

Citations

977

Altmetric

3

View PDF

Add to Library

Data

More

economics department

Privacy statement

Legal statement

Services from the STL Fed

Data, research, apps & more from the  
St. Louis Fed

IDEAS is a RePEc service hosted by the Research Division of the Federal Reserve Bank of St. Louis. RePEc uses  
bibliographic data supplied by the respective publishers.

# Commenting and Annotation in JupyterLab

The screenshot displays the JupyterLab interface. On the left, the 'Datasets' panel lists two files: 'file:///data/adrf-000076-zipcode\_to\_county\_relationship\_file.csv' and 'file:///data/adrf-000079-qwi\_earnings\_age\_19\_21.csv'. The main area shows a Jupyter Notebook with a single cell containing a table of data. The table has columns for 'Year', 'Crop Production', 'Animal Production and Aq', and 'Forestry and Logging'. The data spans from 2000 to 2016. On the right, a comment thread is visible for the dataset 'adrf-000079-qwi\_earnings\_age\_1...'. The thread includes a comment from Brian E. Granger dated Feb 22 10:12pm, stating 'Hmm, I was expected data for the year 2000, anyone understand why it is', and a response from Fernando Perez dated Feb 22 10:15pm, stating 'The data was collected for 2000, but the storage format changed. We are waiting'.

	Year	Crop Production	Animal Production and Aq	Forestry and Logging
1	2000			
2	2001	1116	1382	1580
3	2002	1200	1438	1450
4	2003	1255	1444	1501
5	2004	1309	1517	1579
6	2005	1294	1555	1566
7	2006	1326	1636	1564
8	2007	1425	1782	1510
9	2008	1364	1854	1692
10	2009	1323	1804	1839
11	2010	1343	1783	1712
12	2011	1409	1793	1665
13	2012	1411	1780	1661
14	2013	1482	1853	1952
15	2014	1539	1946	2308
16	2015	1561	1976	3050
17	2016			

\* Early prototype

# Summary

- Three issues for statistical agencies
  - Remember core mission
  - Be realistic
  - Build ecosystem
- Three areas for future investment
  - Utility measures
  - Access
  - Output protection

# Comments welcome

- [Julia.lane@nyu.edu](mailto:Julia.lane@nyu.edu)
- <https://coleridgeinitiative.org>