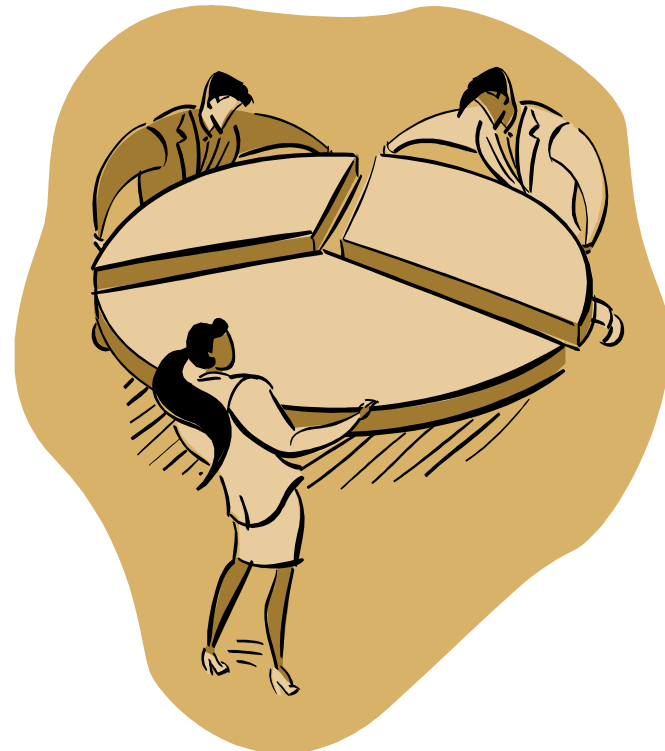




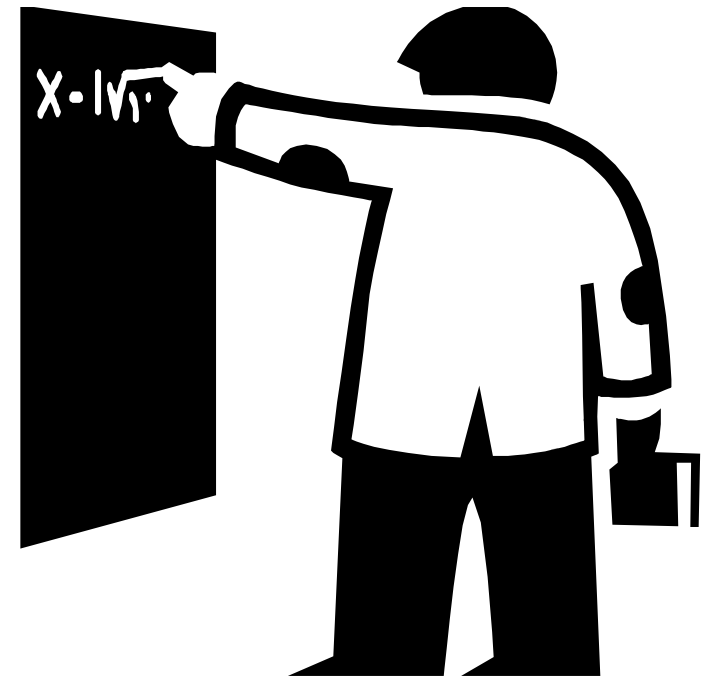
Commercial search engine  
developers and universities: a  
critical time for collaboration in  
the coming age of publicly  
accessible research data

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





- The past and the “recent present”:
  - In universities, researchers have had fairly little incentive to preserve and share their research data with others in the long term, aside from altruism
    - Not to be confused with temporary data collaboration during a project
  - Data sharing and publishing largely not a recognized scholarly activity for the tenure process



# Wolfram DATA SUMMIT 2011



Cornell University  
Cornell Institute for Social and Economic Research

## The potential game changer: requirements by research funding organizations

The image shows a screenshot of the National Science Foundation (NSF) website. The main content area is titled "Dissemination and Sharing of Research Results" and "Announcing a New Grant Program: Digital Humanities Implementation Grants". The text discusses the NSF Data Sharing Policy and the Digital Humanities Implementation Grants (DHIG) program. It mentions that proposals submitted on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". It also states that the DHIG program is designed to fund early, cutting-edge, experimental work that shows great potential. The text is written by Brett Bobley on 6/22/2011 12:57 AM.

Now making research data available becomes *non-optional*

Also new of note: the DHIG program requires that every applicant include both a sustainability plan and a data management plan. You'll note that we have aligned our data management requirements with those of the National Science Foundation to enable you to take advantage of data management resources that your institution may have already developed for applying to the NSF. Please check out the guidelines for more information on this. Over the course of the next year, we will be adding similar requirements to the other ODH grant programs.



# Wolfram DATA SUMMIT 2011



Cornell University  
Cornell Institute for Social and Economic Research

Many research universities have their institutional repositories, chiefly for texts, images, video, audio...

 Cornell

 The HKU Scholars Hub

 The University of Hong Kong

 Australian National University

 HEINRICH HEINE  
UNIVERSITÄT DÜSSELDORF

 UNIVERSITÄTS- UND LANDESBIBLIOTHEK

HHU Düsseldorf → Publikationsservice

**Düsseldorfer Dokumenten- und Publikationsservice**

Hier finden Sie **Hochschulschriften** und andere Dokumente im Volltext, insbesondere aber auch **multimediale Lehr- und Lernmaterialien** wie **Animationen, Simulationen, Audio- und Videomaterial** in verschiedenen Formaten. Der Düsseldorfer Publikationsservice bietet Studierenden und anderen Interessierten einen systematischen und einheitlichen Zugang zu an der Heinrich-Heine-Universität Düsseldorf erstellten oder genutzten digitalen Dokumenten.

**Hinweis:** Ihr Browser muss **Cookies** von diesem Server akzeptieren, damit Sie alle Funktionen nutzen können.

Suche

das Dokument mit der ID  
den Dateibereich mit der ID  
die Person mit der ID  
urn:nbn:de

Erweiterte Suche...  
Personen von A-Z...

 HKU  
ioc

**Hub**

e institutional repository of The University of Hong Kong (U's [Knowledge Exchange](#) initiative, The Hub strives to make their research very visible, with the goal of increasing

Provides exciting opportunities in research



**Hub**

tract and series fields.

... which some now (want to) adapt to accommodate *research data* ... or build separate repositories for that purpose

# Wolfram DATA SUMMIT 2011




Cornell University  
Cornell Institute for Social and Economic Research

**The good:** metadata standards for specific domains of research data do exist, and research universities are increasingly promoting/supporting their use by scientists ...


Biodiversity

For quick access, place your bookmarks here on the bookmarks bar. [Import bookmarks now...](#)

 **Ecological Meta**  
KNB Home Data People In

Ecological Metadata Language (EML) is a meta and for the ecology discipline. It is based on p associated efforts (Michener et al., 1997, Ecol XML document types that can by used in a mo data. Each EML module is designed to describ included with any ecological dataset.

For quick access, place your bookmarks here on the bookmarks bar. [Import bookmarks now...](#)

 **Data Documentation Initiative**

What is DDI? DDI Alliance What's New DDI At Work Controlled Vocabularies

Home

**What is DDI?**

Last Updated: Tue, 2011-05-17 09:48 — ddiadmin

The **Data Documentation Initiative (DDI)** is an effort to create an international standard for describing data from the social, behavioral, and economic sciences. Expressed in [XML](#), the DDI metadata [specification](#) now supports the entire research data life cycle. DDI metadata accompanies and enables data conceptualization, collection, processing, distribution, discovery, analysis, repurposing, and archiving.

Three examples, from  
ecological, life, and  
social sciences

# Wolfram DATA SUMMIT 2011



Cornell University  
Cornell Institute for Social and Economic Research

... so research data is (or at least can be) documented in ways that do (or at least could) allow for “search” in data-targeted ways:

← → ↻ ⌂ knb.ecoinformatics.org/knb/metacat ☆ 🔍

For quick access, place your bookmarks here on the bookmarks bar. [Import bookmarks now...](#)

**Abstract:**

Bird banding data collected over 5 years at the Big Sur Ornithology Lab located in Andrew Molera State Park were analyzed for population trends.

**Keywords:**

• terrestrial

**Additional Information:**

Tenney, Chris. Mist-net banding at the Big Sur Ornithology Lab: a six-year summary w (Master's thesis, Columbia Pacific University).

**License and Usage Rights:**

no restrictions

**Geographic Coverage:**

Geographic Description:	Big Sur Ornithology Lab, Andrew Molera State Park			
Bounding Coordinates:	West:	-121.5833	degrees	
	East:	-121.5833	degrees	
	North:	36.0667	degrees	
	South:	36.0667	degrees	

**Temporal Coverage:**

Begin:	1993-01-01
End:	1997-01-01

**The Knowledge Network for Biocomplexity**

**ICPSR** | INTER-UNIVERSITY CONSORTIUM FOR  
POLITICAL AND SOCIAL RESEARCH

**Smallest Geographic Unit:** municipality

**Geographic Coverage:** [Global](#), [Netherlands](#)

**Time Period:** • 1971--2006

**Date of Collection:** • 1971--2006

**Unit of Observation:** individual

**Universe:** Members of the Dutch electorate at the time of the 1971-2006 parliamentary elections.

**Data Types:** survey data

**Methodology**

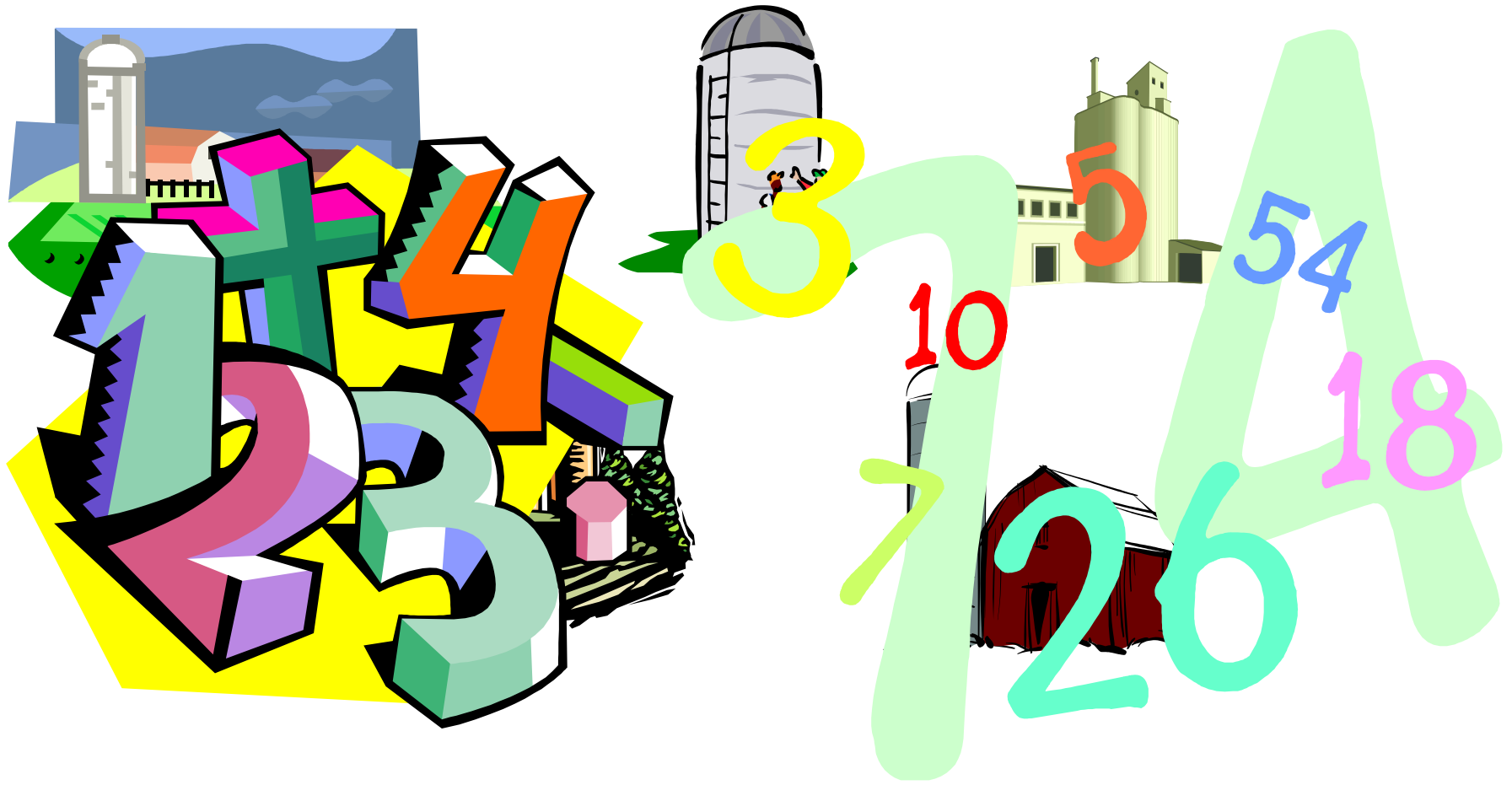
Two examples, from  
ecological and from  
social sciences

# Wolfram DATA SUMMIT 2011



Cornell University  
Cornell Institute for Social and Economic Research

**The bad:** as research data sharing becomes mandatory, university departments, libraries, research centers, associations, consortia, etc. will likely create more and more (different) information silos for more and more (diverse) data



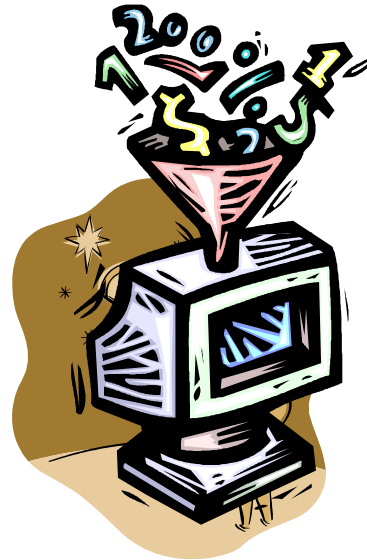
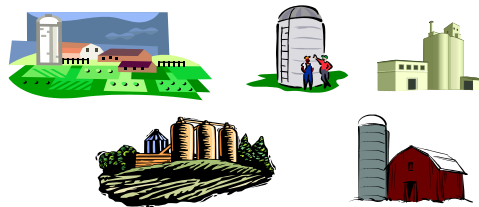
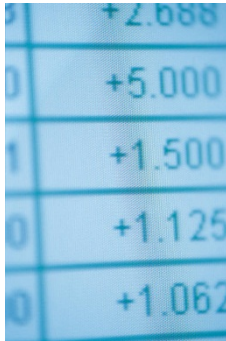


# Wolfram DATA SUMMIT 2011



Cornell University  
Cornell Institute for Social and Economic Research

**The ugly:** the *future* researcher, looking for data for repurposing, replication, ... won't and shouldn't have to know or care much about how this infrastructure came to be ...



... but will want cross-data-repository searching to be **easy, yet focusable**



# Wolfram DATA SUMMIT 2011



Cornell University  
Cornell Institute for Social and Economic Research

Who can create the solutions to these problems? Research-generating institutions (incl. universities) and search engine developers together



+



=





## Thank you for your time & attention!

Stefan Kramer  
stefan.kramer@cornell.edu

William C. Block  
block@cornell.edu

CORNELL UNIVERSITY LIBRARY  
eCommons@Cornell

[eCommons@Cornell](#) >  
[Cornell Institute for Social and Economic Research \(CISER\)](#) >  
[CISER Presentations and Papers](#) >

Browsing "CISER Presentations and Papers" by Title

Jump to: [0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)  
or enter first few letters:

Sort by:  In order:  Results/Page  Authors/Record:

Showing results 1 to 8 of 8

PREVIEW	ISSUE DATE	TITLE	AUTHOR(S)
	23-Sep-2010	<a href="#">Challenges and Opportunities in Social Science Research Data Management</a>	<a href="#">Kramer, Stefan</a>
	21-Mar-2011	<a href="#">Current Trends in Open Repositories &amp; Exemplary Local Solutions: European Perspective</a>	<a href="#">Horstmann, Wolfram</a>
	7-Dec-2010	<a href="#">Data Librarianship: Past, Present, Future, Challenges, Opportunities</a>	<a href="#">Kramer, Stefan</a>
	9-Dec-2010	<a href="#">Development of the next DDI Tools Catalog</a>	<a href="#">Kramer, Stefan</a>
	22-Sep-2010	<a href="#">The Lifecycle of Social Science Research Data: Enabling Discovery through Metadata and Search Tools</a>	<a href="#">Block, William C.; Kramer, Stefan</a>
	28-Mar-2011	<a href="#">The Lifecycle of Social Science Research Data: Improved Discovery through better Metadata and Search Tools</a>	<a href="#">Block, William C.; Kramer, Stefan; Williams, Jeremy</a>
	12-Jun-2011	<a href="#">The new &amp; improved DDI Tools Catalog - for users and developers of tools supporting the Data Documentation Initiative metadata specification</a>	<a href="#">Kramer, Stefan; McNeill, Kate</a>
	11-Mar-2011	<a href="#">Report on Dagstuhl Workshop on Managing Metadata for Longitudinal Data - Best Practices</a>	<a href="#">Block, William C.; Kramer, Stefan</a>

Look for this presentation in:  
<http://ecommons.library.cornell.edu/handle/1813/17391>