Report on Dagstuhl Workshop on Managing Metadata for Longitudinal Data - Best Practices

William C. Block & Stefan Kramer
Cornell Institute for Social and Economic Research (CISER)

Cornell University

- Two previous rounds of working paper authoring at Schloss Dagstuhl in week-long working meetings:
  - 2008: *Best Practices Across the Data Life Cycle*
    http://www.ddialliance.org/resources/publications/working/bestpractices
  - 2009: *Use Cases*
    http://www.ddialliance.org/resources/publications/working/usecases
- From Oct. 17-22, 2010, the topic was **The Data Documentation Initiative (DDI) Standard: Managing Metadata for Longitudinal Data— Best Practices** (http://www.dagstuhl.de/10422)
- Participants in the Dagstuhl 2010 workshop were chosen based on background in a longitudinal data project, DDI expertise, or experience with working on prior DDI Alliance working papers

CISER

Participants in the 2010 Dagstuhl workshop:

- Christian Bilde Andersen, Danish Data Archive (DDA)
- Randy Banks, Institute for Social and Economic Research (ISER), University of Essex
- Bill Block, Cornell Institute for Social and Economic Research (CISER), Cornell University
- Daniel Bontempo, Life Span Institute, University of Kansas
- Tito Castillo, MRC Centre of Epidemiology for Child Health, Institute of Child Health, University College London
- Vicky (Huey-Chi) Chang, Wisconsin Longitudinal Study, University of Wisconsin-Madison
- Benjamin Clark, London School of Hygiene and Tropical Medicine, Tazama Project, Tanzania
- Arofan Gregory, Open Data Foundation (ODaF)
- Sue Ellen Hansen, Institute for Social Research, Survey Research Operation, University of Michigan
- Stan Howald, Wisconsin Longitudinal Study, University of Wisconsin-Madison
- Larry Hoyle, Institute for Policy and Social Research, University of Kansas
- Jeremy Iverson, Algenta Technologies
- Uwe Jensen, GESIS - Leibniz Institute for the Social Sciences
- Douglas Kieweg, Center for Biobehavioral Neurosciences in Communication Disorders (BNCD), University of Kansas
- Neeraj Kumar Kashyap, Vadu Rural Health Program, KEM Hospital Research Centre
- Stefan Kramer, Cornell Institute for Social and Economic Research (CISER), Cornell University
- Hilde Orten, Norwegian Social Science Data Archive (NSD)
- Denise Perpich, Language Acquisition Studies Lab, University of Kansas
- Barry Radler, Institute on Aging, University of Wisconsin-Madison
- Ingo Sieber, German Institute for Economic Research (DIW) Berlin, Socio-Economic Panel Study (SOEP)
- Mary Vardigan, Inter-university Consortium for Political and Social Research (ICPSR)
- Johanna Vompras, University Bielefeld Library, Germany
- Joachim Wackerow, GESIS - Leibniz Institute for the Social Sciences
- Knut Wenzig, National Educational Panel Study (NEPS), University of Bamberg
- Wolfgang Zenk-Möltgen, GESIS - Leibniz Institute for the Social Sciences

CISER

- Monday and Tuesday morning: presentations from longitudinal data projects
- Actual Best Practices paper topics were chosen by all participants during the workshop on Tuesday afternoon, who formed one working group for each topic:
    - *Documenting a Wider Variety of Data* working group. Chair: William C. Block.
    - *Longitudinal Variable Comparison* working group. Chair: Sue Ellen Hansen
    - *Metadata for the Longitudinal Data Lifecycle* working group. Chair: Larry Hoyle
    - *Presenting longitudinal studies to end users* working group. Chair: Stefan Kramer
- The final title of each paper coming out of each group may be different from the "working title" of the group.
- The following four slides excerpt the problem statement/description from each paper's draft on the internal DDI Alliance wiki for the workshop.

Documenting a Wider Variety of Data using the DDI

"This paper looks at the growing variety of data sources in research that are not traditional question-based surveys, and how these can be usefully documented. These data are increasingly being linked with data collected from more traditional surveys, to bring multi-disciplinary perspectives to bear on research questions. This phenomenon is not specific to longitudinal studies, but is a common issue in longitudinal contexts.

….

This paper provides guidance on expanding the capability of the Data Documentation Initiative (DDI) standard to document a wider variety of data resources and suggests improvements that may be incorporated into the DDI standard in future versions."

Authors: Christian Bilde Andersen, William C. Block , Daniel E. Bontempo, Arofan T. Gregory, Stan Howald, Douglas Kieweg, Barry T. Radler

Longitudinal Variable Comparison

"Producers and users of longitudinal data must be able to compare data produced across repeated data collection over time. They also need to be able to evaluate whether repeated observations taken over time believed to be the same are equivalent. This paper proposes best practice for the use of DDI to ensure that there are appropriate metadata to produce documentation to meet these needs."

Authors: Sue Ellen Hansen, Jeremy Iverson, Uwe Jensen, Hilde Orten, and Johanna Vompras

Metadata for the Longitudinal Data Lifecycle

"For this paper longitudinal studies are considered to be those where generation of data and metadata is repeated over time. The data also will include some time dimension. The overall study might involve multiple waves either for a person or population, or might involve ongoing continuous data collection. Some of the issues that are unique to longitudinal studies follow from the repetitive nature of their data collection. Other issues arise simply due to the extended period over which they are conducted, leaving more opportunity for unanticipated events. It is important to realize that studies which are not initially intended to be longitudinal may evolve into longitudinal studies. It is therefore best practice for all studies to structure initial metadata to be compatible with this potential repurposing."

Authors: Fortunato Castillo, Benjamin Clark, Larry Hoyle,
Neeraj Kashyap, Denise Perpich, Joachim Wackerow, Knut Wenzig

Presenting Longitudinal Studies to End Users Effectively Using DDI

"Longitudinal studies are complex and present unique challenges in documenting and delivering data to end users on the Web. Data and metadata from longitudinal studies can be presented in a variety of ways, and there are currently no commonly accepted standards for providing information that users need. It is important to assist prospective users in exploiting the longitudinal data resource effectively.
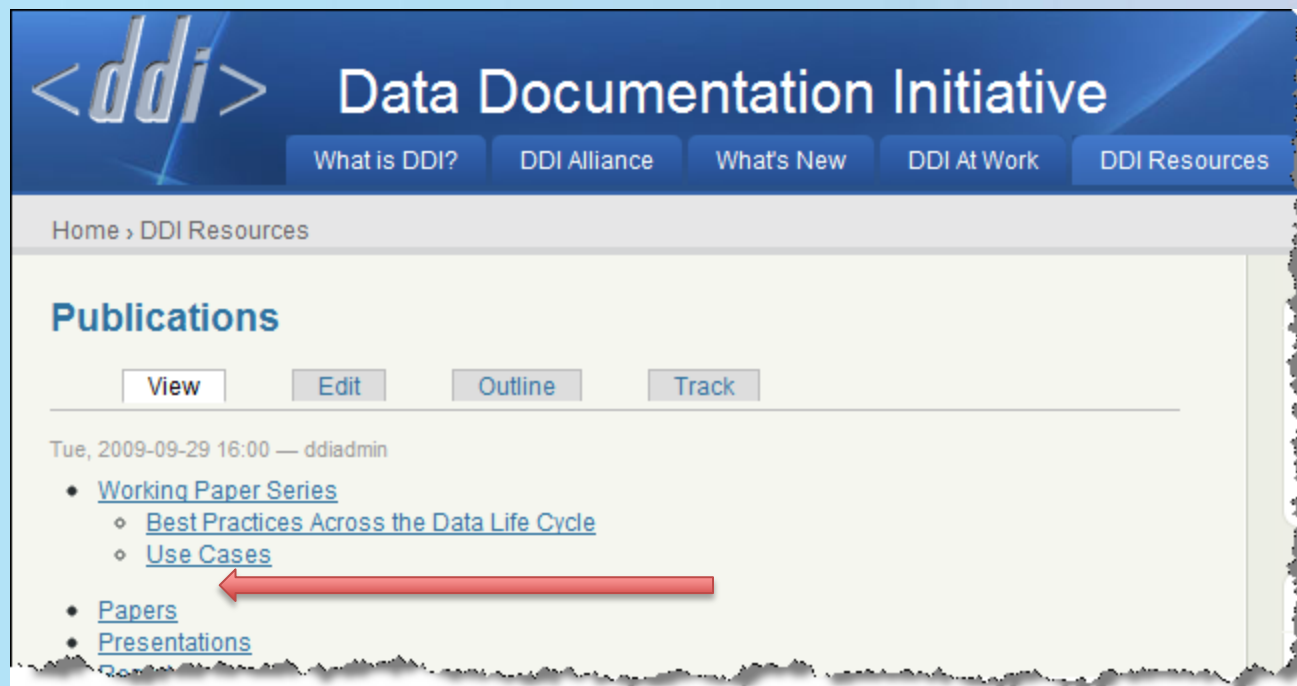
….

This paper is intended to provide implementers and those delivering longitudinal data with recommendations on how to use DDI most effectively to support the presentation of longitudinal studies, most commonly on the Web, and to describe best practice for structuring DDI instances. In addition, the paper suggests improvements to DDI to better support the unique aspects of longitudinal data."

Authors: Randy Banks, Vicky Chang, Stefan Kramer, Ingo Sieber, Mary Vardigan, Wolfgang Zenk-Möltgen

The final papers are expected to be published via the DDI Alliance web site (http://www.ddialliance.org/resources/publications):

At the end of the workshop (on Friday), the participants brainstormed possible future workshop topics

Over three dozen suggestions came up … including:
- Semantic statistics (DDI, SDMX, the semantic web…)
- Longitudinal data, the next step
- Qualitative data

- Data citation
- Delivery of metadata along with data files by data vendors/providers
- DDI for Data Management Plans
- DDI for preservation (relationship to PREMIS, etc.)
- Issues that cannot be addressed within current version of DDI
- Use of metadata in access control, intellectual property of datasets
- Confidentiality of, rights to metadata per se
- Use of DDI with confidential data, embargoing data releases, data curation costs
- Using DDI to drive and validate a process flow within a project

Thank you for your time & attention!

William C. Block

block@cornell.edu

Stefan Kramer

stefan.kramer@cornell.edu