Exploring Ethics in Data and Technology Research

MSU Library

Sara Mannheimer
Scott W. H. Young
Jason Clark
Justin Shanks

Computer Science Seminar Series
February 22, 2016

1. Twitter Sentiment Analysis

2. Personalized Web Search

3. Semantic Web Identities

Ethical Questions

Can vs. should

Personalized services vs. user privacy

Who benefits?

Twitter Sentiment Analysis

Sentiment Analysis + Data Sharing

Identify researchers in 16 disciplines

Pull data from Twitter API

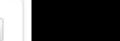Isolate data-related tweets

Code sentiment using Mechanical Turk

Zeynep Tufekci
@zeynep

Newsweek hired company to do "sentiment analysis" on #Gamergate and declared 90% of tweets were "neutral" when, um, truth was "can't tell".

8:05 AM - 28 Oct 2014

14 RETWEETS 10 FAVORITES

Zeynep Tufekci
@zeynep

1-"Sentiment analysis" on Twitter is hard; only 140. 2-
Five min. on the hashtag is enough to realize 90%
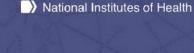neutral is, um, face value absurd.

8:08 AM - 28 Oct 2014

4 RETWEETS 8 FAVORITES

NIH> *Data Science at NIH*

About BD2K

The ability to harvest the wealth of information contained in biomedical Big Data will advance our understanding of human health and disease; however, lack of appropriate tools, poor data accessibility, and insufficient training, are major impediments to rapid translational impact. To meet this challenge, the National Institutes of Health (NIH) launched the Big Data to Knowledge (BD2K) initiative in 2012.

What is Biomedical Big Data?

On this Page...

○ **BD2K Mission Statement:**

● BD2K Program Management Working Groups:

● BD2K Organization & Coordination:

"The ability to harvest the wealth of information contained in biomedical Big Data will advance our understanding of human health and disease."

— NIH Big Data to Knowledge

"Numbers are seductive because they look like answers."

— Nathan Jurgenson

Ethical concerns for social media (big) data research

Perceived Publics

Informed Consent

Digital Labor

Jeffrey Keefer, PhD @JeffreyKeefer   27 Sep

Perhaps people conclude that Internet research of public Tweets and such are fair game, even without the easy ability to anonymize?

#AoIR

Scott W. H. Young
@hei_scott   🐦 Follow

@JeffreyKeefer @bonstewart Good questions re: ethics of social data. I've taken Twitter/Instagram content as public & therefore reusable.

11:46 AM - 27 Sep 2013

1 FAVORITE   ↩ 🔁 ★

Anil Dash
on Jul 24

What Is Public?

It's so simple, right?
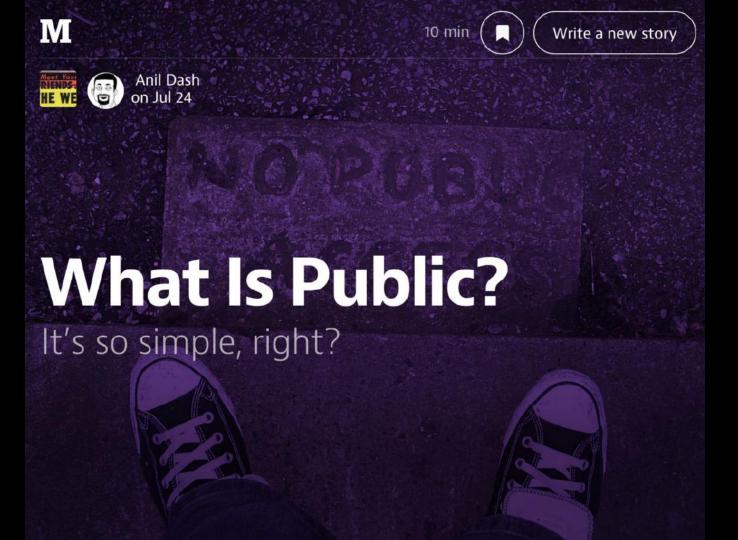
"But the data is already public": on the ethics of research in Facebook

Michael Zimmer

Social media for large studies of behavior

Large-scale studies of human behavior in social media need to be held to higher methodological standards

By **Derek Ruths**[1]* and **Jürgen Pfeffer**[2]

On 3 November 1948, the day after Harry Truman won the United States presidential elections, the *Chicago Tribune* published one of the most famous erroneous headlines in newspaper history: "Dewey Defeats Truman" (*1, 2*). The headline was informed by telephone surveys, which had inadvertently undersampled Truman supporters (*1*). Rather than permanently discrediting the practice of polling, this event led to the development of more sophisticated techniques and higher standards that produce the more accurate and statistically rigorous polls conducted today (*3*).

Now, we are poised at a similar technological inflection point with the rise of online personal and social data for the study of human behavior. Powerful computational resources combined with the availability of massive social media data sets has given rise to a growing body of work that uses a combination of machine learning, natural language processing, network analysis, and statistics for the measurement of population structure and human behavior at unprecedented scale. However, mounting evidence suggests that many of the forecasts and analyses being produced misrepresent the real world (*4–6*). Here, we highlight issues that are endemic to the study of human behavior through large-scale social media data sets

POLICY

different social media platforms (*8*). For instance, Instagram is "especially appealing to adults aged 18 to 29, African-American, Latinos, women, urban residents" (*9*) whereas Pinterest is dominated by females, aged 25 to 34, with an average annual household income of $100,000 (*10*). These sampling biases are rarely corrected for (if even acknowledged).

Proprietary algorithms for public data. Platform-specific sampling problems, for example, the highest-volume source of public Twitter data, which are used by thousands of researchers worldwide, is not an accurate representation of the overall platform's data (*11*). Furthermore, researchers are left in the dark about when and how social media providers change the sampling and/or filtering of their data streams. So long as the algorithms and processes that govern these public data releases are largely dynamic, proprietary, and secret or undocumented, designing reliable and reproducible studies of human behavior that correctly account for the resulting biases will be difficult, if not impossible. Academic efforts to characterize aspects of the behavior of such proprietary systems can provide details needed to begin reporting biases.

The rise of "embedded researchers" (researchers who have special relationships with providers that give them elevated access to platform-specific data, algorithms, and resources) is creating a divided social media research community. Such researchers, for example, can see a platform's inner workings and make accommodations, but may not be able to reveal their corrections or the data used to generate their findings.

REPRESENTATION OF HUMAN BEHAVIOR. *Human behavior and online platform design.* Many social forces that drive the formation and dynamics of human behavior and relations have been intensively studied and are well-known (*12–14*). For instance, homophily ("birds of a feather flock together"), transitivity ("the friend of a friend is a friend"), and propinquity ("those close by form a tie") are all known by designers of social media platforms and, to increase platform use and adoption, have been incorporated in their link suggestion algorithms. Thus, it may be necessary to untangle psychosocial from platform-driven behavior. Unfortunately, few studies attempt this.

Social platforms also implicitly target

Reducing biases and flaws in social media data

DATA COLLECTION

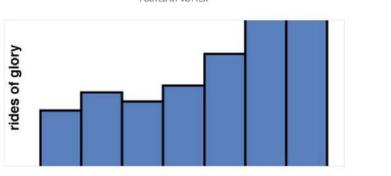- 1. Quantifies platform-specific biases (platform design, user base, platform-specific

#UBERDATA

RIDES OF GLORY

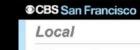MARCH 26, 2012

POSTED BY VOYTEK

Today we're going to get a little emotional. You know that Uber loves you and well, gosh, sometimes it's nice to think that you love us, too. But we know we're not the only ones in your life and we know that you sometimes look for love elsewhere. Well, while you're out loving other human beings, we #UberData nerds are cuddled up with our computers, loving math.

Local

Uber Crunches User Data To Determine Where The Most 'One-Night Stands' Come From

November 18, 2014 1:45 PM

(Photo by David Ramos/Getty Images)

Chanelle Bessette Contributor

I write about news and business trends in technology.

Opinions expressed by Forbes Contributors are their own.

FOLLOW

TECH   11/25/2014 @ 5:36PM   2,466 views

Does Uber Even Deserve Our Trust?

+ Comment Now   + Follow Comments

The Opinion Pages | OP-ED CONTRIBUTORS   💬 2 COMMENTS
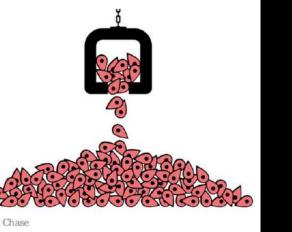
We Can't Trust Uber

By ZEYNEP TUFEKCI and BRAYDEN KING   DEC. 7, 2014

✉ Email

f Share

🐦 Tweet

📁 Save

➡ More

UBER, the popular car-service app that allows you to hail a cab from your smartphone, shows your assigned car as a moving dot on a map as it makes its way toward you. It's reassuring, especially as you wait on a rainy street corner.

Less reassuring, though, was the apparent threat from a senior vice president of Uber to spend "a million dollars" looking into the personal lives

Matt Chase

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer[a,1], Jamie E. Guillory[b,2], and Jeffrey T. Hancock[b,c]

[a]Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of [b]Communication and [c]Information Science, Cornell University, Ithaca, NY 14853

Emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. Emotional contagion is well established in laboratory experiments, with people transferring positive and negative emotions to others. Data from a large real-world social network, collected over a 20-y period suggests that longer-lasting moods (e.g., depression, happiness) can be transferred through networks [Fowler JH, Christakis NA (2008) *BMJ* 337:a2338], although the results are controversial. In an experiment with people who use Facebook, we test whether emotional contagion occurs outside of in-person interaction between individuals by reducing the amount of emotional content in the News Feed. When positive expressions were reduced, people produced fewer positive posts and more negative posts; when negative expressions were reduced, the opposite pattern occurred. These results indicate that

demonstrated that (*i*) emotional contagion occurs via text-based computer-mediated communication (7); (*ii*) contagion of psychological and physiological qualities has been suggested based on correlational data for social networks generally (7, 8); and (*iii*) people's emotional expressions on Facebook predict friends' emotional expressions, even days later (7) (although some shared experiences may in fact last several days). To date, however, there is no experimental evidence that emotions or moods are contagious in the absence of direct interaction between experiencer and target.

On Facebook, people frequently express emotions, which are later seen by their friends via Facebook's "News Feed" product (8). Because people's friends frequently produce much more content than one person can view, the News Feed filters posts, stories, and activities undertaken by friends. News Feed is the primary manner by which people see content that friends share.

Facebook

Facebook emotion study breached ethical guidelines, researchers say

Lack of 'informed consent' means that Facebook experiment on nearly 700,000 news feeds broke rules on tests on human subjects, say scientists
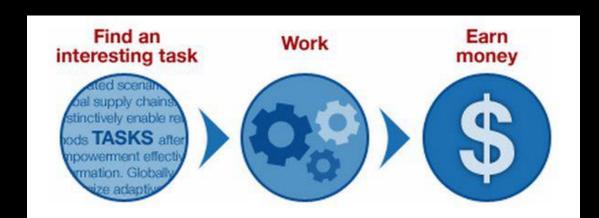
"[The work] was consistent with Facebook's Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research."

— Authors of the Facebook Emotional Contagion Study

"This study is a scandal because it brought Facebook's troubling practices into a realm—academia—where we still have standards of treating people with dignity and serving the common good."

— James Grimmelmann

Utpal Dholakia

My Experience as an Amazon Mechanical Turk (MTurk) Worker

I was surprised by the degree and variety of what I, as an MTurk worker, perceived to be unethical researcher behavior. In answering surveys, I was made to work much longer than promised, refused payment for work I had already done without sufficient explanation or recourse, and I was not provided with an appropriate method to protest or to receive redress.

"In most crowdsourcing platforms, fragmentation of tasks disenfranchises cognitive workers by disconnecting them from the final intellectual work."

—Ayhan Aytes

Guidelines for Academic Requesters

About the project

Version 1.0

"Treat your workers with respect and dignity. Workers are not numbers and statistics. Workers are not lab rats. Workers are people and should be treated with respect." - turker 'T', a Turkopticon moderator
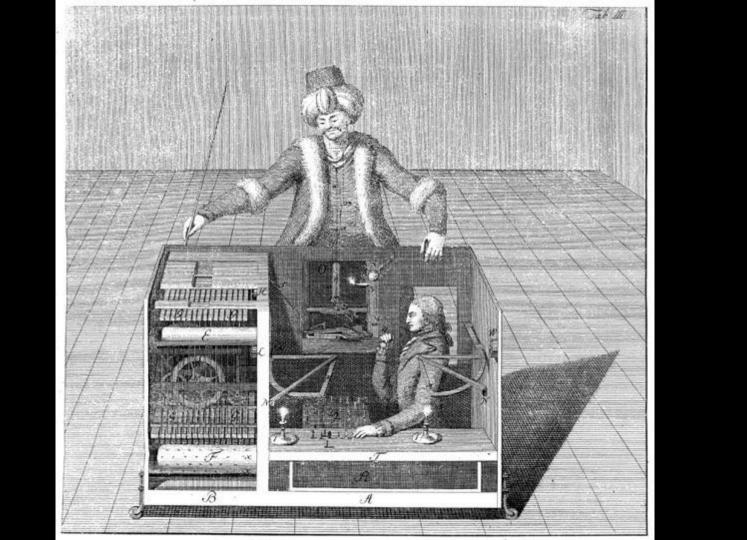
This document consists of this main **Guidelines for Academic Requesters** page, and several subpages with important additional details, which are referenced with "Read more" links at relevant points throughout this main page:

- Basics of how to be a good requester
- Fair payment
- Resources for communicating with requesters
- Links to other resources on AMT and online research ethics

Contents [hide]

http://wiki.wearedynamo.org/index.php/Guidelines_for_Academic_Requesters

Tab. III.

Is there a Ghost in the [Search] Machine?
Improving Search UX using Query Analysis and Machine Cues

Jason A. Clark
Associate Professor
Head, Library Informatics & Computing
Montana State University (MSU) Library

Overview

- What are the "ghosts" of the search act?
- Applying context and intention to Search UX
- Intention mining and anticipatory design in a search prototype
- Questions

What would you like to find?

🔍

- **Search the Catalog**   ○ **Search NYPL.org**

On now
Library for the Performing Arts

NYPL BLOG
In Memory of Harper Lee

JOHN SPOOR BROOME LIBRARY

RESEARCH   **SERVICES**   **COLLECTIONS**   **ABOUT**

Hello there. Looking for something?

Search Library Materials

| CHOOSE A SUBJECT ▾ | | **SEARCH** |

Didn't get the results you wanted? Try our Research page.

Many ways to research: Library Catalog   Research Guides by Subject   Databases A-Z   Journals & Newspapers A-Z   Interlibrary Loan   **Even More...**

● ● ● ○ ○ ○

BOOK A STUDY ROOM

HAPPENING IN THE LIBRARY

🕐 **Today's Hours:**   10:00am–6:00pm   more »

📅 **News & Events**

Blind Date with a Book

NCSU LIBRARIES

FIND   GET HELP   SERVICES   LIBRARIES   ABOUT

ASK US | MY ACCOUNT | HOURS | FAQ | LOG OUT | CHAT NOW 💬

Search All

About this search tool

laser

Search

Articles   Books & Media   Journals   Databases   Databases by Subject   Frequently Asked Questions   Our Website

Articles

Applied physics 1994
Applied Physics.
Journal

Optics & laser technology 1970
Optics & Laser Technology.
Journal

Journal of cosmetic and laser therapy
2001
Journal Of Cosmetic And Laser Therapy.
eJournal Full Text Online ☁

> See all 2,527,153 article results

Books & Media

**Laser Filamentation [electronic resource] :
Mathematical Methods and Models** 2016.
eBook
✔ Online TA1671 -1707 ebook

**First-stage LISA Data Processing and Gravitational
Wave Data Analysis [electronic resource] :...** 2016.
Wang, Yan.
eBook
✔ Online QB460 -466 ebook
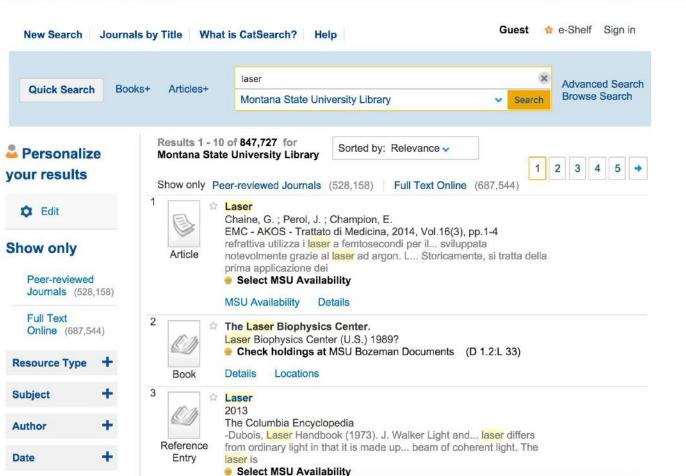
**Advances in MALDI and Laser-Induced Soft
Ionization Mass Spectrometry [electronic resource]**
2016.
eBook
✔ Online QD431 -431.7 ebook

> See all 15,640 book results

Guest ⭐ e-Shelf Sign in

Quick Search Books+ Articles+

laser ✕

Montana State University Library ⌄ Search

Advanced Search
Browse Search

👤 **Personalize
your results**

Results 1 - 10 of **847,727** for
Montana State University Library

Sorted by: Relevance ⌄

1 2 3 4 5 →

Show only **Peer-reviewed Journals** (528,158) | **Full Text Online** (687,544)

⚙ Edit

Show only

Peer-reviewed
Journals (528,158)

Full Text
Online (687,544)

Resource Type ➕

Subject ➕

Author ➕

Date ➕

1 ☆ **Laser**
Chaine, G. ; Perol, J. ; Champion, E.
EMC - AKOS - Trattato di Medicina, 2014, Vol.16(3), pp.1-4
refrattiva utilizza i laser a femtosecondi per il... sviluppata
notevolmente grazie al laser ad argon. L... Storicamente, si tratta della
prima applicazione dei
🟡 **Select MSU Availability**

Article

MSU Availability Details

2 ☆ **The Laser Biophysics Center.**
Laser Biophysics Center (U.S.) 1989?
🟡 **Check holdings at** MSU Bozeman Documents (D 1.2:L 33)

Book

Details Locations

3 ☆ **Laser**
2013
The Columbia Encyclopedia
-Dubois, Laser Handbook (1973). J. Walker Light and... laser differs
from ordinary light in that it is made up... beam of coherent light. The
laser is
🟡 **Select MSU Availability**

Reference
Entry

Librarians

the original search engine

MIND | BODY

The Concept of Mind

Gilbert Ryle

"Such in outline is the official theory. I shall often speak of it, with deliberate abusiveness, as "the dogma of the Ghost in the Machine." I hope to prove that it is entirely false, and false not in detail but in principle. It is not merely an assemblage of particular mistakes. It is one big mistake and a mistake of a special kind. It is, namely, a category mistake."

search query =
literal aspect + inferred aspect

"yellowstone national park"
[search query]

"yellowstone national park"
[literal aspect]

iPhone user, on MT highway
[tacit, inferred aspect]

Search Act => Context + Intention

What are the Ghosts in the Search Act?

How do we enhance services and keep privacy in place?

Cues from the "live" Query

```
$cueWords = array( "about", "above", "across", "after", "afterwards", "again",
"against", "all", "almost", "alone", "became", "because", "become", "becomes",
"becoming", "been", "before", "beforehand", "behind", "being", "below", "beside",
"besides", "between", "beyond", "by", "call", "can", "cannot", "cant", "could",
"couldnt", "cry", "describe", "do", "either", "except", "few", "fill", "find", "found",
"from", "front", "full", "further", "get", "give", "go", "had", "has", "hasnt", "have",
"how", "however", "hundred", "if", "in", "indeed", "interest", "into", "is", "keep",
"might", "mine", "more", "moreover", "most", "mostly", "move", "much", "must", "my",
"myself", "name", "never", "nevertheless", "nor", "not", "nothing", "now", "nowhere",
"of", "off", "often", "on", "once", "one", "only", "onto", "or", "other", "others",
"otherwise", "our", "ours", "ourselves", "out", "over", "own", "part", "per", "perhaps",
"please", "put", "rather", "see", "seem", "seemed", "seeming", "seems", "should",
"show", "side", "since", "so", "some", "somehow", "someone", "something", "sometime",
"sometimes", "somewhere", "still", "such", "system", "take", "temp", "temperature",
"than", "that", "the", "their", "them", "themselves", "then", "thence", "there",
"thereafter", "thereby", "therefore", "therein", "thereupon", "these", "they", "this",
"those", "though", "three", "through", "throughout", "thru", "thus", "to", "together",
"too", "top", "toward", "towards", "under", "until", "up", "upon", "us", "very", "time",
"were", "weather", "what", "whatever", "when", "whence", "whenever", "where",
"whereafter", "whereas", "whereby", "wherein", "whereupon", "wherever", "whether",
"which", "while", "whither", "who", "whoever", "whole", "whom", "whose", "why", "will",
"with", "within", "without", "would", "yet", "you", "your", "yours", "yourself",
"yourselves");
```

```
cueWords = ["about", "above", ... , "can", ...]
query = 'Can I get research help?';
if (in_array(strtolower($query), $cueWords)) {
    echo "Here are some people to talk to: ";
}


OR


# python example
cueWords = ["about", "above", ... , "can", ...]
query = 'Can I get research help?'
query.lower()
if query in cueWords:
    print Here are some people to talk to:
```

Cues from the Machine

HTTP Headers
+
Global Server Variables

en.wikipedia.org/wiki/List_of_HTTP_header_fields

Host: www.lib.montana.edu

Connection: keep-alive

Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8

User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_5) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/38.0.2125.122 Safari/537.36

Referer: http://www.lib.montana.edu/~jason/files.php

Accept-Encoding: gzip,deflate,sdch

Accept-Language: en-US,en;q=0.8

Cookie: is_returning=1; optimizelyEndUserId=oeu1413387638672r0.9280797478277236; optimizelySegments=%7B%22204658328%22%3A%22false%22%2C%22204728159%22%3A%22none%22%2C%22204736122%22% optimizelyBuckets=%7B%7D; s_fid=1162C3214F4A86F8-1B28DA485FC96756; s_lv=1413387991698; __atuvc=0%7C39%2C0%7C40%2C0%7C41%2C0%7C42%2C1%7C43; is_returning=1; __utma=55097015.1071264323.1398267937.1415934660.1415983050.7; __utmz=55097015.1415983050.7.7.utmcsr=googlelutmccn=(organic)lutmcmd=organiclutmctr=(not%20provided); _ga=GA1.2.1071264323.1398267937; __utmt=1; __utma=112212121.694734941.1395729614.1416383505.1416454676.267; __utmb=112212121.2.10.1416454676; __utmc=112212121; __utmz=112212121.1416383505.266.57.utmcsr=googlelutmccn= (organic)lutmcmd=organiclutmctr=(not%20provided)

Page URL: http://www.lib.montana.edu/~jason/files/digital-collections-custom-search-api/index.php

Referring URL: http://www.lib.montana.edu/~jason/files.php

IP Address: 153.90.170.5

Timestamp: Wednesday, November 19, 2014 8:38 pm

Location from the Browser or Network Info

```json
{
    "longitude": -110.9013,
    "latitude": 45.628,
    "asn": "AS13476",
    "offset": "-6",
    "ip": "153.90.170.5",
    "area_code": "0",
    "continent_code": "NA",
    "dma_code": "0",
    "city": "Bozeman",
    "timezone": "America/Denver",
    "region": "Montana",
    "country_code": "US",
    "isp": "Montana State University",
    "postal_code": "59717",
    "country": "United States",
    "country_code3": "USA",
    "region_code": "MT"
}
```

Figure: Screenshot of JSONLint view of values returned from Telize API

Javascript variables from the Browser Load

Am I Unique?

- 🏠 Home
- 🐾 My fingerprint
- 🕐 My timeline
- 🍰 Global statistics
- ❓ FAQ
- 🛡 Privacy policy
- ⚙ Privacy tools
- 🔗 Links
- ◯ About
- 🐙 View on GitHub

We reached **100,000 fingerprints**! We never thought we would reach such a high number! Thanks to all visitors for supporting our research!
We also launched an AmIUnique extension for 🦊 **Firefox** and 🌐 **Chrome** . More details can be found **HERE**.

Learn how identifiable you are on the Internet

Help us investigate the diversity of web browsers

View my browser fingerprint

By clicking on this button, only anonymous data will be collected and a cookie will be stored in your browser for four months. You can find more details in the Privacy Policy.

Spread the word! Share AmIUnique!
Try it on all your devices!

✉ 📘 🐦 🔵 📌 🟠 🔲

Javascript functions:

window.navigator.plugins
window.navigator.platform
window.navigator.cookieEnabled
window.navigator.doNotTrack
new Date().getTimezoneOffset()
window.screen.height
window.screen.width
window.screen.colorDepth

Cues from Analytics Logs

```
deer,
divergent,
indian,
avalanche,
dissertations,
cinnabar,
Election+1845,
Hudson%2C+B.,
montana+hospital+records+1927-1933,
yearbook,
JFK+assassination,
William+Lester+Stockton,
bison,
sounds+of+yellowstone,
dissertation,
Search+our+digital+collections...,
Search+our+digital+collections...,
service+learning,
bozeman+claims+association,
photosynthesis,
Dissertations,
milk+river,
bison+vocalization,
geothermal,
thunder,
swan,
dissertation,
burgess,
Reproductive+biology+of+brown+and+rainbow+trout+below+Hauser+Dam%2C+Missouri+River%2C+with+reference+to+proposed+hydroelectric+peaking,
yearbook,
bison,
leadership+philosophy,
ISSW,
dissertations,
jason+clark,
justin+spinelli,
jason+clark,
Lanny+Real+Bird,
dissertations,
```

arc.lib.montana.edu/digital-collections/search-log.txt

Benedict Evans
@BenedictEvans

All curation grows until it requires search. All search grows until it requires curation.

RETWEETS
584

LIKES
752

3:23 PM - 18 Dec 2015

Applying Contextual and Intentional Cues

Use these cues to show different parts of search index and additional context

This is anticipatory design.

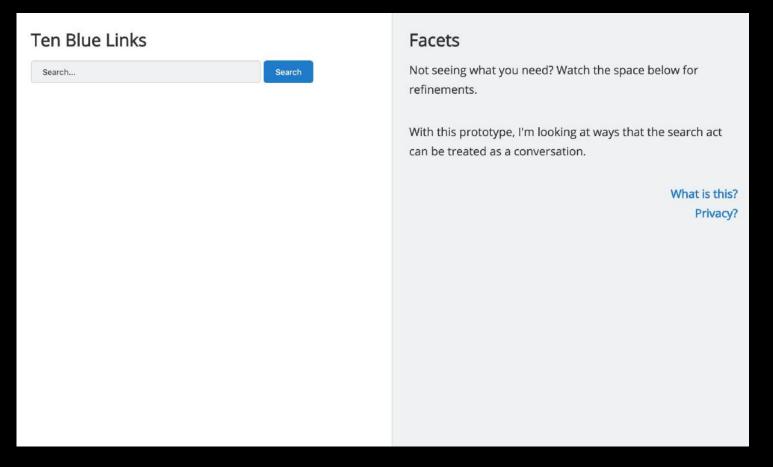This is intention mining.

Search Prototype

Demo
www.lib.montana.edu/~jason/files/search-ux/

Code
github.com/jasonclark/search-ux

Ten Blue Links

Search... **Search**

Facets

Not seeing what you need? Watch the space below for refinements.

With this prototype, I'm looking at ways that the search act can be treated as a conversation.

What is this?

Privacy?

Initial Search Landing Page

Small Result Set (< 30 results)

Large Result Set (> 30 results)

Suggestions Based on Natural Language Question

Ten Blue Links

| research help | | **Search** |

Sort: Relevance or Date

"research help" (About 846 results)

Ask the Library - Ask the Library
ask.lib.montana.edu/

Research and Instruction Services - MSU Library | Montana State ...
www.lib.montana.edu/services/research-and-instruction/

Services - MSU Library | Montana State University
https://www.lib.montana.edu/services/

Home - Psychology Research - Library Guides at Montana State ...
guides.lib.montana.edu/psy

Publication and Data Services - MSU Library | Montana State ...
https://www.lib.montana.edu/services/publication-and-data/

Facets

Not seeing what you need? Watch the space below for refinements.

▶ Categories
▼ "Hand-picked" Topics

A few common topics based on our human-created sitemap and an analysis of our search logs.

articles & research databases, borrow, research help, library hours, yellowstone, parking & directions, montana history, google scholar msu, montana code annotated, mla format 6th edition

▶ Topics by Location
▶ Popular
▶ People
▶ FAQ
▶ Offers

With this prototype, I'm looking at ways that the search act can be treated as a conversation.

Suggestions Based on Result Set Size

Suggestions Based on Local Context

Suggestions Based on Time and Date

Woman Stalked Across 8 Websites By Obsessed Shoe Advertisement

LAWRENCEVILLE, GA—Expressing her growing unease at repeatedly spotting the same picture and text lurking in the corners of her favorite webpages, local woman

Search and Surveillance Workshop Series

Search and Surveillance with Python

▶

Search & Surveillance is a multi-part workshop series that enables scholars from different disciplines to engage fundamental search algorithms that have come to govern the ways in which we access and navigate digital and analogue spaces. Algorithms like Google's Pagerank tend to be experienced through clean interfaces that imply the limitless possibilities of search while actually enforcing boundaries on digital environments through browsing history, links, and lexical patterns. As illustrated in issues ranging from advertising, to political movements, to academic disciplines, search has altered the organizing logic of nearly every type of social structure. This series is designed to seed research projects that examine the consequences of digital media by learning and interrogating the foundational assumptions of burgeoning algorithmic paradigms.

Our first workshop, led by Collin Jennings,

Quaero ergo sum

"I search, therefore I am"

There are Ghosts in the [Search] Machine.

What do we as a profession choose to do with them?

~~Search~~ Surveillance Engine
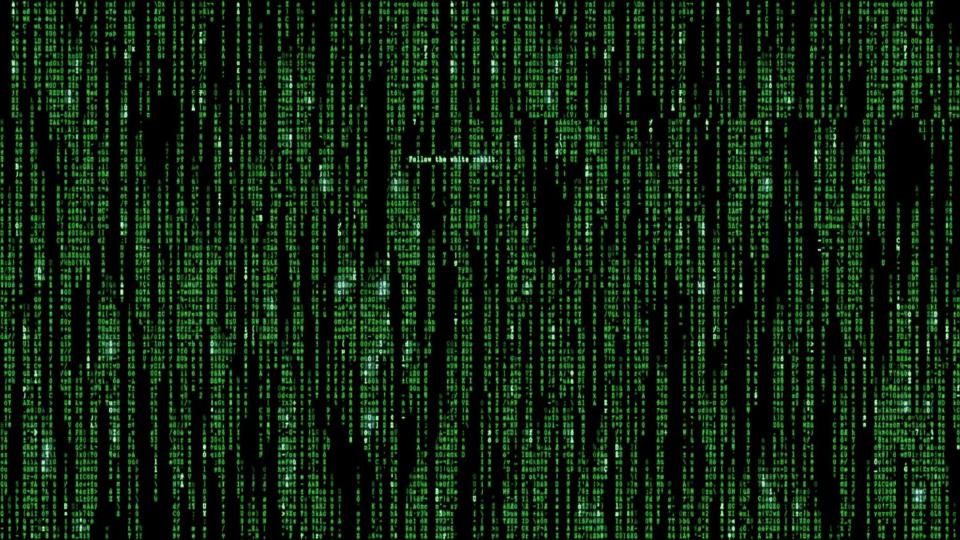
Research - Next Steps

- Systems design with domain experts (research librarians)

- Testing users with eye toward where tolerance for added value and privacy coincide

Semantics of Ethics
or
Ethics of Semantics

Follow the white rabbit.

Sir Timothy John Berners-Lee
b. 1955

Software Agents

"Properly designed, the Semantic Web can assist the evolution of human knowledge as a whole."

~TBL, 2001

Publications

Life Sciences

Cross-Domain

Social Networking

Geographic

Government

Media

User-Generated Content

Linguistics

Linked Datasets as of August 2014

JSON-LD document

JSON-LD data model courtesy of W3C @ goo.gl/AVRjWq

MACHINE LEARNING
POWER TO THE DATA

https://www.google.com/search?q=MSU+Library&biw=1513&bih=747&source=lnt&tbs=cdr%3A1%2Ccd_min%3A8%2F1%2F2013%2Ccd_max%3A9%2F15%2F2013&tbm=...

GOOGLE

montana state university library

Sign in

All   News   Maps   Images   Shopping   More ▾   Search tools

Aug 1, 2013 – Sep 15, 2013 ▾   Sorted by relevance ▾   All results ▾   Clear

The Writing Center - Montana State University
www.montana.edu/writingcenter/hours.html ▾ Montana State University ▾
Aug 9, 2013 - FINALS WEEK HOURS: Monday - Thursday, 9:00 a.m. - 3:00 p.m. – walk-ins only. Renne Library First Floor Commons Monday - Thursday: 9:15 a.m. - 10:00 ...

Dean's Office Administration - Montana State University
www.montana.edu/.../deans_staff.html ▾ Montana State University ▾
Aug 13, 2013 - He's been very involved in the University Library Committee, acting as that group's chair from 1995 to 2003 and again from 2013 to present. For many years, Dr.

[PDF] **Vita - MSU Library - Montana State University**
www.lib.montana.edu/~doralyn/cv.pdf ▾ Montana State University Library ▾
Aug 19, 2013 - Montana State University - Bozeman. Public Administration, 2008. M.S.L.S.. University of North Carolina - Chapel Hill. Library and Information Science, 1994.

Montana State University - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Montana_State_University ▾ Wikipedia ▾
Aug 13, 2013 - Montana State University (MSU) is a public university located in Bozeman, ... These included the 1949 Library Building (now Renne Library), the campus' first ...

Mascot: Champ (bobcat)   President: Waded Cruzado
Students: 15,688 (fall 2015)   Campus: University town; 1,170 acres ...

Subject Librarians | Contact Us | MSU Libraries
https://www.lib.msu.edu/.../subjectlibr... ▾ Michigan State University Libraries ▾
Sep 10, 2013 - Subject Librarians. The table below lists the name, office telephone number and e-mail address of the librarian who specializes in service for each subject area.

Hours | MSU Libraries
https://www.lib.msu.edu/hours/ ▾ Michigan State University Libraries ▾
Sep 9, 2013 - Main Library. (517) 432-6123. The Browsing Collection, Cesar Chavez Collection, Faculty Book Collection, and Oversize Collection are some of the collections ...

Services | Copy Center | MSU Libraries
https://www.lib.msu.edu/.../copycente... ▾ Michigan State University Libraries ▾

Interoperability.

Discoverability.

Visibility.

Authority.

More is better. Always.

Right?

The Reflective Moment

Sociotechnical Considerations

Document 1 (left)

Cloud Publications

Research Article   **Open Access**

Professional and Ethical Issues in Semantic Web

Pankaj Kumar [1] and Dr. A. K. Singh [2]

[1,2] University Department of Mathematics, B. R. Ambedkar Bihar University, Muzaffarpur, Bihar, India

Correspondence should be addressed to Pankaj Kumar, pankaj@glug4muz.org and Dr. A. K. Singh, aksingh@brabu.net

Abstract This paper highlights professional and ethical issues in semantic web environment. As in case of semantic web environment information is more visible to search engines and other websites and so it raises significant privacy issues for users providing personal data on a semantic website. This paper also highlights the importance of informed consent in a semantic web environment. Finally it explores the challenge for developers and designers of semantic web environment and potential approach to meeting these issues in an efficient way.

Keywords *Semantic Web, RDF, OWL, Ethics, Privacy, Informed Consent*

1. Introduction

The main aim of semantic web (1) is to develop a website in which data available on it becomes as much visible to other websites as it is visible to its users. The information available on such websites becomes much more sharable as compared to information available on traditional websites. So the privacy issue for users signing up on these websites becomes more critical. The information provided by users becomes available in the source code of the web pages in very systematic way so that they can be easily scrapped by crawlers and other websites. The information provided by users may even become available in the URL of the web pages of that website. So such situation may become problematic for users. Some of the sensitive information given by the users can also become accessible on other websites. So it may violet the consent given by the user at the time of creating account on that website. Therefore the ethical responsibility of designer and developer of such website become much more than as compared to traditional websites.

2. Introduction to Semantic Web

Traditionally Internet was developed as collection of static web pages. It was simply collection of documents. But now the Internet has become collection of data. The documents available on Internet are collection of data provided by users. Traditionally it was one way communication between the website owner and visitor. But now days, the information available on web pages mostly consists of information provided by visitors itself. So one visitor is interested in searching for information provided by other visitor. The main aim of semantic web is to develop this searching process easy and accurate. For this purpose, the methods of displaying information on web pages are being redefined.

Document 2 (center)

ORIGINAL PAPER

Semantic Web Regulatory Models: Why Ethics Matter

Pompeu Casanovas

Abstract The notion of *validity* fulfils a crucial role in legal theory. In the emerging Web 3.0, Semantic Web languages, legal ontologies, and normative multi-agent systems (nMAS) are designed to cover new regulatory needs. Conceptual models for complex regulatory systems shape the characteristic features of rules, norms, and principles in different ways. This article outlines one of such multilayered governance models, designed for the CAPER platform, and offers a definition of Semantic Web Regulatory Models (SWRM). It distinguishes between *normative*-SWRM and *institutional*-SWRM. It also compares existing principles in privacy by design, linked open data (LOD), legal information institutes (LII), and online dispute resolution (ODR). The article concludes by proposing the notion of Relational Law to summarize the ethical dimension of SWRM. Ethics are the only regulatory way to constitute a *global space*, out of the jurisdictional *public domain* set by national, international, or transnational law, and opposed to the *private* one.

Keywords Validity · Legal theory · Semantic web regulatory models (SWRM) · Legal ontologies · Normative multi-agent systems (nMAS) · Privacy by design (PbD) · Linked open data (LOD) · Online dispute resolution (ODR)

1 Introduction

Collaboration between humans and computers is the key in addressing societal challenges. Democracy and the rule of law can and should benefit from this synergy. The

P. Casanovas (✉)
Institute of Law and Technology (IDT), Department of Political Science and Public Law,
Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona 08193, Spain
e-mail: pompeu.casanovas@uab.cat

P. Casanovas
e-mail: pompeu.casanovas@rmit.edu.au

P. Casanovas
Centre for Applied Social Research, School of Global, Urban and Social Studies, RMIT University,
City Campus, 124 LaTrobe Street, Melbourne, Vic 3000, Australia

🖄 Springer

Document 3 (right)

Computer Ethics in the Semantic Web Age

Aziz Alotaibi
Department of Computer Science
221 University Ave, University of Bridgeport
Bridgeport, CT, USA

Computer ethics can be defined as a set of moral that monitor the use of computers. Similar rules were made for both programmers and users. Issues that were raised in the past have arisen due to the introduction of systems such as Semantic Web. Both programmers and users are obliged to consider phenomenon such as informed consent. In this paper, I will explore the ethical problems that professionals and users with the advent of new technologies especially with privacy concerns and global

Computer ethics; semantic web; privacy concerns; information

I. INTRODUCTION

In the 1940s, the computer ethics field was founded by author Norbert Wiener as academic filed[1]. With advances made in computer science, web technologies provided us with the opportunity to explore a whole new world we were previously unaware of due to our limited resources. However, this knowledge came at a price. Individuals can no longer be assured that their privacy is intact on the web. First, people's privacy is being violated by hackers and computer crime since their privacy is being controlled by other entities, such as agencies or other private enterprises.

Thus, people have share personal information at a larger scale than in the past; for instance, their lifestyle, views about their health, their financial status, their views, their religious affiliation, and gender. Thus, greater paranoia in their minds that their information accessed by computer hackers. From the hacker's view, there should be no restriction to accessing information. However, that is not the case and privacy being strictly controlled through various means apart from being easily shared. Yet, in the new age where information our desire to gather information has been all-time high, we pay for our quicker, faster access by loosening our grip on our privacy and security

II. SEMANTIC WEB

Tim Berners-Lee, at CERN (European Center for Nuclear Research), made a proposal for a system that would enable people to share the information over hypertext system. And in 1990, he wrote the first building HTML[3]. As we know WWW is an acronym for World Wide Web which known as (web). He invented the HTML and HTTP protocol. HTML, short for hypertext markup language, is used to create and

structure documents that can be shown as web pages. HTTP, short for the hypertext transfer protocol, is an application protocol that use to request and transmit files over the internet. And he invented the HTML and HTTP protocol. HTML, stands for hypertext markup language, is used to create and structure documents that can be shown as web pages[4]. HTTP, short for the hypertext transfer protocol, is an application protocol that use to request and transmit files over the internet. Semantic web is known as web 3.0 and an extension to the World Wide Web that enables to share the data.

As Tim Berners-Lee states semantic web is the movement from the web of documents to the web of data [5]and from the human centric to the machine-processable.

TABLE 1.   WEB GENERATION

Document web 1.0	Social web 2.0	Semantic web 3.0	Ubiquitous web 4.0
Document – centric HTML, CSS, Javascript Search engines URL	Data-centric API, web services, AJAX Social network URI	Data-centric RDF,RDFs, OWL, SPARQL Structure data Linked data	User centric Invisible web Smart market Semantic mobile services
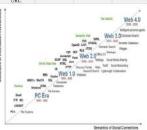
Fig. 1. Source: Radar Networks & Spivack,2007 www.radarnetworks.com

Since these privacy issues already existed prior to the development of newer technologies such as semantic web, its creators and users may question how the introduction of this new platform will affect an already existing problem, i.e., how

Informed Consent**.**

Data Persistence**.**

Privacy**.**

Programmers *¿and?* *¿or?* *¿versus?* Users**.**

ORCID

Connecting Research
and Researchers

FOR RESEARCHERS   **FOR ORGANIZATIONS**   **ABOUT**   **HELP**   **SIGN IN**

DISTINGUISH YOURSELF IN
THREE EASY STEPS

ORCID provides a persistent digital identifier that distinguishes you from every other
researcher and, through integration in key research workflows such as manuscript and grant
submission, supports automated linkages between you and your professional activities ensuring
that your work is recognized. Find out more.

1 **REGISTER** Get your unique ORCID identifier Register now!
Registration takes 30 seconds.

2 **ADD YOUR INFO** Enhance your ORCID record with your
professional information and link to your other
identifiers (such as Scopus or ResearcherID or
LinkedIn).

3 **USE YOUR ORCID ID** Include your ORCID identifier on your Webpage,
when you submit publications, apply for grants, and
in any research workflow to ensure you get credit
for your work.

LATEST NEWS

**Fri, 2016-02-19
ORCID Security
Incident**

**Tue, 2016-02-16
Australian ORCID
consortium officially
launched!**

**Thu, 2016-02-11
Updates to ORCID
Governance:
Member Elections**

ACADEMIA

Join 32,668,121 Academics

Academia is the easiest way to share papers with millions of people across the world for free. A recent study found that papers uploaded to Academia receive a 69% boost in citations over 5 years.

G+ GOOGLE SIGN UP

f FACEBOOK SIGN UP

Don't have Google or Facebook? Sign Up with Email

Job Board   About   Press   Blog   People   Papers   Terms   Privacy   Copyright   💼 We're Hiring!   ❓ Help Center

Find new research papers in: Physics   Chemistry   Biology   Health Sciences   Ecology   Earth Sciences   Cognitive Science   Mathematics   Computer Science   Engineering

#deleteacademiaedu

Column 1:

In reply to Laura Pasquini
Jamison Miller @MillerJamison · 24h
@laurapasquini @paoloman @delaubrarian Would love to #DeleteAcademiaEdu + use an #openaccess inst repository, but are grad stdnts eligible?

Laura Pasquini @laurapasquini · Feb 21
Should you #DeleteAcademiaEdu? Role of commercial services in scholarly communication bit.ly/1mR7pRj by @paoloman #ScholarSunday

Should you #DeleteAcademiaEdu? On the role of commercial service...
Reflecting on the recent surge of criticism about the commercial motives of scholarly social media platform Academia.edu, Paolo Mangiafico argues...
blogs.lse.ac.uk

OA Tracking Project @oatp · Feb 18
Algorithmic accountability in scholarship: what we can learn from #DeleteAcademiaEdu | LSE Media Policy Pr... bit.ly/1SBDRoX #OA

In reply to Ethan Gruber
Stefano Costa @stekosteko · Feb 16
@ewg118 now I manually added abstracts for all papers, do I need to repe... procedure? I thought #deleteacademiaedu was for the lazy 😄😄

Ethan Gruber @ewg118 · Feb 16
Re: #deleteacademiaedu - Academia->@ZENODO_ORG tool is ready for testing! admin.numismatics.org:8080/orbeon/academi... create account on sandbox.zenodo.org

KC @multihyphenate · Feb 12
Startups Sometimes Pretend That They Don't Want to Make Money - Thou... on #deleteacademiaedu digital-science.com/blog/perspecti...

Column 2:

Sierra Williams @sn_will · Feb 12
Good piece to read with #DeleteAcademiaEdu and scholarly infrastructure debates in mind

Bonnie Stewart @bonstewart
Friday reading on human agency in tech design...& choices that build community @EducationPEI #LPAC twitter.com/girlziplooked/...

Stewart Varner @StewartVarner · Feb 11
In the wake of #DeleteAcademiaEdu, this overview of SHARE gives us a few things to think about: acrl.ala.org/techconnect/po...

Ethan Gruber @ewg118 · Feb 11
Now able to get and post JSON in @orbeon #xforms engine! look for #academiaedu -> @ZENODO_ORG migration soon! re: #deleteacademiaed

SPARC Europe @SPARC_EU · Feb 11
#DeleteAcademiaEdu controversy debates role of money in scholarly publishing ihenow.com/1POVzD0

Gaëlle Lortal @Gaelle_Lortal · Feb 10
To #DeleteAcademiaEdu or Not To #DeleteAcademiaEdu? | L'histoire contemporaine à l'ère numérique histnum.hypotheses.org/2589

Chris Millson @asameshimasa · Feb 9
The start of the #DeleteAcademiaEdu debate... What are your thoughts? chronicle.com/article/Schola... #okhe #oa

THE CHRONICLE OF HIGHER EDUCATION

Scholars Criticize Academia.edu Proposal to Charge Authors for Rec...
The social network for sharing academic papers says the idea is just under consideration, not a done deal, but the critics have responded with outr...
chronicle.com

ISTifsttar @ISTifsttar · Feb 9
To #DeleteAcademiaEdu or Not To #DeleteAcademiaEdu? utilité ou non d #réseauxsociaux pour chercheurs histnum.hypotheses.org/2589

MHH @MHoussayH · Feb 8
Ben voilà. Limpide, clair, argumenté. Complètement d'accord #OpenAccess #DeleteAcademiaEdu #OrNot

Column 3:

eLearn Australia @elearnaustralia · Feb 8
Good summary of Academia.edu proposal to charge authors buff.ly/2OC @chronicle #DeleteAcademiaEdu #highered #aussieed

THE CHRONICLE OF HIGHER EDUCATION

Scholars Criticize Academia.edu Proposal to Charge Authors for P...
The social network for sharing academic papers says the idea is just u consideration, not a done deal, but the critics have responded with out
chronicle.com

Ethan Gruber @ewg118 · Feb 8
please help test. i am not sure how many publications academia includes HTML response (at least a few dozen). #DeleteAcademiaEdu

Ethan Gruber @ewg118 · Feb 8
Re: #DeleteAcademiaEdu, have the scraping of academia pub metadata working: admin.numismatics.org/cgi-bin/scrape... sub your profile URI to 1/2

William Morgan, Ph.D @willmorgan66 · Feb 8
Should you #DeleteAcademiaEdu? - Scholarly Communications @ Duke klou.tt/1aastghtckb9s

GeraldineCS @GeraldineCS · Feb 7
Impact of Social Sciences – Should you #DeleteAcademiaEdu? On the commercial services in scholarly Comms blogs.lse.ac.uk/impactofsocial...

Should you #DeleteAcademiaEdu? On the role of commercial serv...
Reflecting on the recent surge of criticism about the commercial motiv scholarly social media platform Academia.edu, Paolo Mangiafico argue
blogs.lse.ac.uk

Column 4:

Liz Covart @lizcovart · Feb 6
#DeleteAcademiaEdu Controversy Debates Role of Money in Scholarly Publishing @insidehighered #Twitterstorians histry.us/1nHZXJE

Voice of Researchers @Research_Voice · Feb 6
.@phillbjones @digitalsci on the recent #deleteacademiaedu row. What do you think? digital-science.com/blog/perspecti... fo.me/5Rine5ydb

CS unibz @cs_unibz · Feb 6
Scholars criticise Academia.edu proposal to charge authors for recommendations. Thoughts? #DeleteAcademiaEdu buff.ly/1Qz367F

Top Left Brick @TopLeftBrick · Feb 5
Go team! Your #DeleteAcademiaEdu means you deleted the one freely-available(w/ login) version of your paper and left the one that costs $35.

Top Left Brick @TopLeftBrick · Feb 5
Okay, know what? #DeleteAcademiaEdu if you feel like it, but move your stuff wherever else it's going to live FIRST.

People are using it.

Digital Science @digitalsci · Feb 5
"Startups sometimes pretend they don't want to make money" ow.ly/XTOx2 #deleteacademiaedu

OA Tracking Project @oatp · Feb 5
Startups Sometimes Pretend That They Don't Want to Make Money - Thoughts on #deleteacademiaedu - Digital S... bit.ly/1L3tPF9 #OA

Katherine McDonald @Katherine_McDon · Feb 5
Part of an effort to move my stuff onto my website, and maybe #DeleteAcademiaEdu (@DrMagennis)

Marie-Anne Paveau @mapav8 · Feb 5
mustread !! // To #DeleteAcademiaEdu or Not To #DeleteAcademiaEdu? | L'histoire contemporaine à l'ère numérique bit.ly/1nRGR3Q

Tom Cutterham @tomcutterham · Feb 5
"Bringing the Economy Back In" profile.nus.edu.sg/fass/poliou/in... #DeleteAcademiaEdu

Peter H. Bent @phbent
The question of how to construct modern economic relations as an object of political theorizing academia.edu/8734097/Bring...

LSE THE LONDON SCHOOL
of ECONOMICS AND
POLITICAL SCIENCE ▪

THE
IMPACT
BLOG

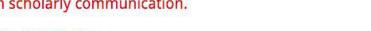Home   Latest   About   Research Book   #AcWri2016   Resources   LSE Comment   Popular   🔍

Should you #DeleteAcademiaEdu? On the role of commercial services in scholarly communication.

Reflecting on the recent surge of criticism about the commercial motives of scholarly social media platform Academia.edu, **Paolo Mangiafico** *argues this is now an ideal opportunity for scholars to make informed choices about their work. If you are comfortable with the trade-offs and risks, and willing to exchange those for the service provided, then don't #DeleteAcademiaEdu. But consider whether alternatives exist that will meet your needs while also preserving your ability to keep control of your work.*

Last week a kerfuffle arose on Twitter about Academia.edu, a social networking site for academics, where many academic authors have profiles, share their publications, and connect with other scholars. You can read about the beginning of the controversy in this article the Chronicle of Higher Education posted on Friday.

The ensuing tweetstorm followed a fairly typical trajectory – moral outrage, call to action, a hashtag, and then of course the inevitable backlash, with each side calling into question the integrity of each

LSE Impact Blog
Community
Survey
Have your say!

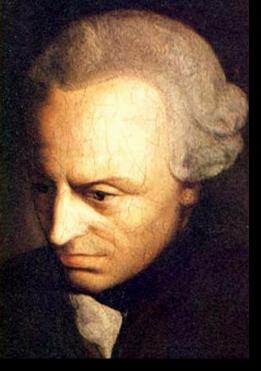**Were Adnan's intentions ethical?**

Is academia.edu an ethical company?

How can we know?

Categorical Imperative
immoral to use another person merely
as a means to an end

a posteriori
justification relies upon experience;
thus, such propositions are logically contingent

Immanuel Kant
1724 - 1804

Remember...

Not new questions.

Answering is not the end goal.

Sit with questions.

Consider implications in digital endeavors.

Framework for Ethical Inquiry

Context

Expectation

Value Analysis

Mannheimer S, Young SWH, Rossmann D (2016) On the Ethics of Social Network Research in Libraries. Journal of Information, Communication, and Ethics in Society 14(2) (forthcoming). http://scholarworks.montana.edu/xmlui/handle/1/9470

Context

❏ What are the data privacy norms and policies in your culture or context?

Expectation

❏   What are your expectations as a researcher conducting your research?

❏   What are users' expectations of data reuse?

Value Analysis

❏ Consider the balance between privacy risks/potential ethical breaches and potential research gains.

❏ Consult with an ethicist, when appropriate and feasible.

How can ethical inquiry impact design and development?

Privacy statements

User testing to confirm expectations

Transparency about how algorithms work

Development that ends if privacy is threatened

Analytical work that ends if exploitation becomes apparent