Statipedia: A Wiki Platform for Collaboration Across Agencies

Peter B. Meyer[1]

[1]Research Economist, Office of Productivity and Technology, U.S. Bureau of Labor Statistics, 2 Massachusetts Ave NE, Washington DC 20212 ; Meyer.peter@bls.gov;

Findings and views expressed here are the author's and do not represent the Bureau.

Abstract

Statipedia is a new wiki for statistical staff across U.S. federal agencies. Government staff can collect definitions, training materials, and reference materials relevant to their work. The author invites others in the federal government to participate.

Key Words: collaboration; software; wiki; platform; government

1. What Statipedia is

Statipedia is a new wiki for statistical staff across U.S. federal agencies. A wiki is a web site where users can edit the pages directly from the Web browser, and can see the past history of versions of these pages. This wiki looks a lot like Wikipedia because it runs the same software.

Statipedia is a pilot project, temporarily authorized by the sponsoring agencies. It is not intended to hold sensitive content, such as pre-release data or private personal information. It is not for the public but rather for U.S. federal staff to work together efficiently.

2. Motivating vision

Because the wiki can be edited quickly and the results viewed from across the agencies, Statipedia can potentially serve as an online workspace for federal staff and also as a reference work with technical and administrative definitions, training materials, basic research sources on methodology, source materials on administrative practice, and source code for useful computer programs. The content can be copied or hyperlinked to.

This effort to pool knowledge together across the agencies has an especially powerful potential in the U.S. federal context. Unlike most other national governments, the U.S. government does not have a single large statistical agency but rather many, divided by federal Department and subject matter. These various agencies must build up similar infrastructures and solve similar problems in parallel, while they operate separately across locations. These organizations are peers, without relations of hierarchical authority with respect to one another. They are not generally in competition so the economic models of competitive industries do not apply completely.

The issues of knowledge sharing in such contexts have been discussed in substantial academic literatures and management literatures associated with such terms as *distributed innovation; open source innovation; user innovation; collective invention; communities of practice; knowledge management; standards; benchmarking; peer production; information commons; open science,*[1] and more. The basic idea we draw from this literature is that the staff of the federal agencies can do better work, and get more done, if they can search a common set of content online; add to it; compare to and copy the work of others; standardize on common designs; and dynamically plan to share future designs.

There are governance issues which, if not handled well, could make the effort infeasible. Later this document discusses how we have addressed these.

Explicit costs (apart from time spent are low, and can be expected remain low as capabilities improve over time. The software is actively developed by the Wikimedia Foundation and others, and is used heavily 24 hours a day. The Environmental Protection Agency made the key decision to adopt the MediaWiki software and other open source programs and tools own enterprise-wide collaboration and to make them usable to partners outside their organization. This gives economies of scale to a number of projects together, of which Statipedia is just one. We believe the platform is overall cheap and furthermore robust to budget uncertainty because it is deeply integrated into EPA's work; shutting it down would be far more expensive than keeping it going.

Other clusters of agencies have adopted similar platforms to address similar sets of opportunities and problems. They were led by the sixteen intelligence agencies which use blogs, wikis, and instant messaging on a common platform that is available to their staff. This set of tools makes up a common "collaboration services" platform, sometimes referred to by the name of its main wiki, Intellipedia. In the foreign affairs context there is now a Diplopedia run by the State Department. The Defense Department's runs a Techipedia for scientific and technical collaboration across its many units and their contractors. OMB runs a wiki and discussion board called MAX which federal government staff can access, intended for budget discussion and other cross-agency discussion and collaboration. We learned from all these cross-agency efforts before starting Statipedia.

2.1 Design objectives

The motivating vision was to make something like Wikipedia widely available to the "statistical community" so interested parties could make things clear to one another. We believed it would then be possible to apply good scientific practices better, quickly, and more easily than we now do. Knowing what online communities look like, we knew what we wanted and could get; one can think of these as design goals or requirements:

[1] On *distributed innovation*, see the works of Wesley Cohen; on *open source innovation*, see Meyer (2007) and Pénin (forthcoming); on *user innovation*, see the works of Eric von Hippel and his many coauthors and students; on *collective invention* see Robert Allen's work; *communities of practice*, *knowledge management*, *benchmarking*, and *standards* are the subject of vast academic literatures; on *peer production*, see the works of Yochai Benkler; on the idea of *information commons*, see Hess and Ostrom's *Understanding the Knowledge Commons* (2006); on *open science*, see the works of Paul David.

- Our platform tools should enable the use of footnotes and the elegant display of equations, since our scientific content uses these. (The equations are stored in the TeX format; in the long run the platform may support the coauthoring of full TeX sections and documents.)

- The platform tools should support dense, frequent, and easy hyperlinking, since we want to enable users to drill down in to the sources of an assertion and jump to related content.

- The platform should not encourage the creation of many enclosed, or secret, subsections. It is hoped we can develop and an open scientific community and did not wish to encourage forms of protection by secrecy. (There did not seem to be much benefit to come from storing *datasets* on the platform in the early going. The activities of the agencies tend to involve *procedures* and *methods*, some of which could usefully be shared widely across the U.S. government. Most listeners assume such a platform would be used for data, however, and this may yet happen.) It had been observed that if fine-grained security and secrecy features were available, some federal staff would be inclined to use them, and this would tend to replicate existing organizations, occupations, ranks, blockages, and categories online. Following many examples, our design preference was to make a platform that was open (among federal users) and simply leave secret information elsewhere. The pilot project will demonstrate, or fail to demonstrate, the usefulness of the new open service.

- The platform should be build on open source software so that our own researchers, scientists, information technology staff, and others could help develop it, add extensions, fix bugs, and adapt it to any special purposes our community developed. Historically scientific and research staff have made important tools and our statistical community should have this capability online. Development of science and technology goes faster with easily shared knowledge; this basic motive has driven the creation of libraries, academic journals, online platforms for communities to share source code, and many other institutions.

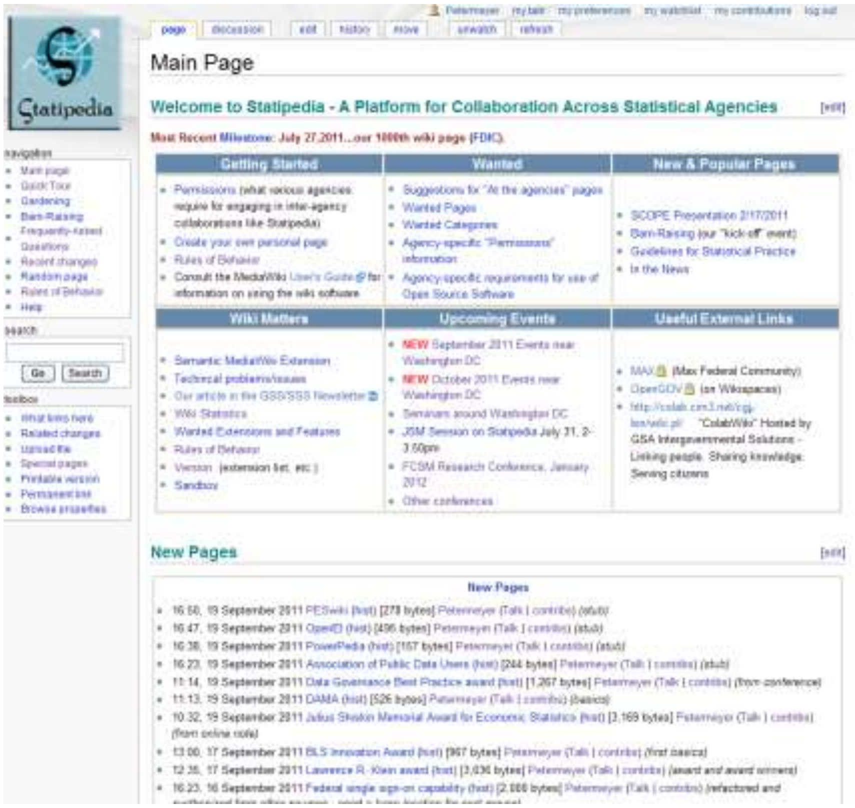
- Apart from explicit design requirements we needed the platform to enable us to easily write and share carefully phrased definitions; to copy one another's posts with designs, procedures, and software; to develop texts material together; and to be searchable so new staff would be enabled to search and learn independently. That conceptual proposal dated back some years in several forms and revisions.[2]

The platform at EPA meets these goals well and hopefully will evolve further in these directions.

3. What is on Statipedia?

There is a front page, which most users will see after they log in. It has links to a spectrum of subjects of likely interest. The user interface is visibly like that of Wikipedia. (See Figure 1.) The link to a user's name at the very goes to his "user page" where a user can describe himself and keep his own text and links to other things. The Rules of Behavior are always just one link away, often in the column at left which is called the "navigation bar." "Recent changes" shows recently edited pages so one can see what is happening, and respond dynamically to others.

[2] Open Source Practices team report (2008); Meyer and Buszuwski (2010); and earlier proposals at BLS and implementations at EPA and other agencies.



Main Page

Welcome to Statipedia - A Platform for Collaboration Across Statistical Agencies [edit]

Most Recent Milestone: July 27,2011...our 1000th wiki page (FDIC).

Getting Started	Wanted	New & Popular Pages
Permissions (what various agencies require for engaging in inter-agency collaborations like Statipedia) Create your own personal page Rules of Behavior Consult the MediaWiki User's Guide for information on using the wiki software	Suggestions for "At the agencies" pages Wanted Pages Wanted Categories Agency-specific "Permissions" information Agency-specific requirements for use of Open Source Software	SCOPE Presentation 2/17/2011 Barn-Raising (our "kick-off" event) Guidelines for Statistical Practice In the News
Wiki Matters	**Upcoming Events**	**Useful External Links**
Semantic MediaWiki Extension Technical problems/issues Our article in the GSS/SSS Newsletter Wiki Statistics Wanted Extensions and Features Rules of Behavior Version (extension list, etc.) Sandbox	NEW September 2011 Events near Washington DC NEW October 2011 Events near Washington DC Seminars around Washington DC JSM Session on Statipedia July 31, 2-3:50pm FCSM Research Conference, January 2012 Other conferences	MAX (Max Federal Community) OpenGOV (on Wikispaces) http://collab.cm3.net/cgi-bin/wiki.pl "CollabWiki" Hosted by GSA Intergovernmental Solutions - Linking people. Sharing knowledge. Serving citizens

New Pages [edit]

New Pages

- 16:50, 19 September 2011 PESwiki (hist) [278 bytes] Petermeyer (Talk | contribs) (stub)
- 16:47, 19 September 2011 OpenID (hist) [496 bytes] Petermeyer (Talk | contribs) (stub)
- 16:38, 19 September 2011 PowerPedia (hist) [157 bytes] Petermeyer (Talk | contribs) (stub)
- 16:23, 19 September 2011 Association of Public Data Users (hist) [244 bytes] Petermeyer (Talk | contribs) (stub)
- 11:14, 19 September 2011 Data Governance Best Practice award (hist) [1,267 bytes] Petermeyer (Talk | contribs) (from conference)
- 11:13, 19 September 2011 DAMA (hist) [526 bytes] Petermeyer (Talk | contribs) (basics)
- 10:32, 19 September 2011 Julius Shiskin Memorial Award for Economic Statistics (hist) [2,169 bytes] Petermeyer (Talk | contribs) (from online note)
- 13:00, 17 September 2011 BLS Innovation Award (hist) [967 bytes] Petermeyer (Talk | contribs) (first basics)
- 12:35, 17 September 2011 Lawrence R. Klein award (hist) [3,036 bytes] Petermeyer (Talk | contribs) (award and award winners)
- 16:23, 16 September 2011 Federal single sign-on capability (hist) [2,000 bytes] Petermeyer (Talk | contribs) (refactored and

Figure 1: Statipedia's main (first) page

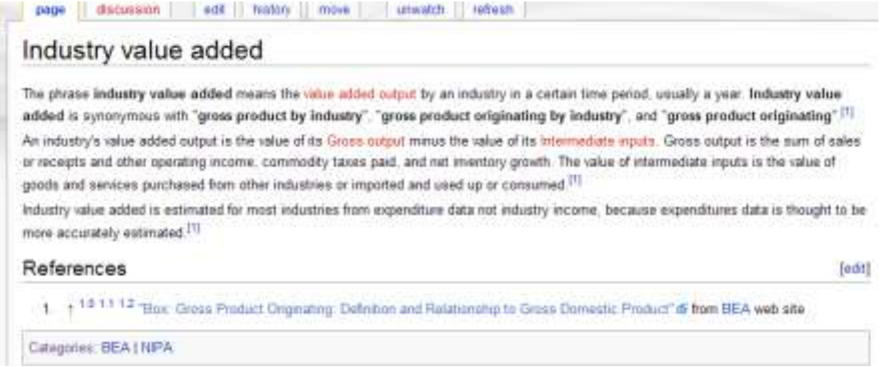
A user can edit almost any page by clicking on the edit tab at the top, which brings up a text editor to change the text of the page. The stored content is in "wikitext" which is a simplified version of HTML, with some special features to enable pages on the same wiki to work well together. Users can see how to insert a sentence but are sometimes confused by the wikitext for tables, mathematics, enhanced text, section headings, and "templates" which include wikitext from other pages. In the long run, the MediaWiki Foundation plans to include a kind of word processor to manage this complexity for the user.

A user can see the history of changes to most of the wiki pages – those with user content on them. Each such content page has a history tab, at the top, which if clicked showed who made the various edits to the page and when the edits were posted. Thus the text in the system can be audited to see which user wrote a particular word and when. All users are identified by name and all edits are associated in the history list with the user who made them and the time and date.

Each content page has an associated "discussion" by users of its content, which may be blank, and which can be found by clicking the tab at the top.

3.1 Definitions

A helpful use for Statipedia is to collect definitions of technical terms used in statistical agencies. An example is below. "Industry value added," is a term used across agencies (by BEA and BLS for example) and whose definition is a technical matter. If there were variant definitions, they could all be listed here. The user is invited to click on the footnote to "drill down" to the more exact formal definition; searching Statipedia thus offers a service to get to formally cite-able sources. (Statipedia is not usually an ideal source to cite, itself, because (a its content is dynamic; (b its content is unpublished and as a workspace would be expected to have mistakes; (c its authorship is mixed; and (d most readers do not have access to it.



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 "**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 net 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.[1]

Industry value added is estimated for most industries from expenditure data not industry income, because expenditures data is thought to be more accurately estimated.[1]

References [edit]

1. ↑ 1.0 1.1 1.2 "Box: Gross Product Originating: Definition and Relationship to Gross Domestic Product" from BEA web site

Categories: BEA | NIPA

Figure 2: An example definition, with footnote and categories.

One could envision adding such a definition to Wikipedia for the benefit of the general public, but (a) the term is narrowly focused and may not be of general encyclopedic interest, and (b) many federal staff are not permitted to edit Wikipedia from work. Statipedia is a natural repository to develop such content for reuse on Wikipedia or in publications at some later time. Content on Statipedia is reviewed as users pass by it, which should help bring out improvements.

Definitions related to classification systems are potentially especially useful to federal statisticians because categories for industry, occupation, region, health condition, and so forth) are so frequently used by analysts and researchers in the federal system, and also are developed by the staff of the statistical offices in cross-agency projects. For an example proposed use in the Statipedia context, see Meyer and Buszuwski (2010).

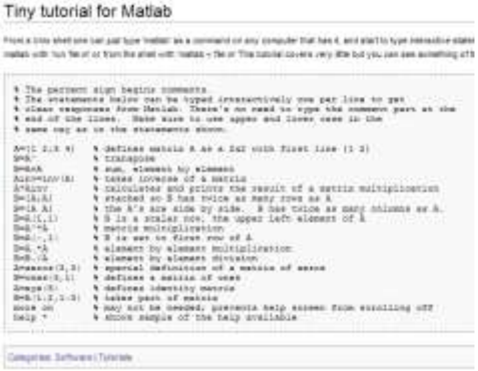
3.2 Source code

It is possible to share source code on wiki pages. Examples include self-contained interpreted analytical programs and subroutines in any computer language.

For larger projects made up of multiple files, many programmers, or complicated multi-file procedures to build the program or execute it, a wiki is not a very good way to share code because it does not conveniently package the group of files together and understand their different types. The opportunity of storing open-source software in repositories has been addressed by public Web sites including SourceForge, Tigris.org, and Github. Open source programmers choose to use such open sites partly to find coauthors. Government agencies often do not permit their staff to post code to such platforms. The Defense Department has set up its own analogous site, called Forge.mil. A civilian-side analog for the statistical agencies is not yet generally available to my knowledge. When it is, this will be a better way to store and share source code, and can be expected to make it easier to develop software.[3]

3.3 Training materials

Training materials can be gathered together on Statipedia, and collectively developed. Basic training materials can be written by Federal government staff, for federal government staff, and made available before or during lectures.

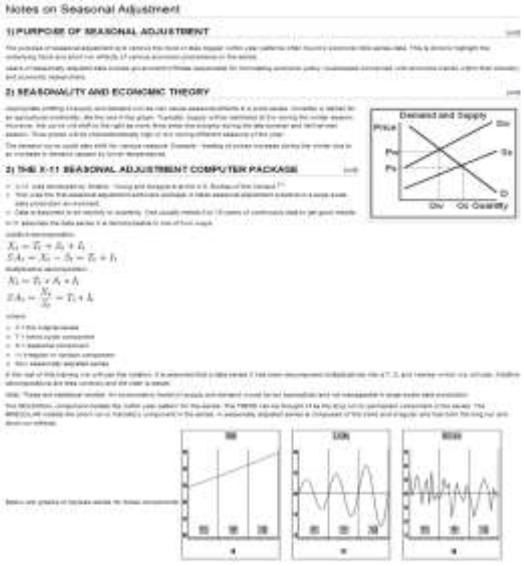


Tiny tutorial for Matlab

```
% 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=B./A          % element by element division
B=zeros(3,3)    % special definition of a matrix of zeros
B=ones(3,1)     % defines a matrix of ones
B=eye(3)        % 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



Notes on Seasonal Adjustment

1) PURPOSE OF SEASONAL ADJUSTMENT

2) SEASONALITY AND ECONOMIC THEORY

3) THE X-11 SEASONAL ADJUSTMENT COMPUTER PACKAGE

Overview of multifactor productivity and index numbers

Topics

Basics of Economic Indexes

Basics of Production Functions

$Y = f(X_1, X_2, \ldots, X_n, t)$

Specifications of the Form of the Production Function

Econometrics vs. Growth Accounting

The Growth Accounting Approach: Mathematical assumptions about the production function

Growth Accounting I: The Math

Figure 3: Three pages with training materials available on Statipedia

3.4 Library of statistical agency activities – past, present, and future

We collect information on our past and current statistical offices, programs and their activities such as press releases. We collect information on seminars and conferences relevant to Statipedians, especially in the Washington, DC area.

Statipedia has a collection of information on past surveys and censuses conducted by government agencies and other institutions around the world. All this can be organized as a glossary or encyclopedia. As part of their intellectual capital the U.S. statistical agencies can thus have access to a joint library and archive of such information.

[3] This argument has been made by the Open Source Practices team report (2008) and by Meyer and Buszuwski (2010).

Administrative information from other countries gives us perspectives on alternative ways of doing things. Other governments collect different information from which we can get findings and results that we do not create for ourselves and our own populations.

Statipedia has a special category for comparing procedures across agencies, called “At the agencies” pages, where staff from different agencies may describe their practices or procedures.

3.5 Organizing by categories and in bibliographies

As on Wikipedia, a page on Statipedia may be in many formal “categories” and users can quickly see lists of categories, subcategories, pages within a category. Example categories include “BLS,” “Argentina,” and “survey methodology.” A user can quickly assign pages into a category or remove them from a category. Some pages describe source materials with a summary and example citation. Thus some categories are themselves like bibliographies.

3.6 Developing new materials

The categories discussed above have information that is already established within the statistical system. On the wiki we can also synthesize works of new knowledge. We can gather customer questions and statements of critical perspectives on our work, and collect and rephrase our answers to them. We can collect research questions and address them with experimental findings, and draft academic papers. We can collect information on new (hot topics together and figure out what to do.

4. Growth over time and anticipated effects

Statipedia and similar platforms can have meaningful effects on knowledge management within government, and result in more efficient scientific communities. The presence of such platforms makes easier the development of shared source material, and expands the number of reference points, or points of view, who examine it and have a chance for input. We hope it will improve mutual awareness and peer review across disciplines and organizations. It can improve our community’s skills with open-source tools of growing relevance. It can reduce duplication of effort, and specialists within the agencies can serve a larger audience; users can find them on this platform. Such platforms can ease training and turnover.

4.1 Discoverability versus dissemination

"Dissemination of information" sounds like a good thing, but it can lead to being overwhelmed by overwhelming amounts of email, with large attachments. The receiver may wish to read them at some point in the future. A wiki offers something different from dissemination: "discoverability". Using such a system, information from the various documents can be put in a shared space where users can find it when they are ready to look for it and use it.

4.2 Working the wiki way

Wikis have a culture in which text and ideas are cited in a fragmentary way; this contrasts to a world of government statisticians in which documents are complete and authored. Some pages may appear to look like finished articles but normally this arises from a patchwork of changes over time. It will take time to become used to the fragmentary scheme.

It is routinely reported that 90% of the users of a wiki do not contribute to it, and only a few contribute a lot. This is to be expected on Statipedia also. Hopefully people who benefit from the works of others can find them however.

Statipedia is not intended to deliver social networking functions (like govloop or facebook for example. It is possible to store pictures and other media but not easy, and it is not meant for sending messages.

4.3 Morale effects

Statistical staff in government often feel constrained by various limitations. For example, they often do not know the practices of other agencies, and feel it is not worth the trouble to find out. But taken together, the agencies have vast, diverse expertise and capability, economies of scale and scope, knowledge of data, and great computer resources. Platforms like Statipedia can make it possible to benefit from our large scale without requiring a formal reorganization.

4. Aspects that could be better

It takes days to get a login to Statipedia, and it takes time to log in; then after 15 minutes of activity one is logged out. In the future we hope to have quicker and more automatic login procedures for government staff using government computers.

Many users would like a text editor which fits the content better. One is under development by the Wikimedia Foundation which runs the Wikipedias, and we anticipate being able to use it in late 2012.

5. Governance and principles of administration

Statipedia's administrators have learned from other wiki installations to follow certain principles:

- Benefits rise as communities of interest emerge
- Meet open technical standards, such as HTML, TeX.
- Copy the procedures and designs used on other wikis. There is a co-evolution of the community, the technology, and the platform. That is, tomorrow's community, technology, and platform are functions of today's, but are not the same.
- Our job is to serve and empower staff, invite voluntary participation, but not to require them or force them to use the platform.
- Our platform is scientifically oriented, so we want to encourage and enable users to (a) anchor their discussions to sources, evidence, theory; (b) enable readers to drill down toward these sources of information; and (c) write for broad, open audiences, not organization-specific ones or insiders to a particular field.

5.1 Rules and guidelines of user conduct

Rules of conduct are online. We keep such statements limited since we do not see misbehavior yet. The core rules are: (a) Do not post sensitive information. Such as private information about other individuals, early/pre-release/sensitive data, or computer passwords; (b) Do not publish Statipedia content outside Statipedia without the permission of its authors. (Generally the information should be treated like information emailed to you); (c) Respect the rules of the EPA portal: In short, this is a U.S. federal computer system, with rules like other such systems, and our work here is tracked; (d Uuse professional language and conduct.

There are also guidelines, similar to those of Wikipedia:

- Assume the good intentions of others. Understand that they will change the content too.
- Boldly add new content or correct mistakes.
- Write in plain language
- cite and link to source materials, even if they are not accessible to all users. The source material does not need to be available to all users; an inaccessible named source is better than no source; listing a person as a contact to the source is better than no source.
- Clearly distinguish between statements of opinion and statements of fact; after more experienced, we will have a formal approach to this.

6. Conclusions

Statipedia is a wiki for statistical staff across U.S. Federal agencies, with methodology source material. It is a pilot project, not for the public. It is not for sensitive content. Its materials are growing and I expect it to be useful in the long run to the staffs of the statistical agencies at large.

Acknowledgements

Much of this has been developed with my collaborators Michael Messner (of the EPA), James Buszuwski (of BLS), BLS's Open Source Practices Team. We have had the support of Barry Nussbaum (EPA) and several managers at BLS. I learn from Wikipedians a lot.

References

Allen, Robert C. 1983. Collective invention. *Journal of Economic Behavior and Organization* 4: 1-24.

Cohen, Wesley M., and Daniel Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 128-152.

David, Paul A. 1998. Common Agency Contracting and the Emergence of "Open Science" Institutions. *American Economic Review* 88(2), 15-21.

Hess, Charlotte and Elinor Ostrom. 2006. Introduction. In Charlotte Hess and Elinor Ostrom, eds. *Understanding Knowledge as a Commons*. MIT Press.

Meyer, Peter B., and James A. Buszuwski. 2010. Statipedia: a platform for collaboration across statistical agencies. Federal Conference on Statistical Methodology paper. http://www.fcsm.gov/events/papers2009.html

Meyer, Peter B.; James A. Buszuwski; Jean Fox; Daniel Murphy; Curtis Reid; Daryl Slusher; Mark Thomas; and Elliot Williams. 2008. Open Source Practices Team Report. Bureau of Labor Statistics internal report.

Meyer, Peter B. 2007. Network of tinkerers: a model of open-source innovation. U.S. Bureau of Labor Statistics working paper 413. http://www.bls.gov/ore/pdf/ec070120.pdf

Pénin, Julien. Open source innovation: Towards a generalization of the open source model beyond software, Revue d'économie industrielle (forthcoming).

von Hippel, Eric. 2006. *Democratizing innovation*. MIT Press.