FCSM



# RESEARCH CONFERENCE

**December 1-3, 2015**

**Washington, DC**

# Abstract Submission Template

The conference provides a forum for experts from around the world to discuss and exchange current research and methodological knowledge relevant to statistical programs. **Papers must be original and not previously published or disseminated.**

**Deadline:** Submit the abstract via email to [**fcsm.conf@gmail.com**](mailto:fcsm.conf@gmail.com) by **May 5, 2015**. Late submissions will not be accepted.

Please note that authors/ presenters are responsible for their travel expenses, but the registration fee is waived for one presenter per paper.

**1. Author Contact Information**

**Primary Author Information**

Please provide the following information for the primary author:

|  |  |
| --- | --- |
| **First Name** | Lars |
| **Last name** | Vilhuber |
| **Organization** | Cornell University |
| **Mailing address** | 352 East Ives Hall, Ithaca, NY 14853 |
| **Country** | USA |
| **Phone** | 607-330-5743 |
| **Email** | Lars.vilhuber@cornell.edu |
| **Fax** |  |

**Secondary Author(s) Information**

Please provide the full name, affiliation, and email for each co-author, if applicable. Add additional rows if necessary.

|  |  |
| --- | --- |
| **Name and Affiliation** | **Email** |
| Benjamin Perry, Cornell University | bap63@cornell.edu |
| Venkata Kambhampaty, Cornell University | vkambhampaty@cornell.edu |
| Kyle Brumsted, McGill University | kjb245@cornell.edu |
| William C. Block, Cornell University | block@cornell.edu |

**2. Topic Area:**

### (Select from list: Survey design and data collection, Evaluation, Cross-cutting topics, Estimation and Analysis, Other – please specify)

|  |
| --- |
| Other – data dissemination, metadata standards |

**3. Keywords (enter up to 5)**

|  |
| --- |
| **DDI, metadata, crowdsourcing, dissemination** |

**4. Briefly describe the research hypothesis or problem your presentation addresses (limit 120 words)**

|  |
| --- |
| Recent years have shown the power of user-sourced information evidenced by the success of Wikipedia and its many emulators. Agencies on the other hand publish metadata on their data products, but have little opportunity to get structured feedback on that metadata (codebooks) from their users. |

**5. What methods and/or data are used to answer the question or solve the problem? (limit 120 words)**

|  |
| --- |
| As part of our Comprehensive Extensible Data Documentation and Access Repository (CED²AR) infrastructure, we demonstrate a prototype of crowdsourced metadata (codebooks), using DDI-C and supplemental XML. The system allows for any number of network connected instances (web or desktop deployments) of the CED²AR DDI editor to concurrently create and modify metadata. The backend transparently handles changes, and frontend has the ability to separate official edits (by designated curators of the data and the metadata) from crowd-sourced content. Features of the software allow official curators to merge in select enhancements into revisions of official documentation. CED²AR ingests and publishes widely-used DDI metadata standards, and can thus easily augment existing metadata publishing systems. |

**6. Briefly describe what the audience will take away from your presentation (how is it applicable to other researchers?). (limit 120 words)**

|  |
| --- |
| Creating and augmenting metadata is a labor-intensive endeavor. Harnessing collective knowledge from actual data users can supplement officially generated metadata. Using standards-compliant ingest and publication methods, our proposed crowd-sourced metadata system can greatly enhance user acceptance of documentation, user-engagement in the documentation process, and reduce agency costs while improving the quality of data documentation. This development is part of the NSF-Census Research Network node at Cornell University. |

**7. Please indicate the submission type:**

\_\_\_\_ Contributed paper

\_\_\_\_ Organized session paper\* (please list other papers included in the proposed session below)

\_\_x\_\_ Technical software demonstration\*\*

**\*Organized sessions**: Proposals for organized sessions should include 3 papers and a discussant OR 4 papers around a topical area but should represent more than one agency and/or program. Complete the template for each paper in the proposed session. The session organizer should email the abstract files together in one email to FCSM. Also, include the session title, session description, and discussant name (if applicable) in your submission.

**\*\*Technical software demonstration:** There will be an opportunity to present brief demonstrations of new software or data access tools. Note that these should be informational demonstrations and not marketing displays.

**8. Abstract (limit 300 words):**

Please use the following formatting instructions when preparing your abstract:

Font type: Arial 12pt.

Font style Title – bold type

Text – regular type

Case: Upper and lower-case letters for all headings and text

Line spacing: Single

Paragraph format: No initial indent, use blank line between paragraphs

Authors: Please list each author’s name under the abstract title followed by their affiliation in parentheses

### **Crowdsourcing Codebook Enhancements: A DDI-based approach**

Benjamin Perry (Cornell University), Venkata Kambhampaty (Cornell University), Kyle Brumsted (McGill University), Lars Vilhuber (Cornell University), & William C. Block (Cornell University)

Recent years have shown the power of user-sourced information evidenced by the success of Wikipedia and its many emulators. Agencies on the other hand publish metadata on their data products, but have little opportunity to get structured feedback on that metadata (codebooks) from their users. Creating and augmenting metadata is a labor-intensive endeavor. Harnessing collective knowledge from actual data users can supplement officially generated metadata. As part of our Comprehensive Extensible Data Documentation and Access Repository (CED²AR) infrastructure, we demonstrate a prototype of crowdsourced metadata (codebooks), using DDI-C and supplemental XML. The system allows for any number of network connected instances (web or desktop deployments) of the CED²AR DDI editor to concurrently create and modify metadata. The backend transparently handles changes, and frontend has the ability to separate official edits (by designated curators of the data and the metadata) from crowd-sourced content. Features of the software allow official curators to merge in select enhancements into revisions of official documentation. CED²AR ingests and publishes widely-used DDI metadata standards, and can thus easily augment existing metadata publishing systems. Using standards-compliant ingest and publication methods, our proposed crowd-sourced metadata system can greatly enhance user acceptance of documentation, user-engagement in the documentation process, and reduce agency costs while improving the quality of data documentation. This development is part of the NSF-Census Research Network node at Cornell University.