Statipedia: A wiki platform for collaboration across agencies

At Federal Reserve Board, 18 Sept 2013

Peter B. Meyer
Office of Productivity and Technology, BLS
(but not speaking for the agency)

...with lots of advice from others, notably Michael Messner, Office of Water, EPA
Jim Buszuwski, BLS/OCWC
A look at Statipedia

Wikis across government agencies
   - Intellipedia, Diplopedia, OMB's MAX
   - Many other wikis in government

design differences

More from Statipedia

Futures
Vision: share methodology information online

Statipedia is a wiki for statistical staff across agencies

- An online searchable reference and workspace
- With methodology source material
- Pilot project
- Not for the public (outside .gov)
- Not for sensitive content (e.g. data before release)
Knowledge pooling

We can

- Share definitions
- Copy one another’s procedures
- Develop materials together
- Enable new staff to search and absorb quickly

Development of science and technology goes faster with easily shared knowledge

Platform is cheap and robust to budget uncertainty
Looks like Wikipedia.

User can edit a page directly.
A user can see the history of changes to a wiki page.
We can define common terms

Industry value added

The phrase *industry value added* means the *value added output* by an industry in a certain time period, usually a year. "Industry value added" is synonymous with *industry value added": "gross product by industry", "gross product originating by industry", and "gross product originating".[1]

An industry’s value added output is the value of its *Gross output* minus the value of its *Intermediate inputs*. Gross output is the sum of sales or receipts and other operating income, commodity taxes paid, and not inventory growth. The value of intermediate inputs is the value of goods and services purchased from other industries or imported and used up or consumed.[2]

Industry value added is estimated for most industries from expenditure data not industry income, because expenditures data is thought to be more accurately estimated.[3]

Future improvement

- follow up the sources to the source cited here

References

1. ↑ "Box: Gross Product Originating: Definition and Relationship to Gross Domestic Product" from BEA web site
2. ↑ "Box: Gross Product Originating: Definition and Relationship to Gross Domestic Product" from BEA web site
3. ↑ "Box: Gross Product Originating: Definition and Relationship to Gross Domestic Product" from BEA web site

Categories: BEA | NIPA
Describe and search classification systems
Can share source code
“BLS” is a category, with subcategories

<table>
<thead>
<tr>
<th>Category: BLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>This category is for pages related to the Bureau of Labor Statistics.</td>
</tr>
</tbody>
</table>

**Subcategories**

This category has the following 2 subcategories, out of 2 total.

**B**
- Statisticians at BLS

**C**
- CPS

**Pages in category "BLS"**

The following 59 pages are in this category, out of 59 total.

<table>
<thead>
<tr>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>790 program</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATUS</td>
</tr>
<tr>
<td>American Time Use Survey</td>
</tr>
<tr>
<td>At the agencies: Availability of journal archives</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template: At the agencies</td>
</tr>
<tr>
<td>Federal Economic Statistics Advisory Committee</td>
</tr>
<tr>
<td>Office of Productivity and Technology</td>
</tr>
<tr>
<td>BLS Handbook of Methods</td>
</tr>
<tr>
<td>BLS Handbook of Methods chapter 15 on International Price Indexes</td>
</tr>
<tr>
<td>At the agencies: Bayesian Statistics</td>
</tr>
<tr>
<td>Benefit incidence</td>
</tr>
<tr>
<td>At the agencies: Browsers</td>
</tr>
<tr>
<td>Bureau of Labor Statistics</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C cont.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census of Fatal Occupational Injuries (CFOI)</td>
</tr>
<tr>
<td>At the agencies: Cloud Computing</td>
</tr>
<tr>
<td>Collective Bargaining Agreements (statistical)</td>
</tr>
<tr>
<td>Commodities and Services Survey</td>
</tr>
<tr>
<td>Compensation and Working Conditions</td>
</tr>
<tr>
<td>Confidential Information Protection and Statistical Act (CIPSEA)</td>
</tr>
<tr>
<td>At the agencies: Confidentiality and Data Dissemination</td>
</tr>
<tr>
<td>Current Employment Statistics</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffusion index</td>
</tr>
<tr>
<td>Discouraged workers</td>
</tr>
<tr>
<td>Displaced workers</td>
</tr>
<tr>
<td>Duration of unemployment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>At the agencies: Editing Wikipedia</td>
</tr>
<tr>
<td>At the agencies: Email address structure</td>
</tr>
<tr>
<td>Employed persons</td>
</tr>
<tr>
<td>Employment Situation</td>
</tr>
</tbody>
</table>
Recent changes to the wiki
History of our kind of work

1848 Census of Egypt

Egypt conducted its first national census in 1848. This was under the reformist nominally-Ottoman ruler called Mohammed Ali, or Mohammed Ali Pasha.

Sources:

Quoting from the abstract:
In format and data collected, the Egyptian census registers appear to have been compiled in accordance with population counts undated at the same time in the central Ottoman Empire, though the former have the important advantage of including a count of women. These registers contain data on household composition, sex, age, religion, ethnicity, place of origin, marriage and polygamy, slavery, migration, economic activity, physical infirmities, and—In the towns—the type and fiscal status of dwellings. Although most of the sources for this period are biased toward representing the wealthier strata, the census registers offer an extraordinary view of all elements in society.
- Fergus, Philippe. Family and household in mid-nineteenth century Cairo in Deumur, Beshara, ed. Family history in the Middle East: household, property, and gender
- Salah, Mohamed. Historical Claims of Inter-Social Difference: Evidence from 19th and 20th Century Egypt. University of Southern California, Dept of Economics

2004 Economic Census of China

The 2004 Economic Census of China covered Chinese industries except for agriculture, forestry, animal husbandry, and fishing.\[1\]

The resulting data were used for China’s National Accounts.

The 2004 Economic Census came closer to completeness and meeting international standards than the usual annual data, and previous Censuses, had done:

1. Earlier censuses covered services less extensively partly because intangible service activities did not meet the Marxist definition or concept of industry.\[2\] For information on the National Accounts before 2004 see the pages on the Development of China’s National Accounts.
2. Government, administrative and institutional units were more completely covered based on financial information in the Economic Census which included a questionnaire beyond the usual statistics of government compensation and financial statements of those units.\[3\]
3. Small businesses which had not registered with the Administration for Industry and Commerce (perhaps 40% of them) were much more likely to be covered by the 2004 Economic Census than earlier sources. This came about because of a questionnaire on individual businesses.\[4\]
4. A classification into 94 industries was possible, finer that in previous data.\[5\]
5. Coverage of auxiliary establishments whose activities were different from that of the main enterprise improved because of a questionnaire to establishments.\[6\]
6. A related Household Survey estimated 11 categories of consumer expenditure for rural households and 12 for urban households.\[7\]

Reference

1. ↑ 1.0 1.1 1.2 1.3 1.4 1.5 XU Xianchun. 2006. Recent Progress in China on the SINA and Challenges for the New Revision. Paper prepared for the 29th General Conference of the International Association for Research in Income and Wealth, at Joensuu, Finland, on Aug 20-26, 2006.

History of BLS Mass Layoff Statistics

The Mass Layoff Statistics program at BLS began in 1984.\[8\]

A Mass Layoff is defined to be an event that causes 50 or more unemployment insurance claims to be filed within a 5 week period regarding a single employer.

Data from the program are available monthly since April 1995.\[9\]

The program releases data monthly, around the 22nd of the month. It is administered by BLS/OEUS with a staff person at each of BLS’s regional offices.

Mass layoff data are regularly redistributed by the Bureau of National Affairs, for example in this update covering March 2011.\[9\]

References

1. ↑ Timeline of BLS history, drawing from BLS Major Milestones on BLS Intranet
Can share info on seminars in DC area
<table>
<thead>
<tr>
<th>Conference</th>
<th>When</th>
<th>Where</th>
<th>Dates submissions due (yyyy-mm-dd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011 Joint Statistical Meetings</td>
<td>July 30 - August 4, 2011</td>
<td>Miami Beach Convention Center</td>
<td>2011-02-01</td>
</tr>
<tr>
<td>FedCASIC Workshop</td>
<td>22-24 March 2011</td>
<td>Wash DC, at BLS</td>
<td>2011-02-01 (? 50 word abstract: Dan Gillman)</td>
</tr>
<tr>
<td>Federal Forecaster Conference</td>
<td>21 April 2011</td>
<td>Wash DC, at BLS</td>
<td>2011-02-01</td>
</tr>
<tr>
<td>International Symposium on Forecasting</td>
<td>June 26-29, 2011</td>
<td>Prague, The University of Economics</td>
<td>2011-03-04</td>
</tr>
<tr>
<td>Wikimania 2011</td>
<td>Aug 4-7 2011 (plus meetings Aug 2-4)</td>
<td>Haifa, Israel</td>
<td>2011-04-30 Call for participation</td>
</tr>
<tr>
<td>ICES IV</td>
<td>11-14, June 2012</td>
<td>Montreal, Canada</td>
<td>2011-03-01 (not set at this writing)</td>
</tr>
<tr>
<td>SHOT (Society for the History of Technology) 2011</td>
<td>3-6 Nov 2011</td>
<td>Cleveland</td>
<td>2011-spring (not set at this writing)</td>
</tr>
<tr>
<td>2012 Society of Labor Economists</td>
<td>4-5 May 2012</td>
<td>Hyatt Regency Chicago</td>
<td>2011-fall</td>
</tr>
<tr>
<td>ASSA Annual Meeting 2012 (economists)</td>
<td>6-8 Jan 2012</td>
<td>Chicago</td>
<td>2011-01-31? Not stated yet on line of site.</td>
</tr>
<tr>
<td>Predictive Analytics World including workshops</td>
<td>2011-03-13 (March 13-19 2011)</td>
<td>San Francisco</td>
<td>Didn’t see how to submit an abstract.</td>
</tr>
</tbody>
</table>

Track conferences on our topics
Can compare procedures across agencies

This is one of many “At the agencies” pages
Can share simple training materials in the wiki

## Tiny tutorial for Matlab

From a Unix shell one can just type 'matlab' as a command on any computer that has it, and start to type interactive statements such as those below. One could also put them in a file with the .m extension to run them from within matlab with 'run file.m' or from the shell with 'matlab < file.m' This tutorial covers very little but you can see something of the language.

% The percent sign begins comments.
% The statements below can be typed interactively one per line to get
% clear responses from Matlab. There's no need to type the comment part at the
% end of the lines. Make sure to use upper and lower case in the
% same way as in the statements shown.

A=[1 2;3 4]  % defines matrix A as a 2x2 with first line [1 2]
B=A'        % transpose
B=A+A       % sum, element by element
Ainv=inv(A) % takes inverse of a matrix
A*Ainv      % calculates and prints the result of a matrix multiplication
B=[A;A]     % stacked so B has twice as many rows as A
B=[A A]     % the A's are side by side. B has twice as many columns as A.
B=A(1,1)    % B is a scalar now, the upper left element of A
B=A'*A      % matrix multiplication
B=A(:,1)    % B is set to first row of A
B=A.*A      % element by element multiplication
B=A./A      % element by element division
A=zeros(3,3) % special definition of a matrix of zeros
B=ones(3,1) % defines a matrix of ones
A=eye(5)    % defines identity matrix
B=A(1:2,1:3) % takes part of matrix
more on     % may not be needed; prevents help screen from scrolling off
help        % shows sample of the help available

Categories: Software | Tutorials
And detailed training materials searching

- links and footnotes that can be clicked for more info

- equations can be copied
Recording and discussing critical perspectives

"Real" unemployment rate

The headline measure of unemployment is a market measure, meaning that it counts as unemployed those people who are (a) not employed, (b) want to work, (c) are available to work, and (d) are actively looking for work; a behavioral criterion, defining who is in the market looking for a matching partner.

These are some common criticisms of this measure. People naturally want some measure that covers all those people who want to work and are not working including those who are not actively looking. This last group is called the discouraged workers sometimes.

BLS measures a variety of concepts of unemployment. A key fact in this discussion is that they track very much in parallel, so they tell the same business cycle story. For evidence see this graph and table from Haugen (2009)\(^1\), which has a valuable discussion of the history and current set of measures.

---

<table>
<thead>
<tr>
<th>Measure abbreviation</th>
<th>Definition</th>
<th>2008 average</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-1</td>
<td>Persons unemployed 15 weeks or longer, as a percent of the civilian labor force</td>
<td>2.1%</td>
</tr>
<tr>
<td>U-2</td>
<td>Job losers and persons who completed temporary jobs, as a percent of the civilian labor force</td>
<td>3.1%</td>
</tr>
<tr>
<td>U-3</td>
<td>Total unemployed, as a percent of the civilian labor force (official unemployment rate)</td>
<td>5.8%</td>
</tr>
<tr>
<td>U-4</td>
<td>Total unemployed plus discouraged workers, as a percent of the civilian labor force plus discouraged workers</td>
<td>6.1%</td>
</tr>
<tr>
<td>U-5</td>
<td>Total unemployed, plus discouraged workers, plus all other marginally attached workers, as a percent of the civilian labor force plus all marginally attached workers</td>
<td>6.8%</td>
</tr>
<tr>
<td>U-6</td>
<td>Total unemployed, plus discouraged workers, plus all marginally attached workers, plus total employed part time for economic reasons, as a percent of the civilian labor force plus all marginally attached workers</td>
<td>10.5%</td>
</tr>
</tbody>
</table>

References

Institutions

Statistical Community of Practice and Engagement (SCOPE)

The Statistical Community of Practice and Engagement (SCOPE) is a joint effort of statistical agencies sponsored by OMB’s statistical office and the ICSP. SCOPE will knit some IT services of the agencies together for efficiency and security -- sometimes called “horizontal IT consolidation.”

From OMB/IOIRA[2]: “The President’s FY 2011 budget proposes an initiative for the Economic Research Service to serve as the Program Management Office for a Statistical Community of Practice (SCOP, pronounced SCOPE). Statistical agencies would participate in this system-wide initiative on a voluntary basis. The program management office would coordinate system-wide projects on standards, practices, policies, and protocols concerning targeted barriers and inconsistencies that now exist across agencies. SCOP will permit the Federal statistical community to more easily identify and share best practices for Federal statistical activities, resulting in significant data quality, security, and operating efficiency and productivity gains in the statistical activities of our Nation.”

This is beginning (as of 2010-2011) to include FedStats, the FCSM web site, and could include Statipedia in the future. It does not include Data.gov at this point. The SCOPE project management office is at the USDA/ERS. Contact people include Ron Bianchi (USDA), Marilyn Stastny (NCES), Karl Sudmuns (USDA), and Paul Bugg (OMB). Online cooperation occurs on OMB’s MAX[3]

Two million dollars were allocated in the President’s FY 2011 budget proposal for ERS to run SCOPE.[4] SCOPE may require a “waiver of vertical IT consolidation [internal to agencies] to allow . . . integration across statistical agencies.”[5]

One objective is to make a Line of Business for SCOPE and/or the statistical agencies. (This may ease acquisition). Another is to streamline Paperwork Reduction Act Information Collection Requests. Another objective is to reduce procurement costs and difficulties by doing it together, another is to adopt open source software, another is to standardize on metadata, another is to offer joint data access to sensitive microdata. A clear comparison is made to other countries where there is a single big statistical agency. One analogous success is the Food Atlas App which resulted in Dec 2009 from data shared by ERS, NCHS, and Census.[6]

Productivity and Costs news release

The Productivity and Costs release from BLS report on labor productivity across large segments of the U.S. economy. Here are the archived releases including the most recent one: http://www.bls.gov/etr/database/archives/prodrl.htm#1915

In the 3 Feb 2011 releases[1] it was estimated that U.S. nonfarm business sector labor productivity increased at a 2.8% annualized rate from the third quarter of 2010 to the fourth quarter. This is calculated mainly from these estimates: output grew at a 4.5% annualized clip and hours-worked grew at a 1.7% clip between these periods. The nonfarm business sector covers about 75% of the total economy, and we do not have good ways to measure the output of most of the rest.

History

News releases from this program are online back to 1963. Some data series from the program go back to 1867[2].

References

1. Productivity and Costs Release, 3 Feb 2011
2. Online last data from LDC labor productivity and costs program

Categories: Press releases | BLS | Productivity measurement | Principal Federal Economic Indicators | Statistical programs
Hot topics

- Simpler: the future of government, by Cass Sunstein
- Open Data guidance
- Web APIs from statistical agencies
- Google’s data language
New issues
Quick technical discussion

Source code control systems
Quick survey, taken mainly from wikipedia/Template/Revision_control_software and wp/List of revision control software and wp:Comparison of revision control software and of revision_control_software wp:List of revision control software

- Year marks the date of first stable release.
- Systems with names in italics are no longer maintained or have planned end-of-life dates.
- Most-used are at top, where "planned" is a mix of most-used, most actively developed now, most advanced, most usable, most important.
In 2004 the U.S. intelligence agencies reviewed their communication tools and processes (Andrus, 2005; Burton, 2005)

They added web tools and work spaces to their network

Intellipedia wiki across agencies starting 2006

Blogs, tagging, search, messaging, videos

easier search across 16 agencies
Diplopedia

- Intended for sharing procedures (Bronk and Smith 2010)
- E.g. for staff at new posts
- Article count grew linearly with time after 2006 start
Page counts grow linearly over time

Growth of Canadian government’s internal wiki GCpedia

- 33,000+ registered users
- 2,500+ contributors
- 1,300+ active users
- 18,700+ content pages
- 775,000+ edits
- 15,300,000+ page views

As of July 2012
OMB’s MAX Federal Community

- Used first for budget work, then for other groups
- Runs Confluence software
- Can recognize NASA computers - A key capability for the future
- costs drop once agencies can use one another’s services
Many other examples

US federal govt
- Powerpedia at DOE
  - 2500 users, 6K articles, 27K pages
- DoDTechipedia
- FBI’s Bureaupedia
- EPA wikis
- DOL wiki
- OSHApedia
- NIH wikis
- NITRD’s wiki

Others doing governance
- Canada’s GCPEDIA (all civil svc)
- Australia’s GovDex (all civil svc)
- UK’s Civil Wiki (all civil svc)
- OECD’s Wikiprogress
- Eurostat’s Statistics Explained
- City wikis
- Katrina Help Wiki
- OpenWetWare
- GovLoop
Different wiki designs & uses (1)

Broad govt-wide vs. focused knowledge pool

- GCPedia is available across Canadian civil service
- Not all US federal staff have a wiki available & allowed
  - And, what they have requires frequent login
  - Or is not visible across most of government
- Key usability criterion:
  - If I email a link to the wiki, will it work for recipient?
  - In a winning all-government system, yes.
- Cost/complexity of many wikis, admins, namespaces
Different wiki designs & uses (2)

Storing **documents** versus **wiki-text**

- Documents are self-contained, often finished binary files
  - may be hard to search for; authored; dated; formatted; may have intro, conclusion; may have focused audience; may be official; updates require coordination or permission; common on SharePoint & Confluence

- Wikitext is searchable and open
  - fragmentary; multi-authored; unlikely to be official or definitive; can be updated and link to latest sources; open format; can manipulate with extensions, import/export

- Both needed; some mismatch in what’s used/available
Reference work: Statipedia

Past or comparative cases
- 2004 Economic Census of China
- Hungarian Central Statistical Office
- Hungarian industry classification TEÁOR
- History of BLS Mass Layoff Statistics

New technical subjects
- Wikidata
- Data.gov
- Dataset Publishing Language
- Semantic Mediawiki

Other wikis
(to learn from)

Events
- November 2013 Events
- Conference submission due dates
- Notes on presentations

Agency practices & policies
- Editing Wikipedia for work

Critiques and FAQs

Source code
Norms for admins and users

- **Scale up:** Benefits rise as communities of interest emerge

- **Meet open technical standards**
  - HTML, TEX, wiki-text links, extensions
  - Participate in software development
  - Co-evolution: Today’s community, technology, platform
    ➔ Tomorrow’s community, technology, platform

- **Voluntary** participation; serve and empower staff

- Anchor discussions to sources, evidence, theory
  - Enable drilling down to sources

- Address broad audiences -- not organization-specific
Better knowledge management

- More efficient/scientific communities
  - Shared source material, more reference points
  - Mutual awareness and peer review
- Reduce duplication of effort
- Specialists serve larger audience
  - Users can find them on this platform
- Ease training and turnover
- Less email? ➔ “discoverability” not “dissemination”

TOGETHER, the agencies have vast, diverse expertise and capability, economies of scale and scope, knowledge of data, and great computer resources
Challenges & Opportunities

- **Logging in** each time is a burden
  - Would benefit from federal “single sign-on”
  - Like OMB MAX Community does with NASA PCs
- **Editing wikitext** will be easier in new versions
- **More content** will help (now 1100 pages)
- Push from management
  - After above issues addressed further
- Has not been used for an actual project
Conclusions and future steps

Statipedia is a wiki for statistical staff across agencies

- With methodology source material
- Not for the public (but could include non-.gov)
- Not for sensitive content
- Content is growing

In future

- Can adopt and develop more tools and content
- Share source code better (not just on wiki)
- Make it more quickly and easily available and usable
Contact Information

Meyer.Peter@bls.gov
202-691-5678

Messner.Michael@epa.gov
202-564-5268
“Management visions often do not match the benefits delivered by successful wikis.” -- Grudin & Poole (2010)

- Citizens get their information from wikis
  - We in govt are in the information services business
  - It’s relevant to know the science and practice
  - Writing for Wikipedia meets our mission directly

- Empowerment
  - Benefit to our personnel – e.g. in enabling to publish good work
  - Users find specialists, who then serve larger audience
  - **Together**, agencies have vast, diverse resources, expertise, capability, and economies of scale and scope
More purposes and futures (2)

- **Efficient scientific communities (knowledge management)**
  - Support references to evidence ➔ cultural change
  - Shared source material, reference points, mutual awareness, peer review (reduce fiefdoms)
  - Support tools development by our own staff
  - Info on platforms - “discoverable” not “disseminated”
  - Could reduce email strain

- **Support innovation from outside-in**
  - Compare practices between agencies
  - Give workers broad opportunity/right to wikis
    - like IG’s office; like freedom of speech
  - Emergent or open-source type innovation (a research area)
Rhetorics for wikis in govt

- Openness, transparency (translucency)
  - Overcoming stovepipes/silos
  - Reduce duplication of effort
  - Clarify & unify procedures
  - Ease training, turnover, emergency management

Rhetorics against

- System’s insecure or uncertified
- Organizational culture doesn’t fit
  - Wiki would encourage sloppy or undignified handling of info
  - It’s not our job to make a “knowledge base”
Update

- Content & user logins grew by about 1/3
  - We have 1000 pages
  - 33 users have made an edit
  - Now including federal contractors
- New calendar scheme is working well
- Developed content for Wikipedia (Tornqvist index)
- Presentation on wikis-in-government
- Updated software to MediaWiki 1.17.
To get innovation and improvement in government . . .

It’ll help to have a wiki that says what governments do.