Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier

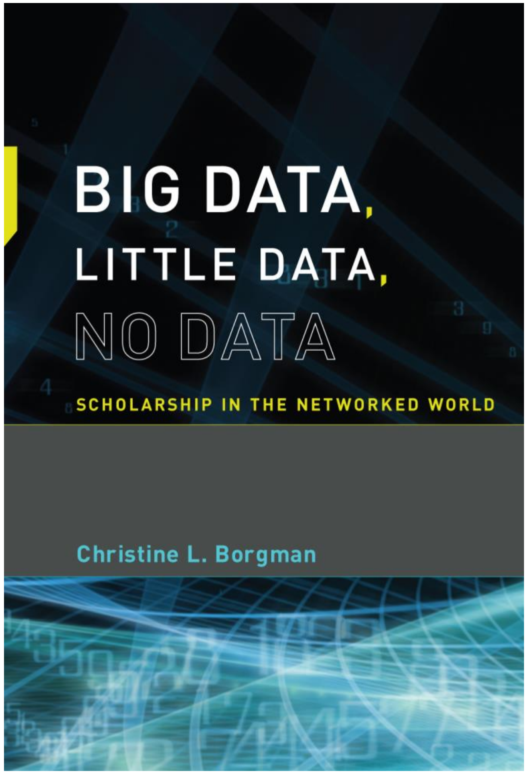
BTLJ
Berkeley Technology Law Journal

Christine L. Borgman

Distinguished Research Professor
Director, Center for Knowledge Infrastructures, UCLA
http://christineborgman.info
@scitechprof

Visiting Scholar, Harvard, October 2018
- Center for Astrophysics
- Data Science Initiative
- Berkman Klein Center for Internet & Society

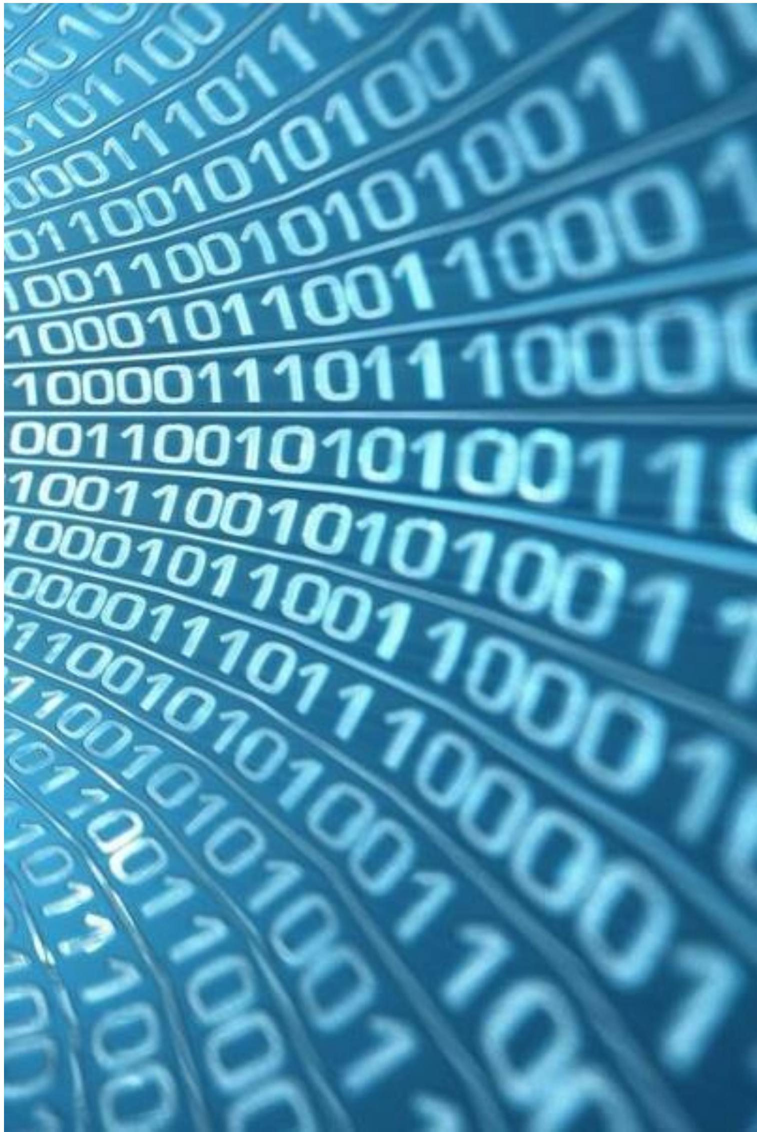
Berkman Klein Center for Internet & Society
Data Science Initiative
https://cyber.harvard.edu/events/2018-10-09/open-data-grey-data-and-stewardship
October 9, 2018



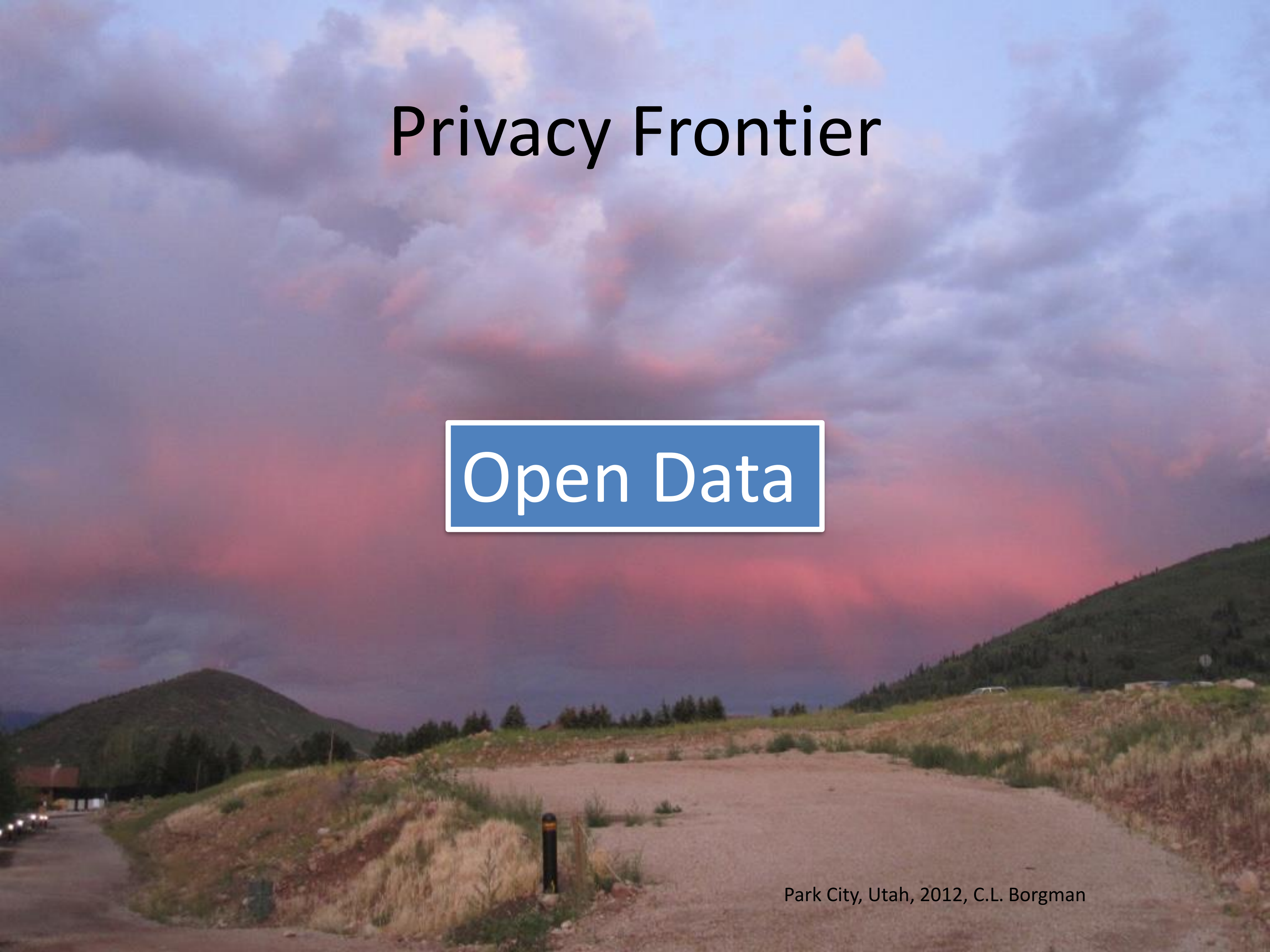
MIT Press, 2015

Universities in a Data-Rich World

- Exploit data for missions
  - Research
  - Teaching
  - Services
- Sustain trust of community
  - Privacy
  - Academic freedom
  - Stewardship and governance



Privacy Frontier



Open Data

Park City, Utah, 2012, C.L. Borgman

THE WHITE HOUSE
WASHINGTON

Open access to data

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

中央研究院
ACADEMIA SINICA

- Research Councils of the UK
- European Union
- Australian Research Council
- U.S. Federal research policy
- Taiwan, China, India…
- Individual countries, funding agencies

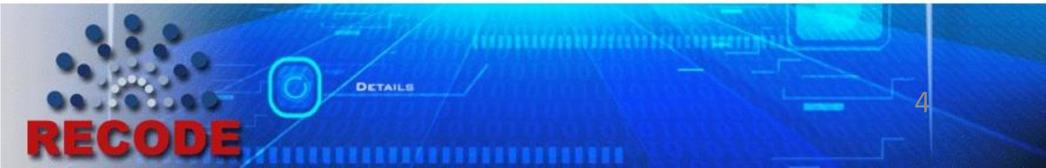
CHINESE ACADEMY OF SCIENCES

Supported by wellcometrust

Australian Government
National Health and Medical Research Council

National Science Foundation
WHERE DISCOVERIES BEGIN

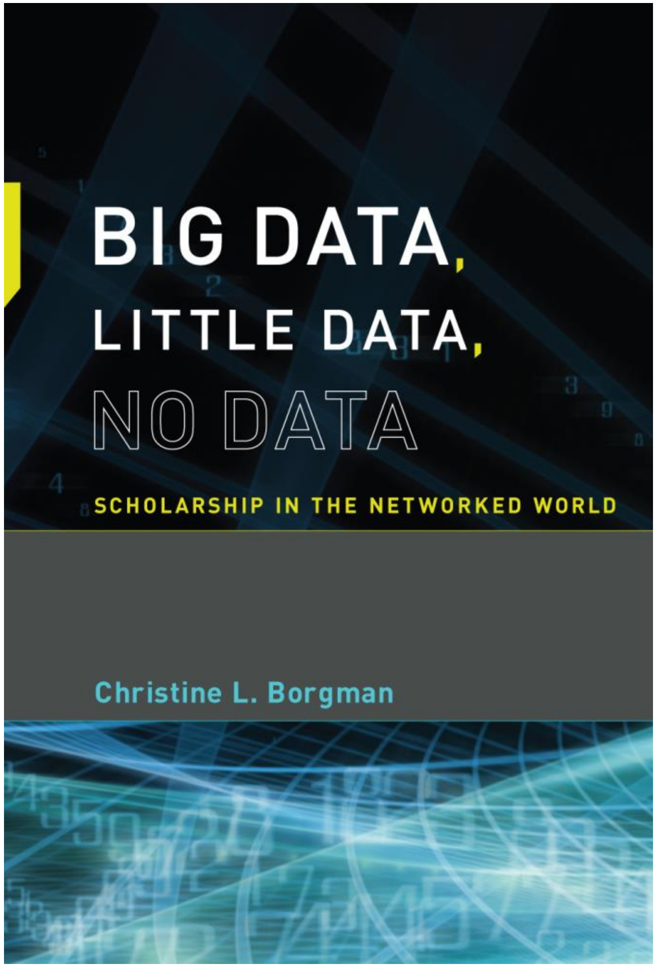
Policy RECommendations for Open Access to Research Data in Europe



NIH National Institutes of Health
Turning Discovery Into Health

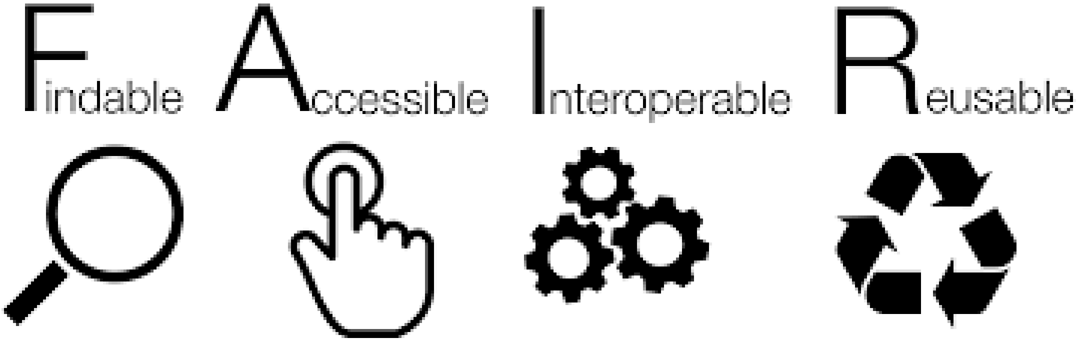
Why Open Access to Research Data?

- To reproduce research
- To make public assets available to the public
- To leverage investments in research
- To advance research and innovation



Open Data Practices

- Deposit datasets in a data archive
- Link dataset to journal article or publication
- Publish data documentation
  - – Research protocols
  - – Codebooks
  - – Software
  - – Algorithms



Wilkinson, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, http://dx.doi.org/10.1038/sdata.2016.18

https://upload.wikimedia.org/wikipedia/commons/a/aa/FAIR_data_principles.jpg

http://www.genome.gov/dmd/img.cfm?node=Photos/Graphics&id=85327

Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

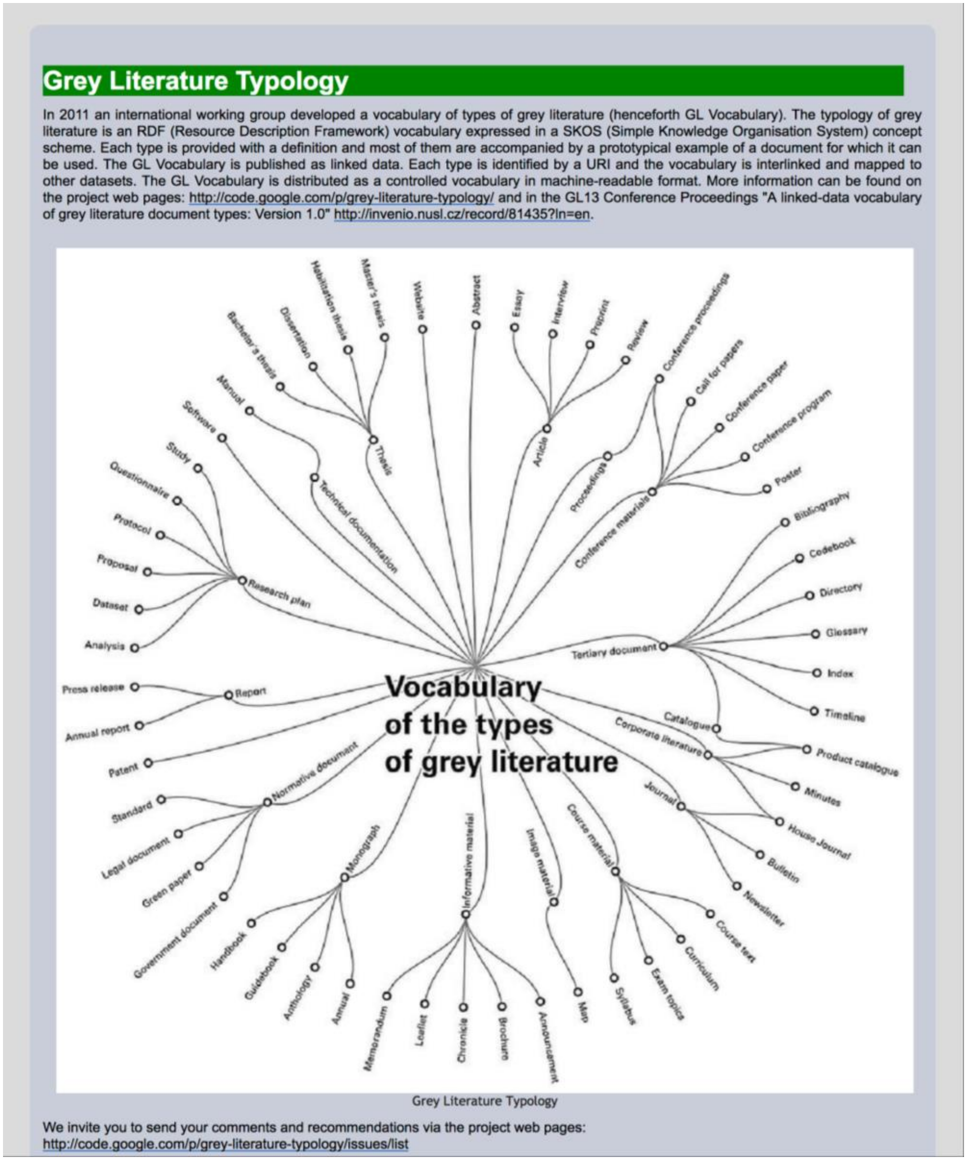
C.L. Borgman (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press

Publications

Dégradation ER mRNA...
Estrogen Receptors in Human Nontarget Tissues: Biological and Clinical Implications*
DANIEL R. CIOCCA AND LAURA M. VARGAS ROIG
1.3.2.1
Effects of Low Dose Oral Contraceptives on Ver
and Low Density Lipoprotein Metabolism
Abstract
1225
The Estradiol-stimulated Lipoprotein Receptor of Rat Liver

http://www.cse.psu.edu/hpcl/images/publications.jpg

Grey Literature



Grey Literature Typology

In 2011 an international working group developed a vocabulary of types of grey literature (henceforth GL Vocabulary). The typology of grey literature is an RDF (Resource Description Framework) vocabulary expressed in a SKOS (Simple Knowledge Organisation System) concept scheme. Each type is provided with a definition and most of them are accompanied by a prototypical example of a document for which it can be used. The GL Vocabulary is published as linked data. Each type is identified by a URI and the vocabulary is interlinked and mapped to other datasets. The GL Vocabulary is distributed as a controlled vocabulary in machine-readable format. More information can be found on the project web pages: http://code.google.com/p/grey-literature-typology/ and in the GL13 Conference Proceedings "A linked-data vocabulary of grey literature document types: Version 1.0" http://invenio.nusl.cz/record/81435?ln=en.

Grey Literature Typology

We invite you to send your comments and recommendations via the project web pages: http://code.google.com/p/grey-literature-typology/issues/list

- Reports
- Working papers
- Conference papers
- Preprints
- Patents
- Datasets
- Audio
- Video
- Slides
- Posters
- Codebooks
- Course syllabi
- Proposals
- Memos

http://www.greynet.org/

Stuart Miles: FreeDigitalPhotos.net

Grey Data

HIPAA

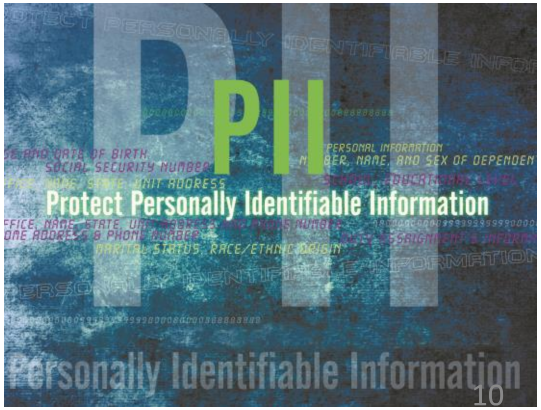
Health Insurance Portability and Accountability Act

https://www.linkedin.com/pulse/hipaa-privacy-rule-compliance-understanding-new-rules-syed-najaf

- Student applications
- Registrar records
- Learning management systems
- University ID cards: library, health, recreation, dorms, food service, transportation…
- Academic personnel dossiers
- Regulation and compliance data
- Staff surveys
- Sensor networks
- Security cameras
- Network traffic
- Street traffic…

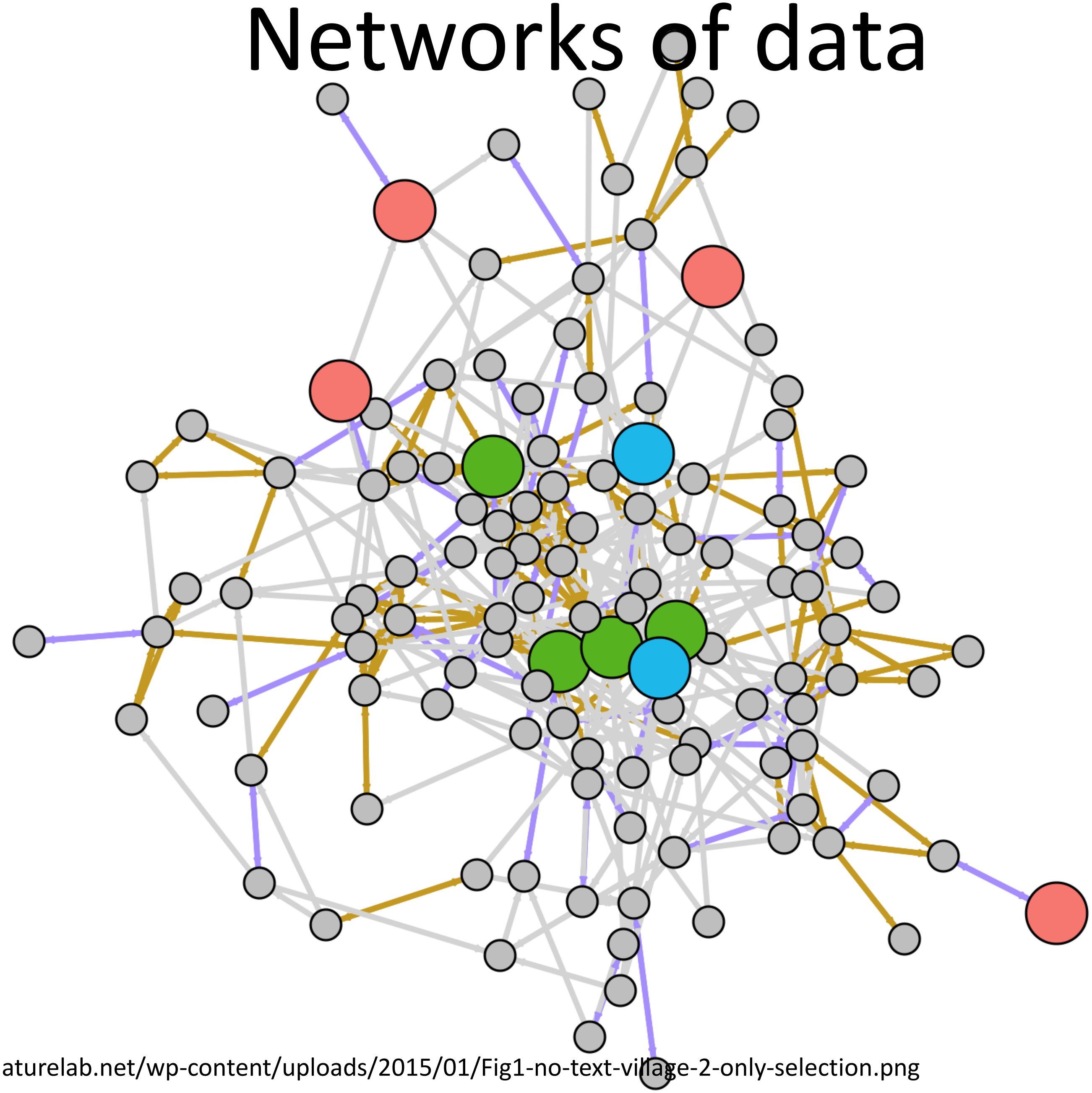
RESPECT FOR PERSONS • JUSTICE • BENEFICENCE •
UNIVERSITY OF WASHINGTON
IRB
HUMAN SUBJECTS DIVISION

DEPARTMENT OF EDUCATION • UNITED STATES OF AMERICA
FERPA
Family Educational Rights and Privacy Act


PII
Protect Personally Identifiable Information
Personally Identifiable Information
SOCIAL SECURITY NUMBER
PERSONAL INFORMATION
MARITAL STATUS, RACE/ETHNIC ORIGIN

http://www.aetc.af.mil/News/Article-Display/Article/559551/think-before-sending-protecting-pii/

Networks of data



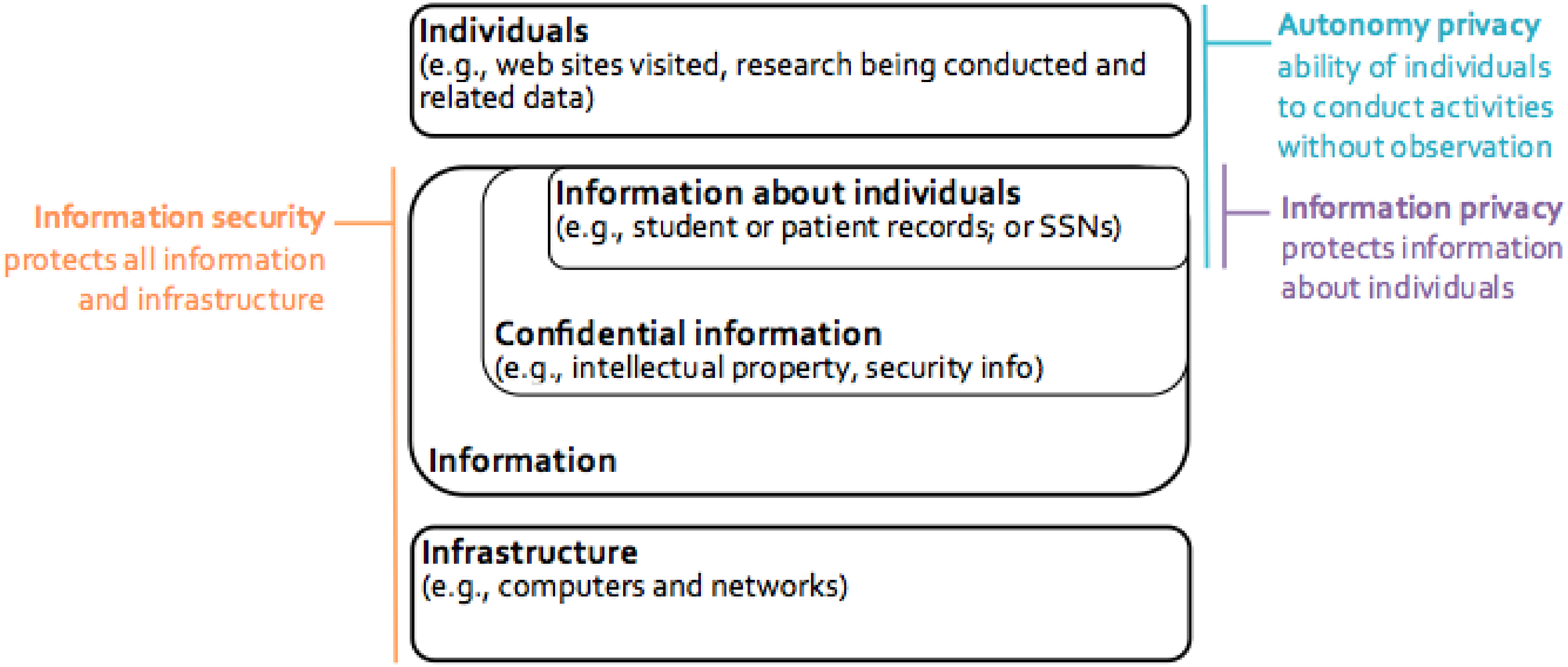
http://humannaturelab.net/wp-content/uploads/2015/01/Fig1-no-text-village-2-only-selection.png

University Responsibilities for Data

- Privacy
- Academic and intellectual freedom
- Stewardship and governance

Park City, Utah, 2012, C.L. Borgman

Information and Autonomy Privacy



Individuals
(e.g., web sites visited, research being conducted and related data)

Information about individuals
(e.g., student or patient records; or SSNs)

Confidential information
(e.g., intellectual property, security info)

Information

Infrastructure
(e.g., computers and networks)

Information security
protects all information and infrastructure

Autonomy privacy
ability of individuals to conduct activities without observation

Information privacy
protects information about individuals

UCOP Privacy and Information Security Initiative. (2013). http://ucop.edu/privacy-initiative/

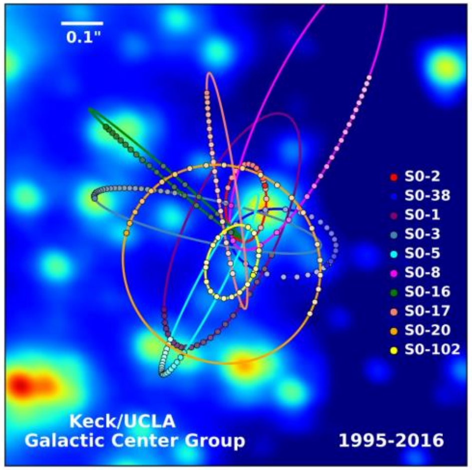
Information Privacy

https://aglearn.usda.gov/customcontent/OCIO/USDA-PII-Lite-Web/index.html

Autonomy Privacy

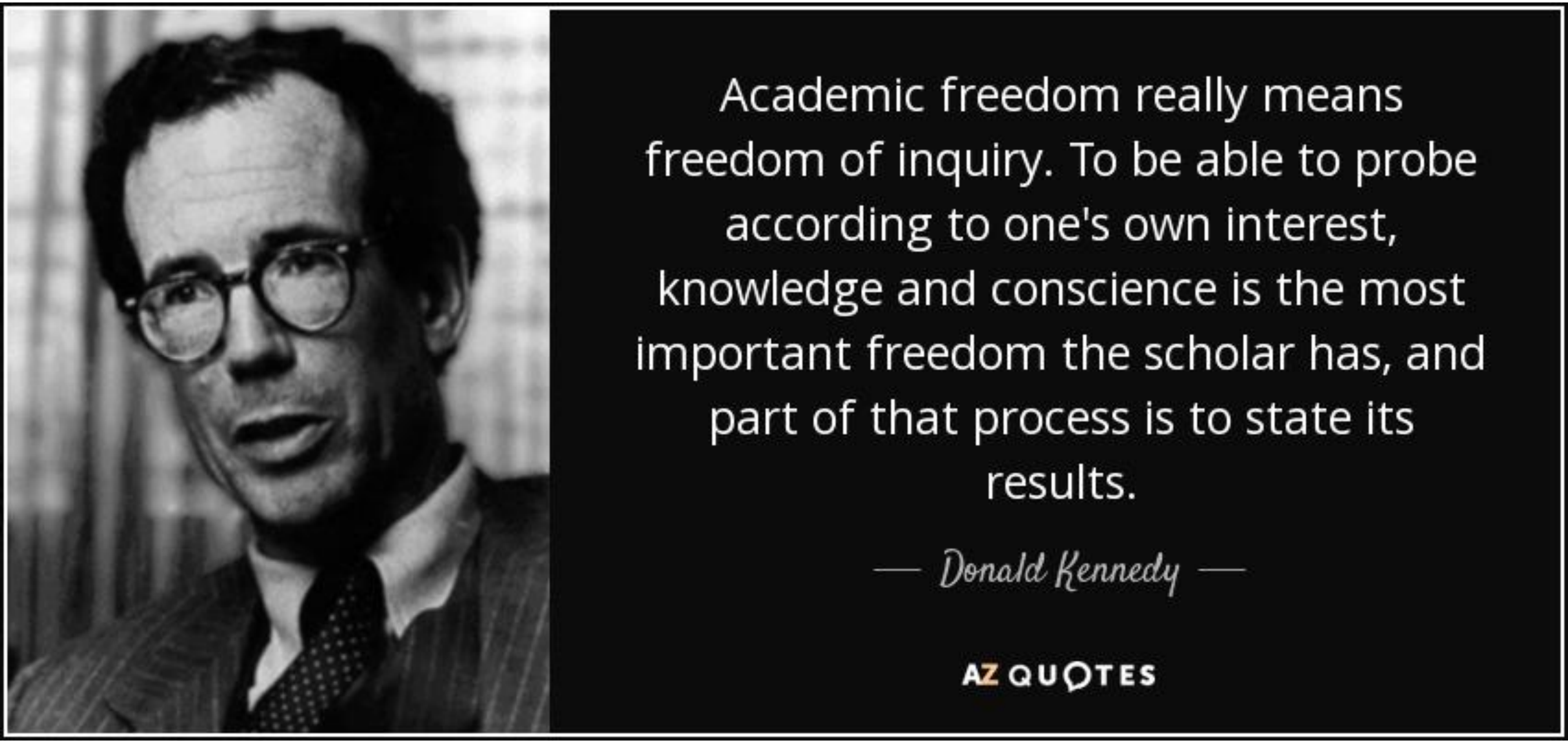
- Ability of individuals to conduct activities without surveillance
  - Intellectual inquiry
  - Conducting research
  - Classroom discussions
  - Searching for information
  - Email, web browsing
  - Reading ….

Title: Leader Mario Savio sounding off, Date: Nov. 9, 1964. Collection: San Francisco News-Call Bulletin Newspaper Photograph Archive (Free Speech Movement Selection), Owning Institution: UC Berkeley, Bancroft Library, Source: Calisphere. Date of access: November 10 2017 19:09, Permalink: https://calisphere.org/item/ark:/13030/ft400005ht/



http://www.galacticcenter.astro.ucla.edu/images_science.html

Academic Freedom



Academic freedom really means freedom of inquiry. To be able to probe according to one's own interest, knowledge and conscience is the most important freedom the scholar has, and part of that process is to state its results.

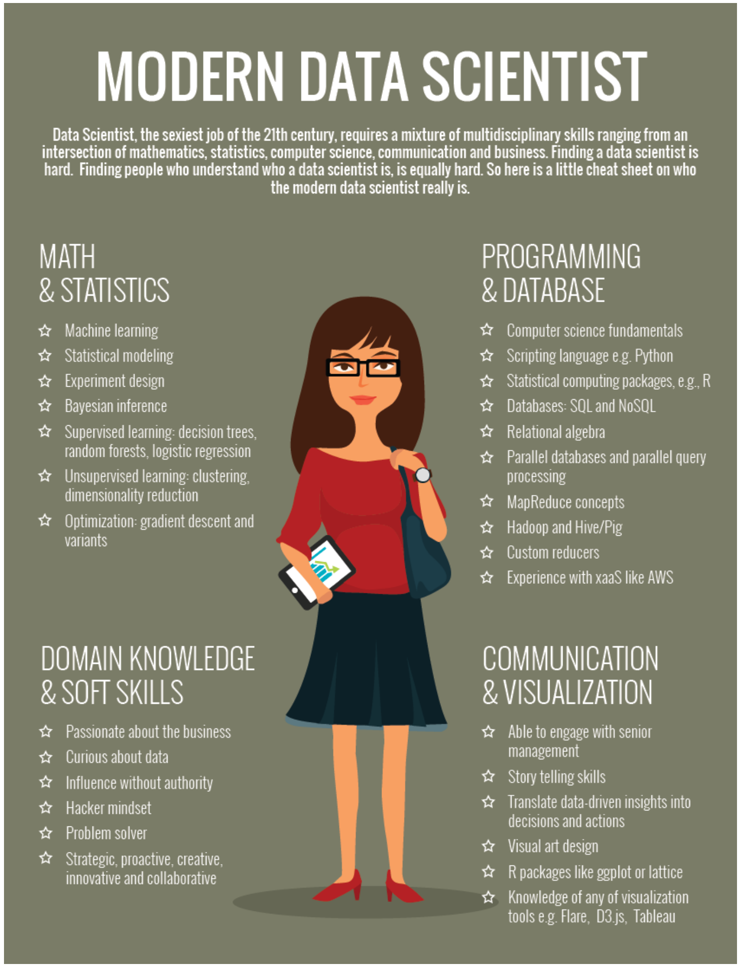
— Donald Kennedy —

AZ QUOTES

http://www.azquotes.com/quote/907028

Stewardship and Governance

- Protect
  - Privacy: information, autonomy
  - Academic freedom
- Secure infrastructure
- Data management
  - **F**indable
  - **A**ccessible
  - **I**nteroperable
  - **R**eusable
- Governance
  - Principles
  - Processes



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

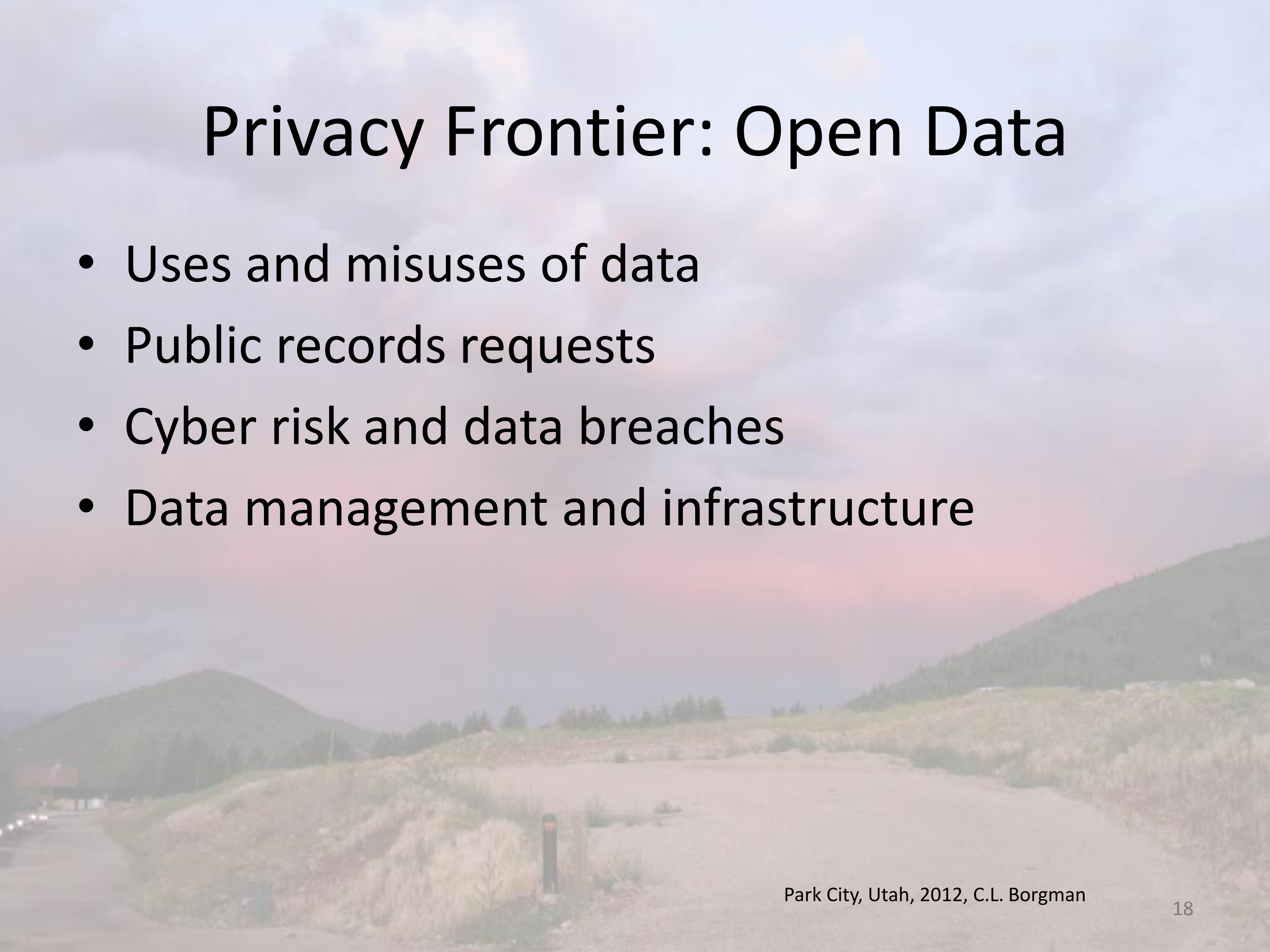
COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

https://github.com/okulbilisim/awesome-datascience

Privacy Frontier: Open Data

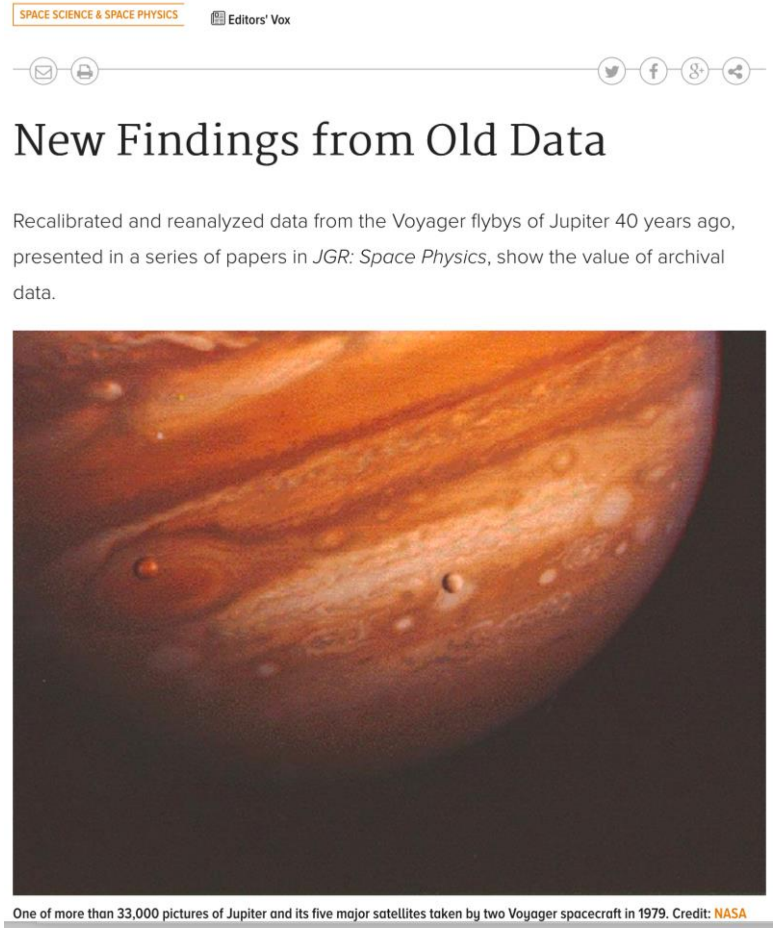
- Uses and misuses of data
- Public records requests
- Cyber risk and data breaches
- Data management and infrastructure



Park City, Utah, 2012, C.L. Borgman

Uses and Misuses of Data

Reuse



SPACE SCIENCE & SPACE PHYSICS

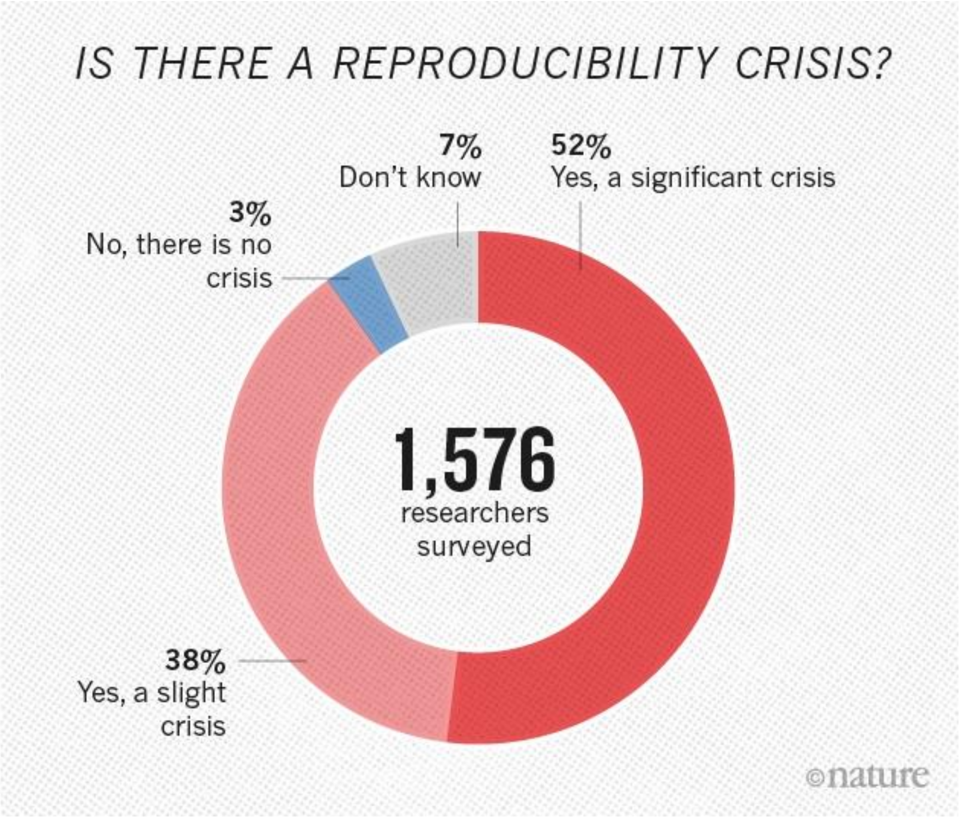
Editors' Vox

New Findings from Old Data

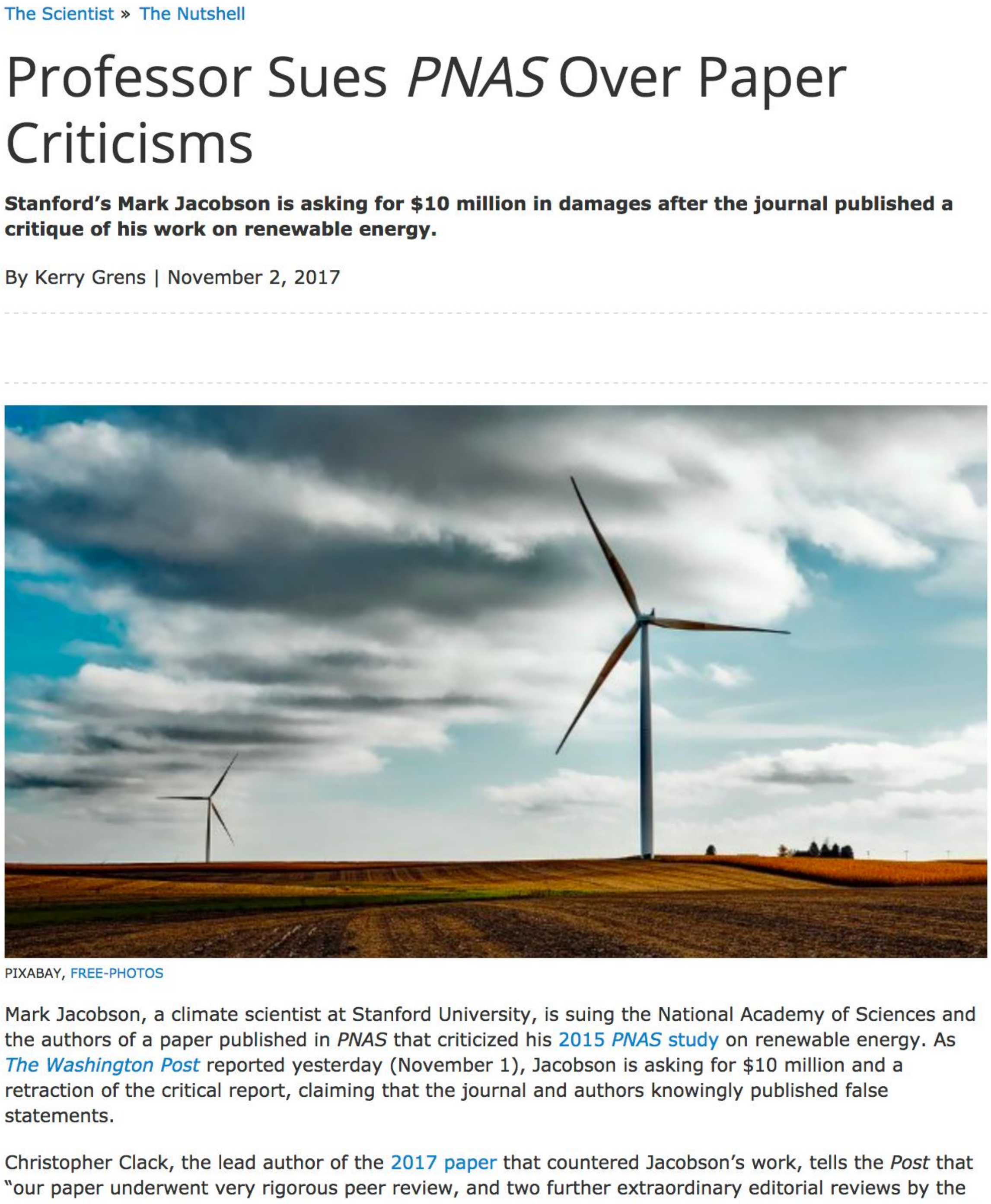
Recalibrated and reanalyzed data from the Voyager flybys of Jupiter 40 years ago, presented in a series of papers in *JGR: Space Physics*, show the value of archival data.

One of more than 33,000 pictures of Jupiter and its five major satellites taken by two Voyager spacecraft in 1979. Credit: NASA

Reproducibility



http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970



The Scientist » The Nutshell

Professor Sues *PNAS* Over Paper Criticisms

Stanford's Mark Jacobson is asking for $10 million in damages after the journal published a critique of his work on renewable energy.

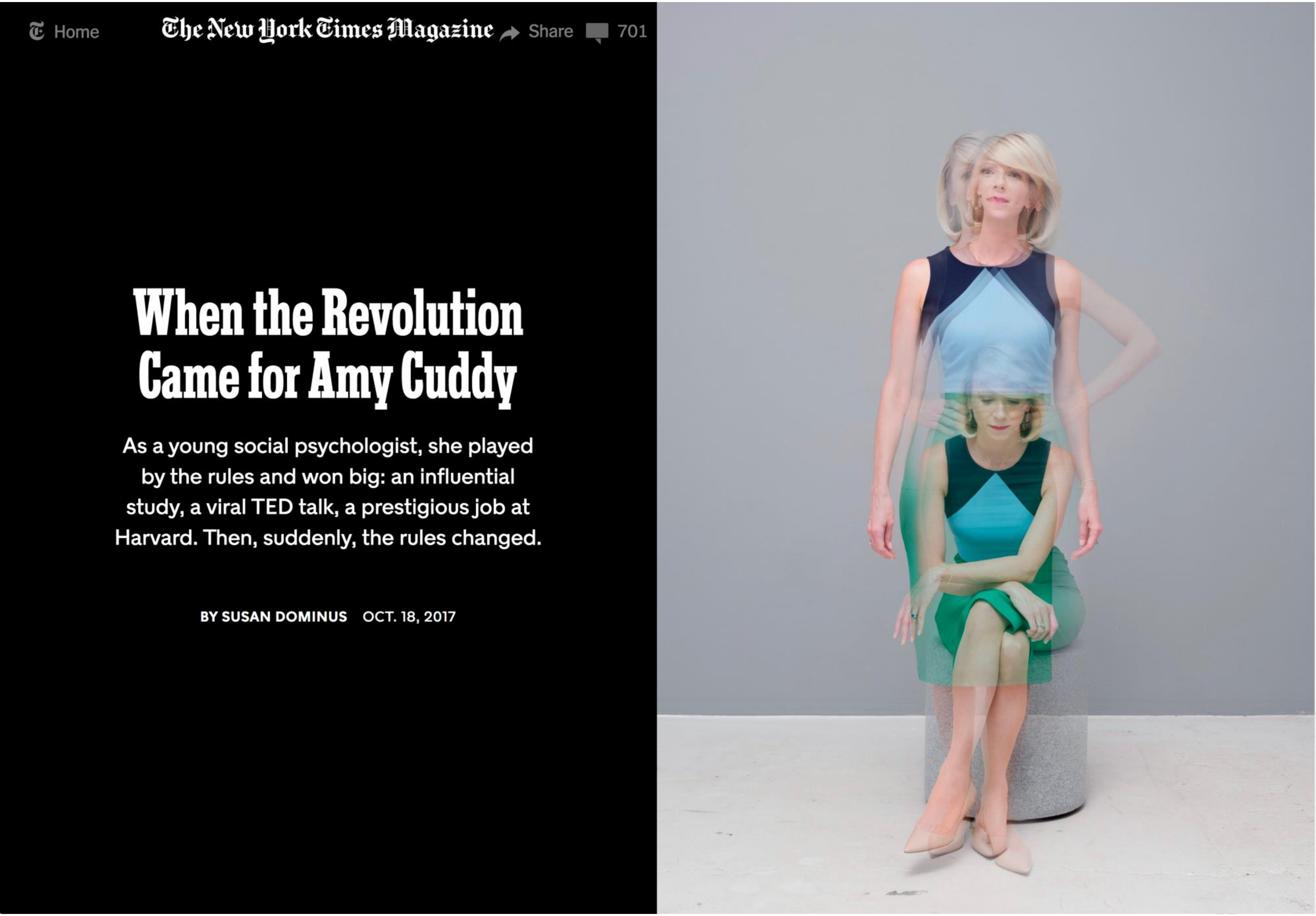
By Kerry Grens | November 2, 2017

PIXABAY, FREE-PHOTOS

Mark Jacobson, a climate scientist at Stanford University, is suing the National Academy of Sciences and the authors of a paper published in *PNAS* that criticized his 2015 *PNAS* study on renewable energy. As *The Washington Post* reported yesterday (November 1), Jacobson is asking for $10 million and a retraction of the critical report, claiming that the journal and authors knowingly published false statements.

Christopher Clack, the lead author of the 2017 paper that countered Jacobson's work, tells the *Post* that "our paper underwent very rigorous peer review, and two further extraordinary editorial reviews by the

Update (February 21, 2018): At a hearing yesterday in the District of Columbia Superior Court, a judge heard testimony from National Academy of Sciences lawyers, who were asking her to dismiss the defamation lawsuit. According to Retraction Watch, the attorneys argued that PNAS *is protected by a law designed to preserve speech that's in the public interest. Jacobson's lawyer disagreed, but the judge has yet to make her decision.*



Home

The New York Times Magazine

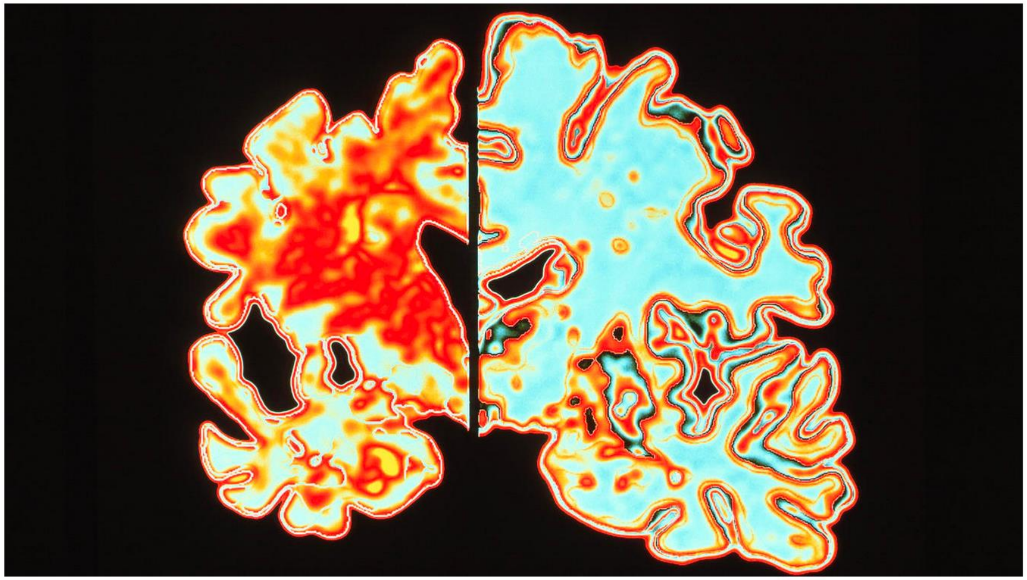
Share 701

When the Revolution Came for Amy Cuddy

As a young social psychologist, she played by the rules and won big: an influential study, a viral TED talk, a prestigious job at Harvard. Then, suddenly, the rules changed.

BY SUSAN DOMINUS OCT. 18, 2017

UC San Diego sues USC and scientist, alleging conspiracy to take funding, data

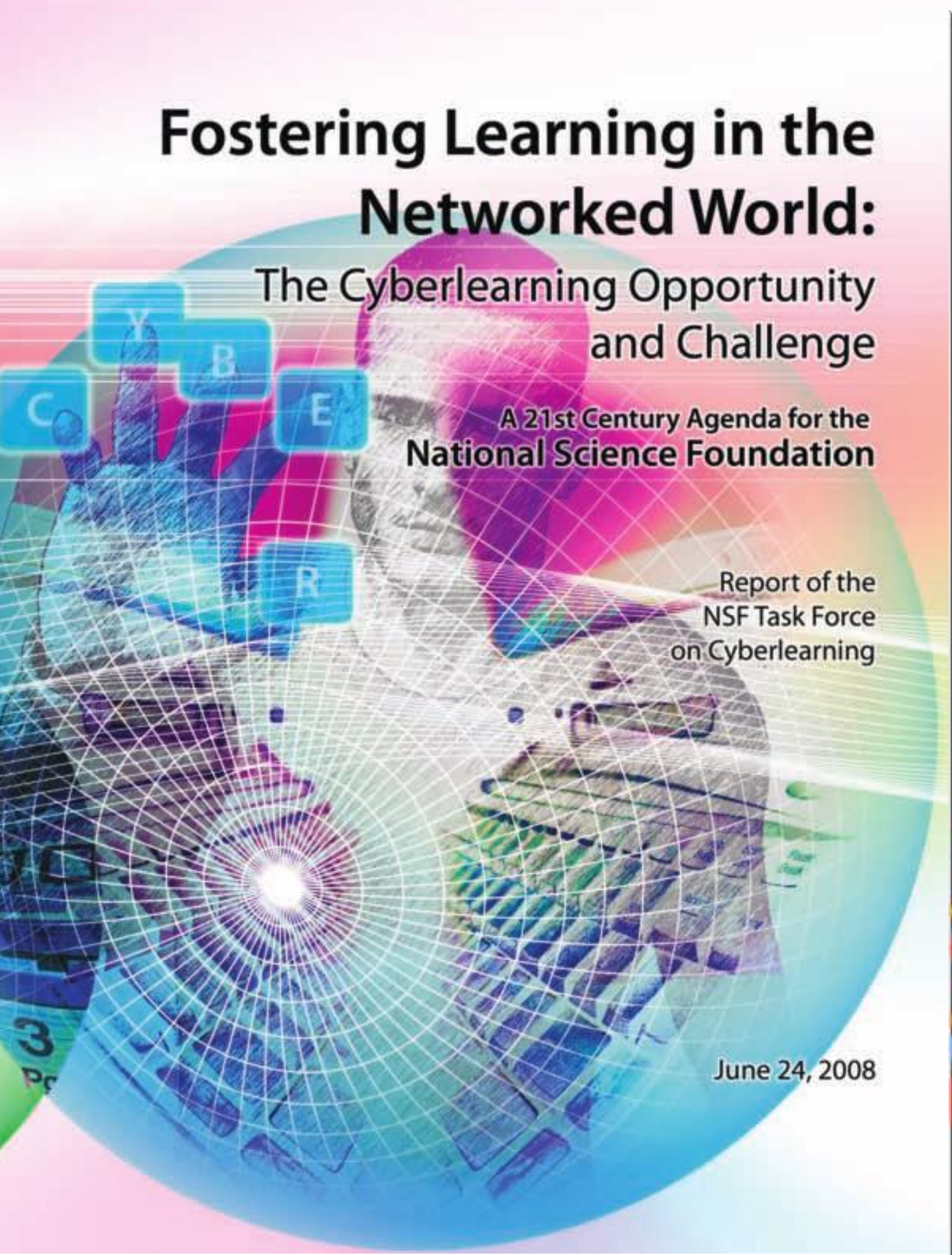


An ultrasound comparison of a brain of a patient with Alzheimer's disease, left, and a normal brain, right. (Pasieka / Getty Images)

By Bradley J. Fikes

JULY 5, 2015, 5:55 PM

UC San Diego has sued USC and a nationally recognized Alzheimer's disease researcher, alleging that they illegally



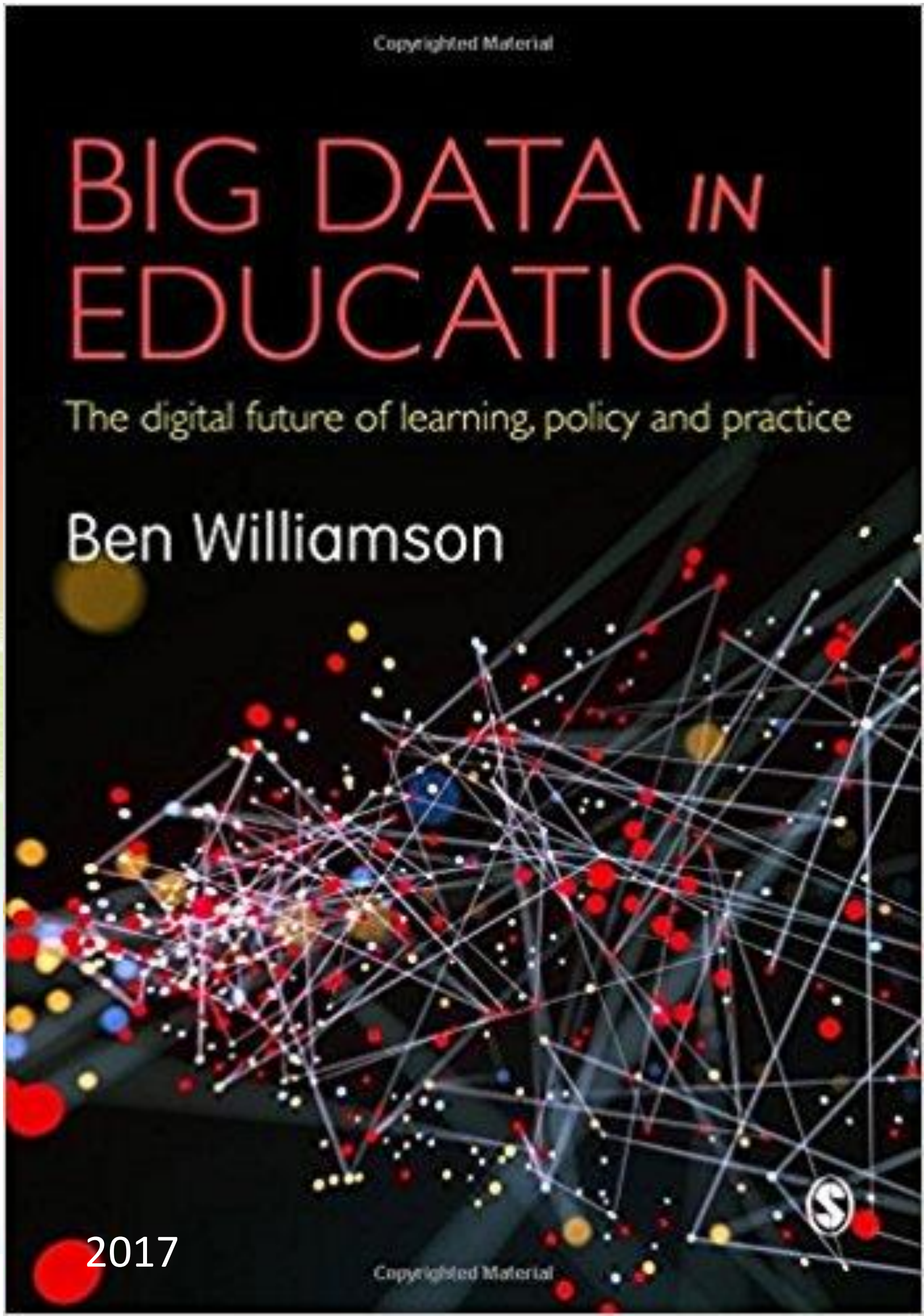
Fostering Learning in the Networked World:

The Cyberlearning Opportunity and Challenge

A 21st Century Agenda for the
National Science Foundation

Report of the
NSF Task Force
on Cyberlearning

June 24, 2008



Copyrighted Material

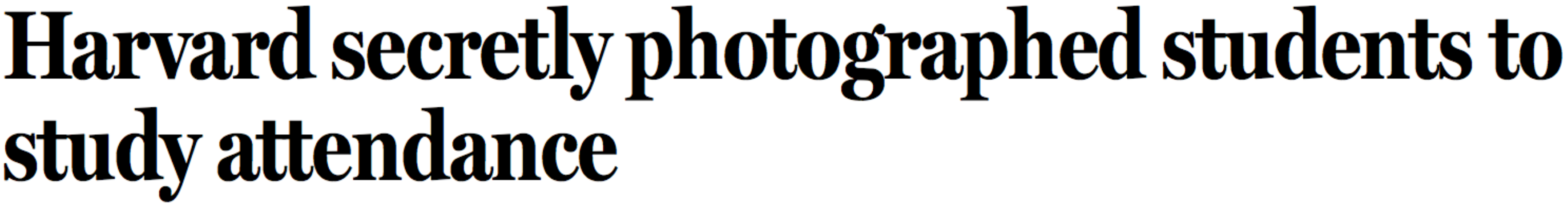
BIG DATA IN EDUCATION

The digital future of learning, policy and practice

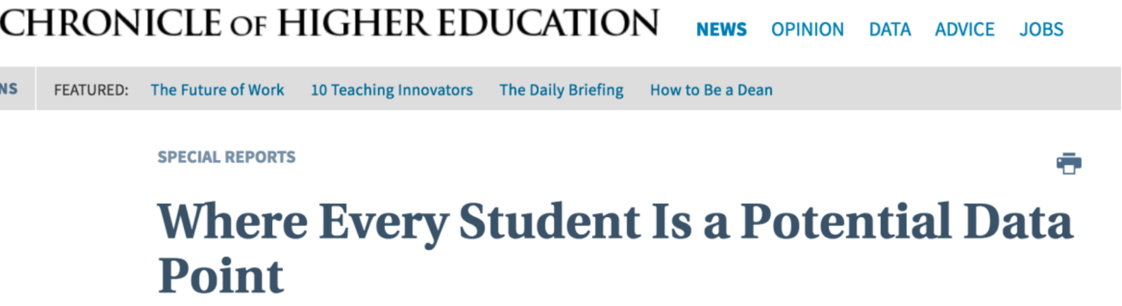
Ben Williamson

2017

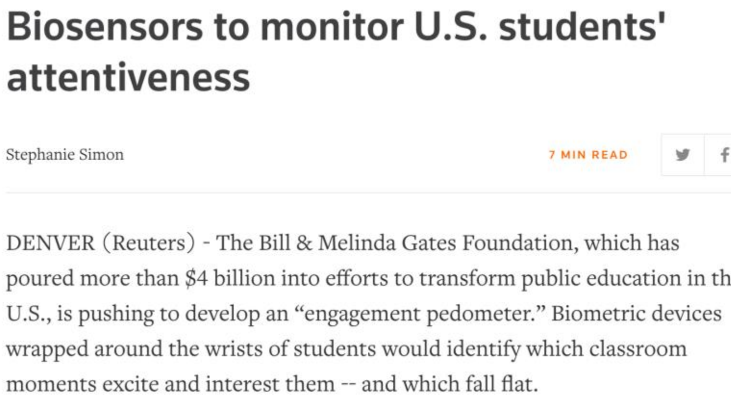
Copyrighted Material



Harvard secretly photographed students to study attendance



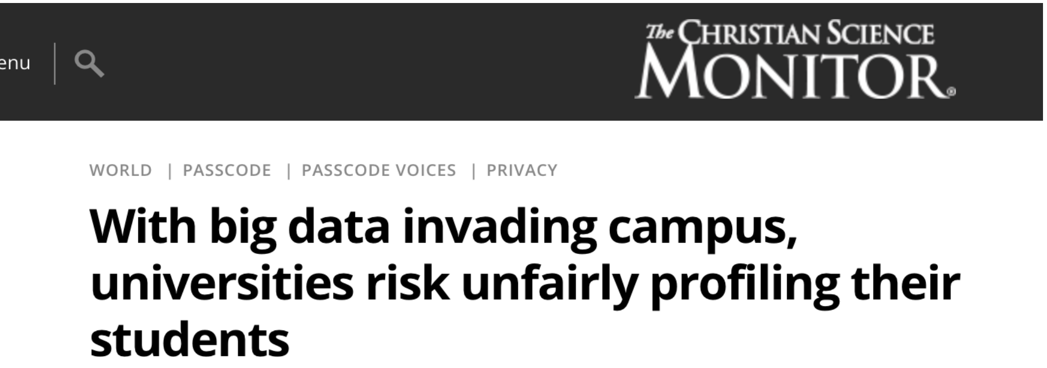
Where Every Student Is a Potential Data Point



Biosensors to monitor U.S. students' attentiveness

Stephanie Simon

DENVER (Reuters) - The Bill & Melinda Gates Foundation, which has poured more than $4 billion into efforts to transform public education in th U.S., is pushing to develop an “engagement pedometer.” Biometric devices wrapped around the wrists of students would identify which classroom moments excite and interest them -- and which fall flat.



enu
The CHRISTIAN SCIENCE MONITOR
WORLD | PASSCODE | PASSCODE VOICES | PRIVACY

With big data invading campus, universities risk unfairly profiling their students

Libraries: Right to read anonymously

Cohen, J. E. (1996). A Right to Read Anonymously: A Closer Look at “Copyright Management” In Cyberspace. *Connecticut Law Review*, *28*, 981–1039.

ALA American Library Association

Freedom to Read Foundation

Free People Read Freely ®

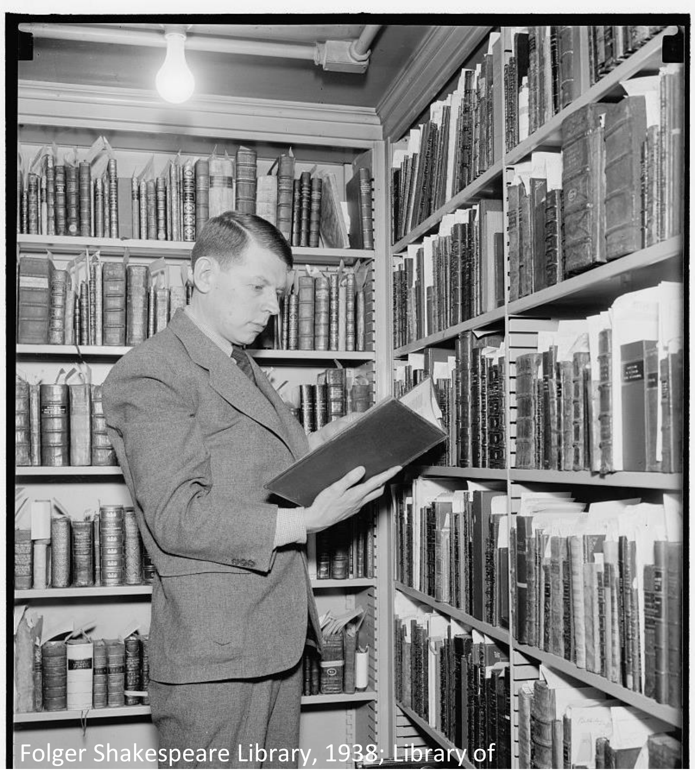
http://cdn.loc.gov/service/pnp/fsa/8d02000/8d02800/8d02865v.jp

Privacy

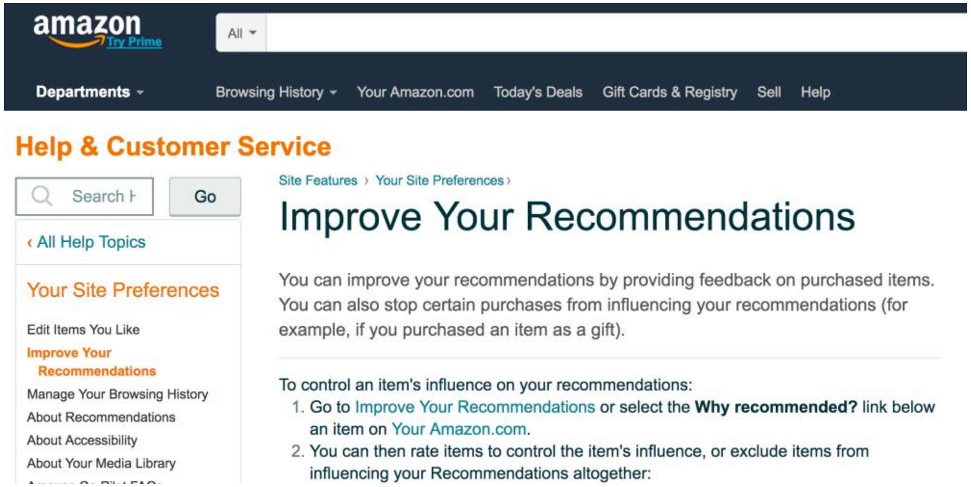
An Interpretation of the Library Bill of Rights

“In a library (physical or virtual), the right to privacy is the right to open inquiry without having the subject of one’s interest examined or scrutinized by others.”

Publishers: Tracking and recommendations



Folger Shakespeare Library, 1938; Library of



Help & Customer Service

Your Site Preferences

- Edit Items You Like
- Improve Your Recommendations
- Manage Your Browsing History
- About Recommendations
- About Accessibility
- About Your Media Library

Site Features › Your Site Preferences ›

Improve Your Recommendations

You can improve your recommendations by providing feedback on purchased items. You can also stop certain purchases from influencing your recommendations (for example, if you purchased an item as a gift).

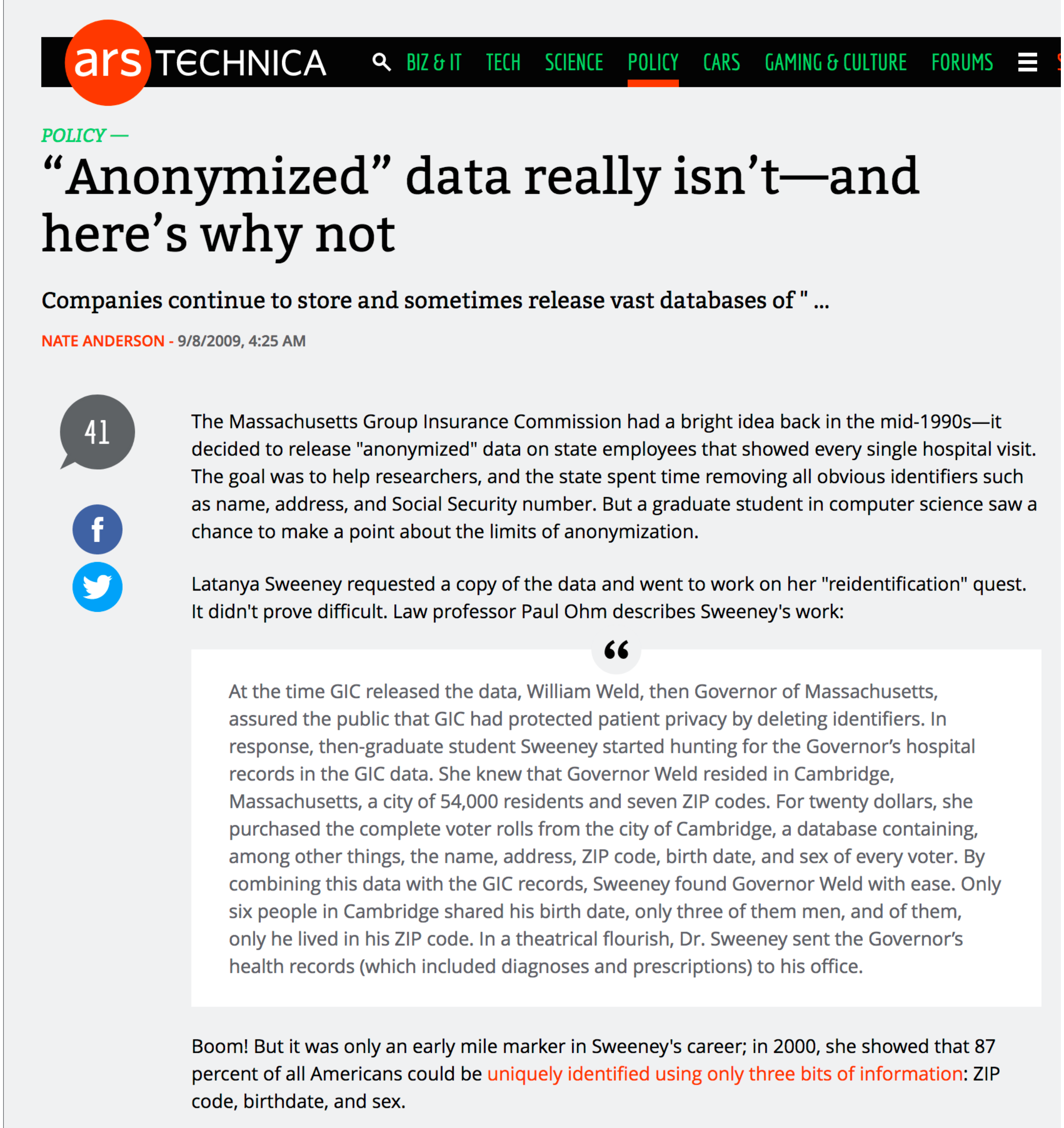
To control an item's influence on your recommendations:

1. Go to Improve Your Recommendations or select the **Why recommended?** link below an item on Your Amazon.com.
2. You can then rate items to control the item's influence, or exclude items from influencing your Recommendations altogether:

The rise of reading analytics and the emerging calculus of reader privacy in the digital world

Clifford Lynch

f i ® s t
m ¤ ñ d @ ¥
PEER-REVIEWED JOURNAL ON THE INTERNET



POLICY —

“Anonymized” data really isn’t—and here’s why not

Companies continue to store and sometimes release vast databases of " ...

NATE ANDERSON - 9/8/2009, 4:25 AM

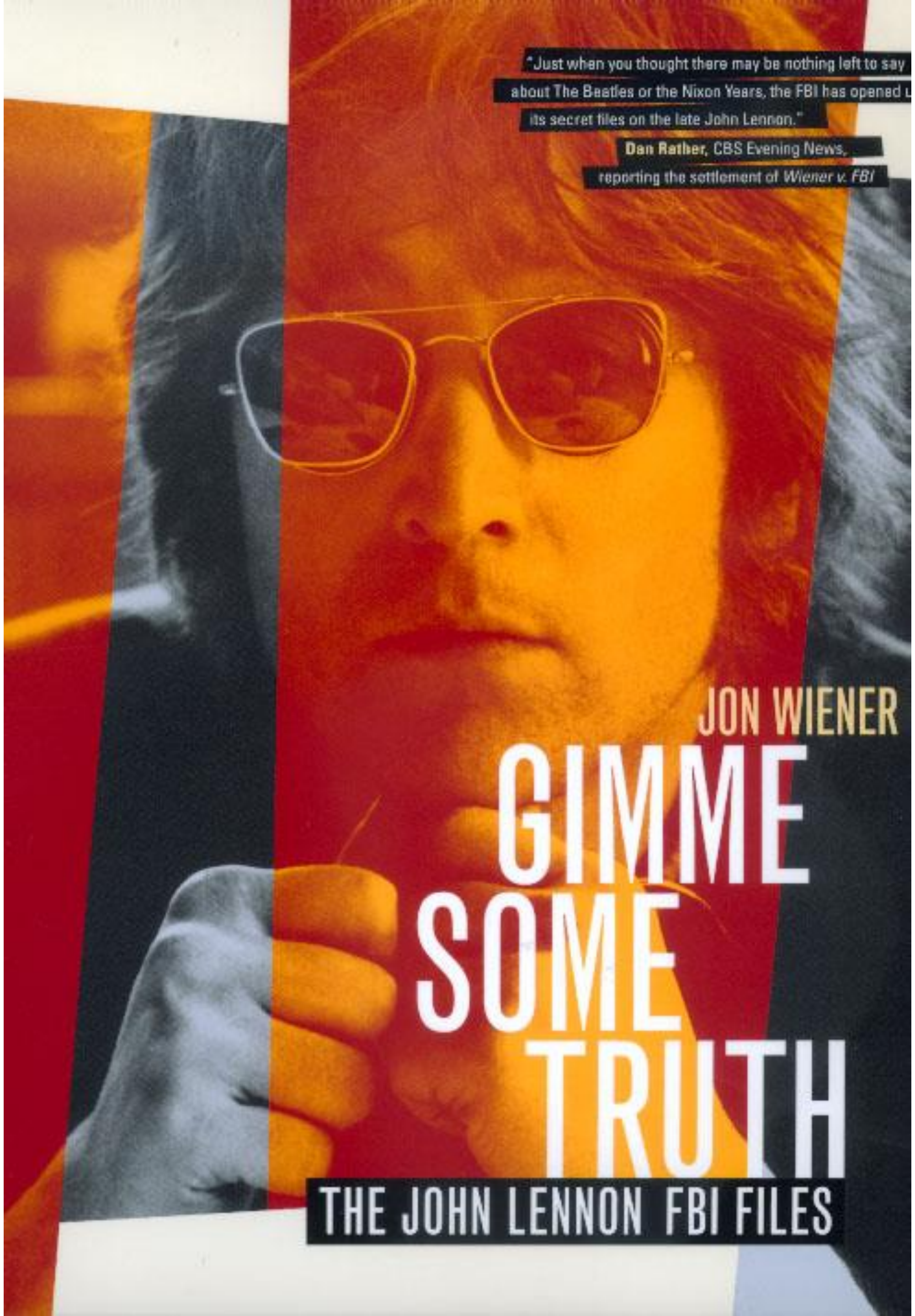
The Massachusetts Group Insurance Commission had a bright idea back in the mid-1990s—it decided to release "anonymized" data on state employees that showed every single hospital visit. The goal was to help researchers, and the state spent time removing all obvious identifiers such as name, address, and Social Security number. But a graduate student in computer science saw a chance to make a point about the limits of anonymization.

Latanya Sweeney requested a copy of the data and went to work on her "reidentification" quest. It didn't prove difficult. Law professor Paul Ohm describes Sweeney's work:

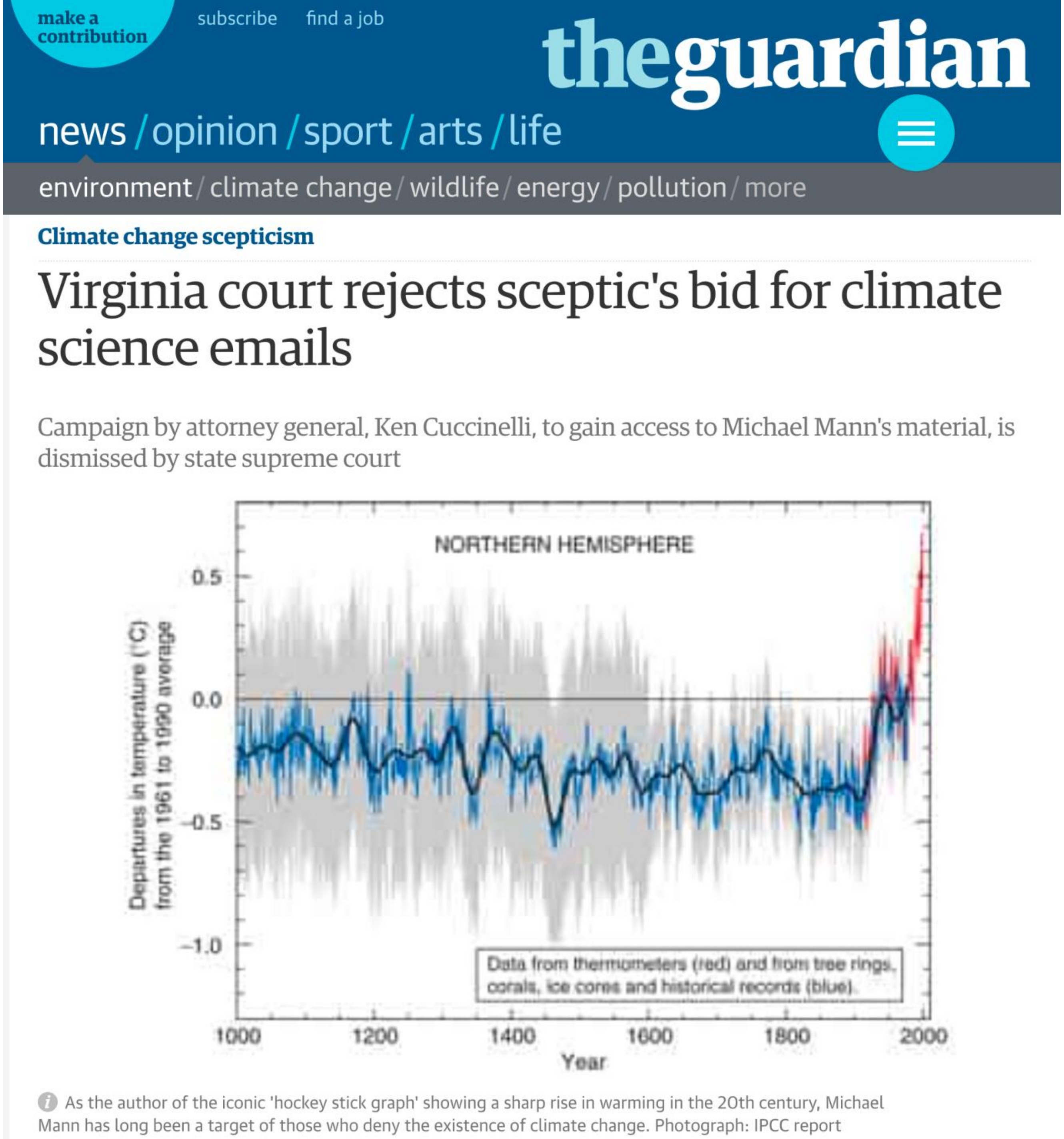
> At the time GIC released the data, William Weld, then Governor of Massachusetts, assured the public that GIC had protected patient privacy by deleting identifiers. In response, then-graduate student Sweeney started hunting for the Governor’s hospital records in the GIC data. She knew that Governor Weld resided in Cambridge, Massachusetts, a city of 54,000 residents and seven ZIP codes. For twenty dollars, she purchased the complete voter rolls from the city of Cambridge, a database containing, among other things, the name, address, ZIP code, birth date, and sex of every voter. By combining this data with the GIC records, Sweeney found Governor Weld with ease. Only six people in Cambridge shared his birth date, only three of them men, and of them, only he lived in his ZIP code. In a theatrical flourish, Dr. Sweeney sent the Governor’s health records (which included diagnoses and prescriptions) to his office.

Boom! But it was only an early mile marker in Sweeney's career; in 2000, she showed that 87 percent of all Americans could be uniquely identified using only three bits of information: ZIP code, birthdate, and sex.

Public Records Requests


"Just when you thought there may be nothing left to say about The Beatles or the Nixon Years, the FBI has opened u its secret files on the late John Lennon."
Dan Rather, CBS Evening News, reporting the settlement of Wiener v. FBI
JON WIENER
GIMME SOME TRUTH
THE JOHN LENNON FBI FILES

Goldenberg, S. (2012, March 2). Virginia court rejects sceptic's bid for climate science emails. *The Guardian*.


make a contribution
subscribe
find a job
theguardian
news / opinion / sport / arts / life
environment / climate change / wildlife / energy / pollution / more

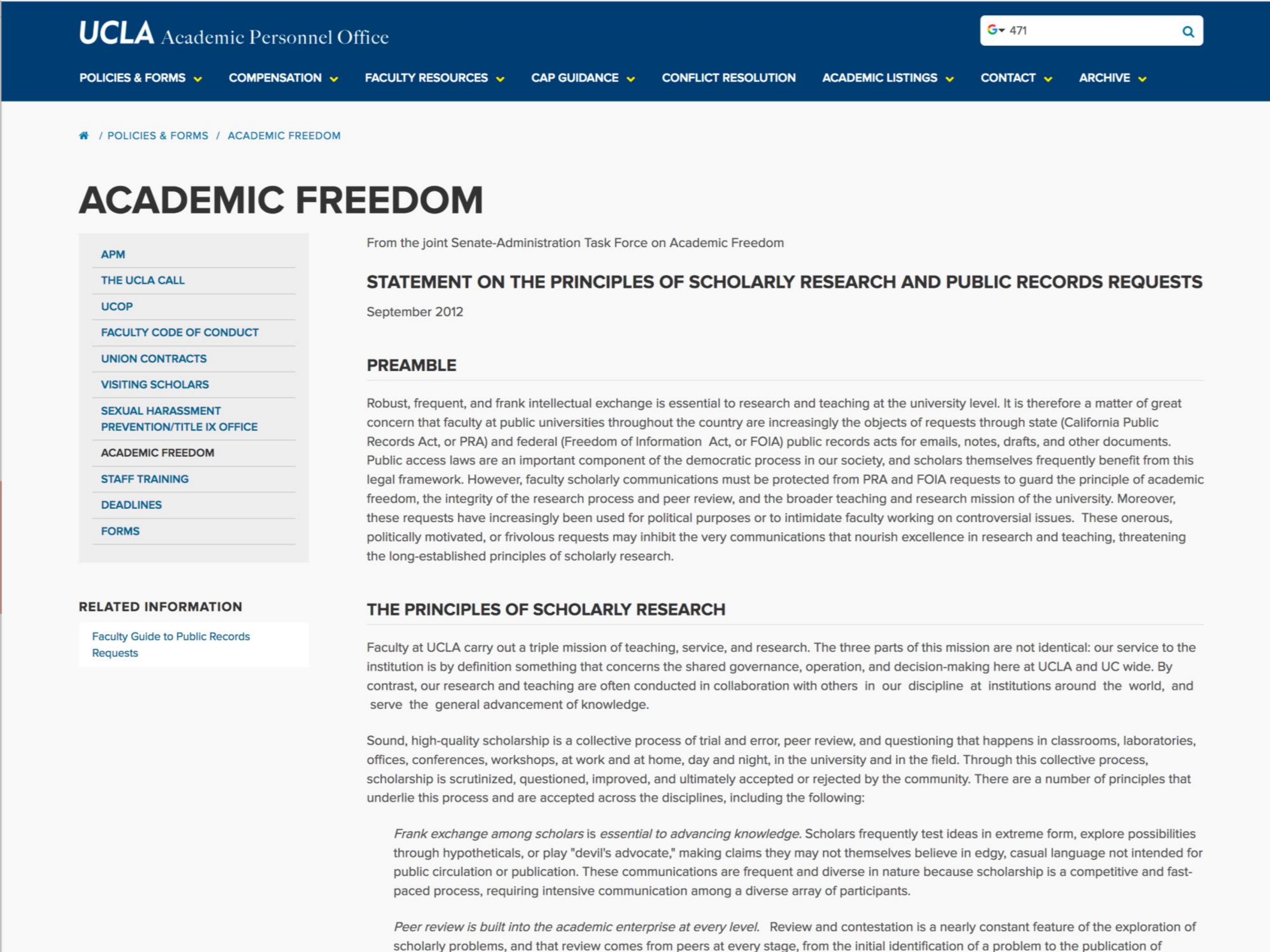
Climate change scepticism

Virginia court rejects sceptic's bid for climate science emails

Campaign by attorney general, Ken Cuccinelli, to gain access to Michael Mann's material, is dismissed by state supreme court

NORTHERN HEMISPHERE
Departures in temperature (°C) from the 1961 to 1990 average
0.5
0.0
−0.5
−1.0
Data from thermometers (red) and from tree rings, corals, ice cores and historical records (blue).
1000
1200
1400
1600
1800
2000
Year

As the author of the iconic 'hockey stick graph' showing a sharp rise in warming in the 20th century, Michael Mann has long been a target of those who deny the existence of climate change. Photograph: IPCC report



ACADEMIC FREEDOM

From the joint Senate-Administration Task Force on Academic Freedom

STATEMENT ON THE PRINCIPLES OF SCHOLARLY RESEARCH AND PUBLIC RECORDS REQUESTS

September 2012

PREAMBLE

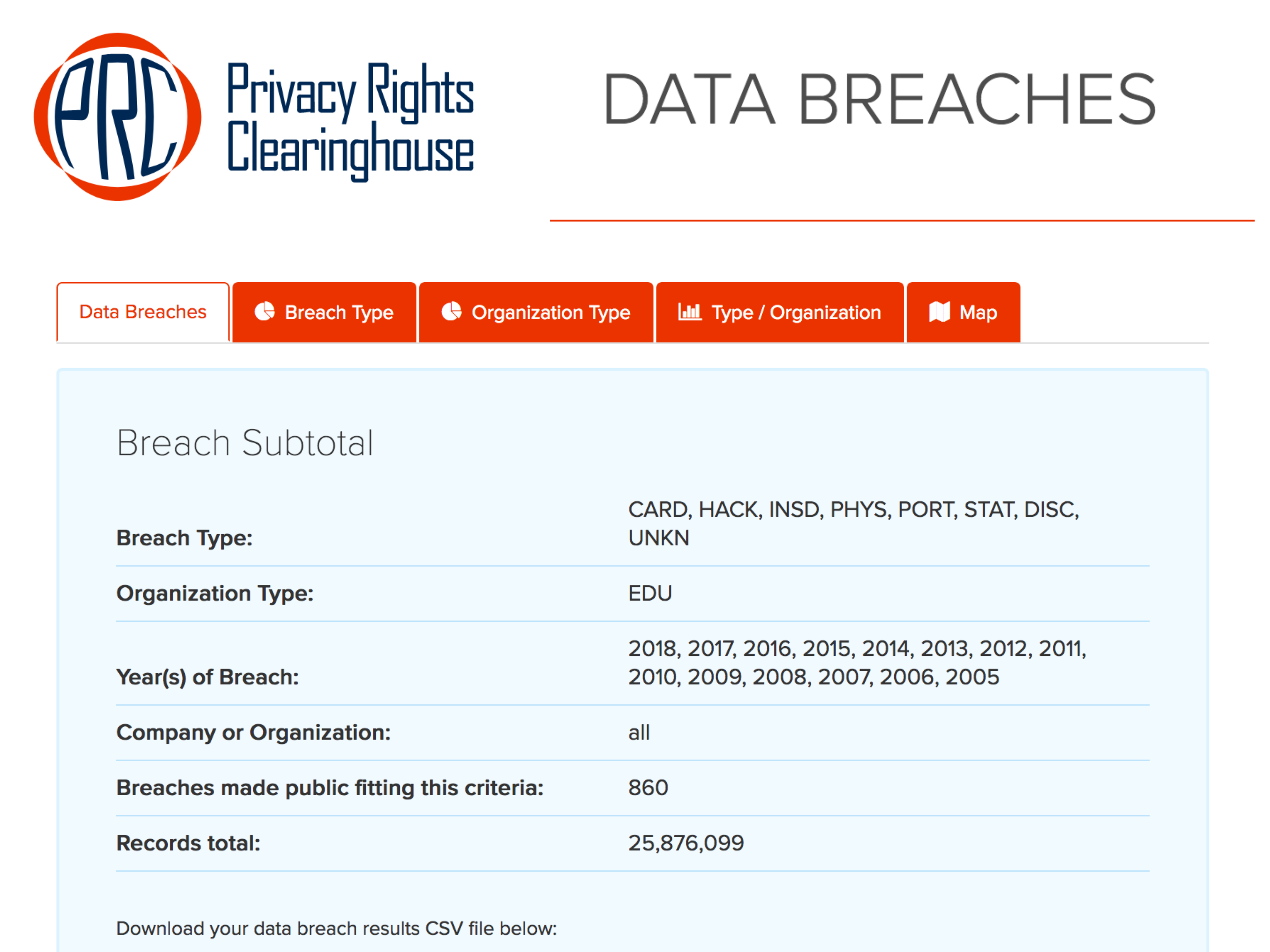
Robust, frequent, and frank intellectual exchange is essential to research and teaching at the university level. It is therefore a matter of great concern that faculty at public universities throughout the country are increasingly the objects of requests through state (California Public Records Act, or PRA) and federal (Freedom of Information Act, or FOIA) public records acts for emails, notes, drafts, and other documents. Public access laws are an important component of the democratic process in our society, and scholars themselves frequently benefit from this legal framework. However, faculty scholarly communications must be protected from PRA and FOIA requests to guard the principle of academic freedom, the integrity of the research process and peer review, and the broader teaching and research mission of the university. Moreover, these requests have increasingly been used for political purposes or to intimidate faculty working on controversial issues. These onerous, politically motivated, or frivolous requests may inhibit the very communications that nourish excellence in research and teaching, threatening the long-established principles of scholarly research.

THE PRINCIPLES OF SCHOLARLY RESEARCH

Faculty at UCLA carry out a triple mission of teaching, service, and research. The three parts of this mission are not identical: our service to the institution is by definition something that concerns the shared governance, operation, and decision-making here at UCLA and UC wide. By contrast, our research and teaching are often conducted in collaboration with others in our discipline at institutions around the world, and serve the general advancement of knowledge.

Sound, high-quality scholarship is a collective process of trial and error, peer review, and questioning that happens in classrooms, laboratories, offices, conferences, workshops, at work and at home, day and night, in the university and in the field. Through this collective process, scholarship is scrutinized, questioned, improved, and ultimately accepted or rejected by the community. There are a number of principles that underlie this process and are accepted across the disciplines, including the following:

> *Frank exchange among scholars* is *essential to advancing knowledge.* Scholars frequently test ideas in extreme form, explore possibilities through hypotheticals, or play "devil's advocate," making claims they may not themselves believe in edgy, casual language not intended for public circulation or publication. These communications are frequent and diverse in nature because scholarship is a competitive and fast-paced process, requiring intensive communication among a diverse array of participants.
>
> *Peer review is built into the academic enterprise at every level.* Review and contestation is a nearly constant feature of the exploration of scholarly problems, and that review comes from peers at every stage, from the initial identification of a problem to the publication of

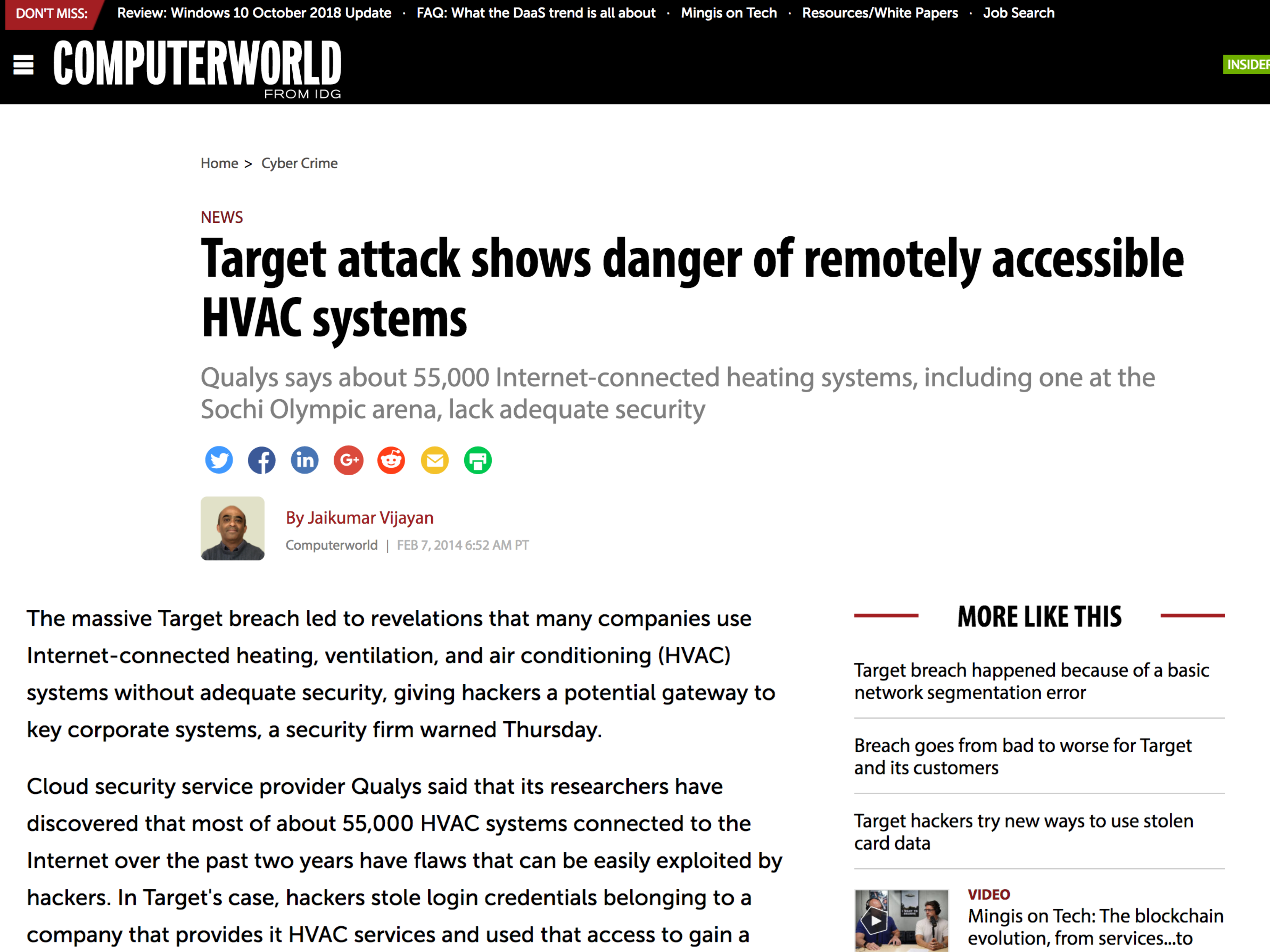


DATA BREACHES

Breach Subtotal

Breach Type:	CARD, HACK, INSD, PHYS, PORT, STAT, DISC, UNKN
Organization Type:	EDU
Year(s) of Breach:	2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005
Company or Organization:	all
Breaches made public fitting this criteria:	860
Records total:	25,876,099

Download your data breach results CSV file below:



COMPUTERWORLD
FROM IDG

Home > Cyber Crime

NEWS

Target attack shows danger of remotely accessible HVAC systems

Qualys says about 55,000 Internet-connected heating systems, including one at the Sochi Olympic arena, lack adequate security

By Jaikumar Vijayan
Computerworld | FEB 7, 2014 6:52 AM PT

The massive Target breach led to revelations that many companies use Internet-connected heating, ventilation, and air conditioning (HVAC) systems without adequate security, giving hackers a potential gateway to key corporate systems, a security firm warned Thursday.

Cloud security service provider Qualys said that its researchers have discovered that most of about 55,000 HVAC systems connected to the Internet over the past two years have flaws that can be easily exploited by hackers. In Target's case, hackers stole login credentials belonging to a company that provides it HVAC services and used that access to gain a

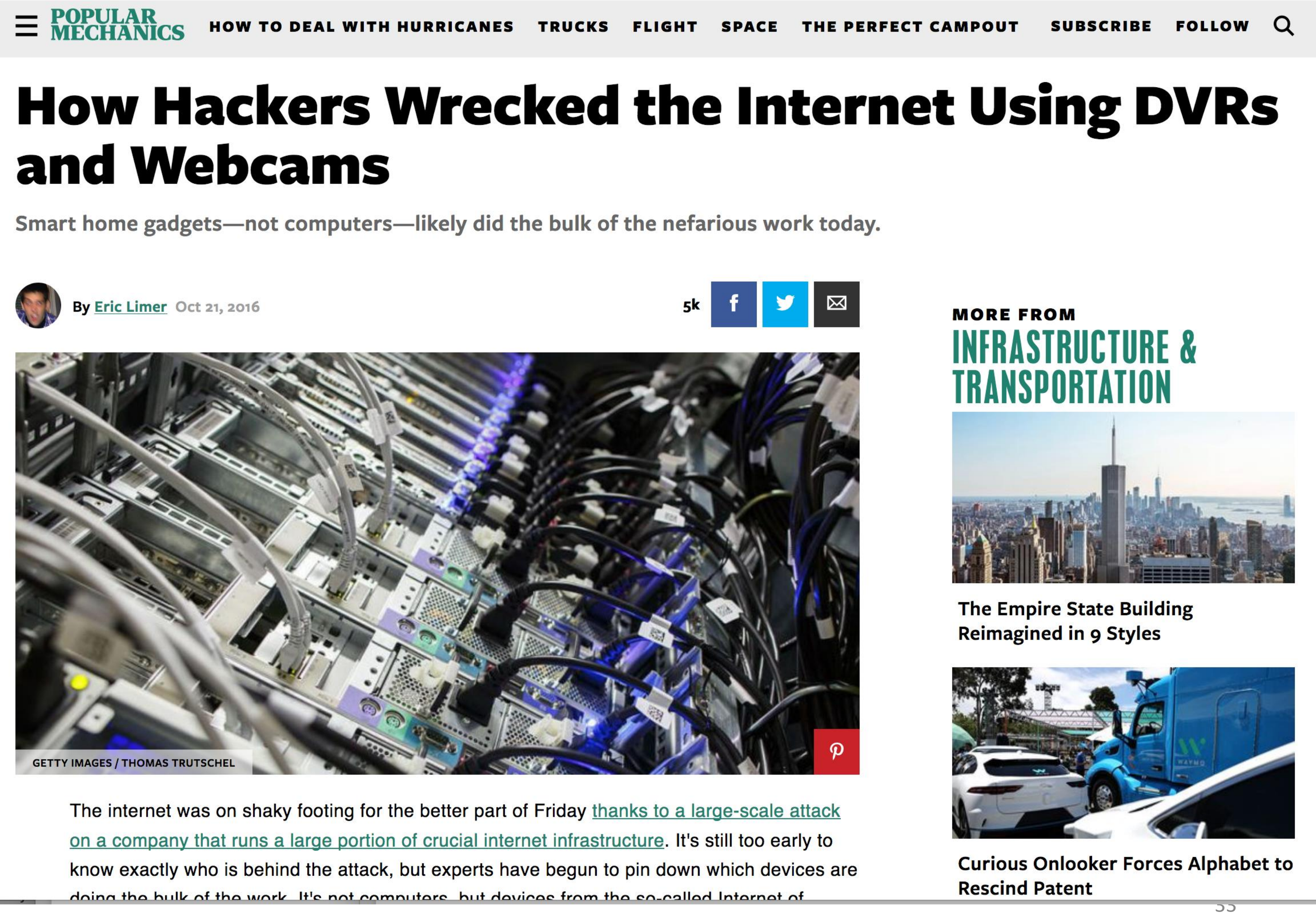
MORE LIKE THIS

Target breach happened because of a basic network segmentation error

Breach goes from bad to worse for Target and its customers

Target hackers try new ways to use stolen card data

VIDEO
Mingis on Tech: The blockchain evolution, from services...to



How Hackers Wrecked the Internet Using DVRs and Webcams

Smart home gadgets—not computers—likely did the bulk of the nefarious work today.

By Eric Limer Oct 21, 2016

GETTY IMAGES / THOMAS TRUTSCHEL

The internet was on shaky footing for the better part of Friday thanks to a large-scale attack on a company that runs a large portion of crucial internet infrastructure. It's still too early to know exactly who is behind the attack, but experts have begun to pin down which devices are doing the bulk of the work. It's not computers, but devices from the so-called Internet of

MORE FROM

INFRASTRUCTURE & TRANSPORTATION

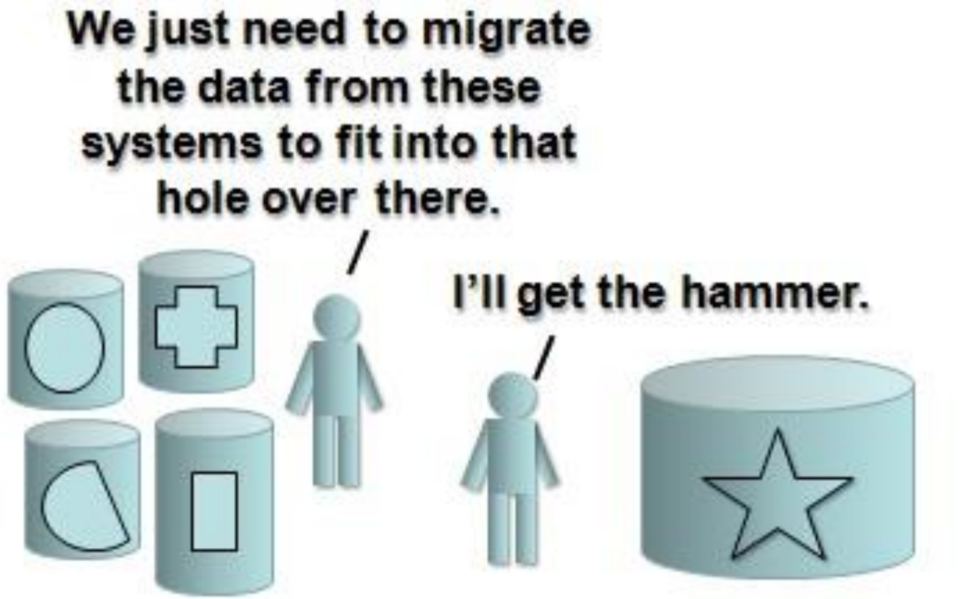
The Empire State Building Reimagined in 9 Styles

Curious Onlooker Forces Alphabet to Rescind Patent

Data Stewardship

http://www.information-age.com/cloud-computing-pharmaceutical-industry-123462676/

Getty Research Institute



We just need to migrate the data from these systems to fit into that hole over there.
I'll get the hammer.

http://www.datamartist.com/data-migration-part-1-introduction-to-the-data-migration-delema

Mount Wilson Solar Observatory, 2017

http://gsa.rice.edu/

Graduate students

https://med.nyu.edu/our-community/life-nyu-school-medicine/life-postdoc

Post-doctoral fellows

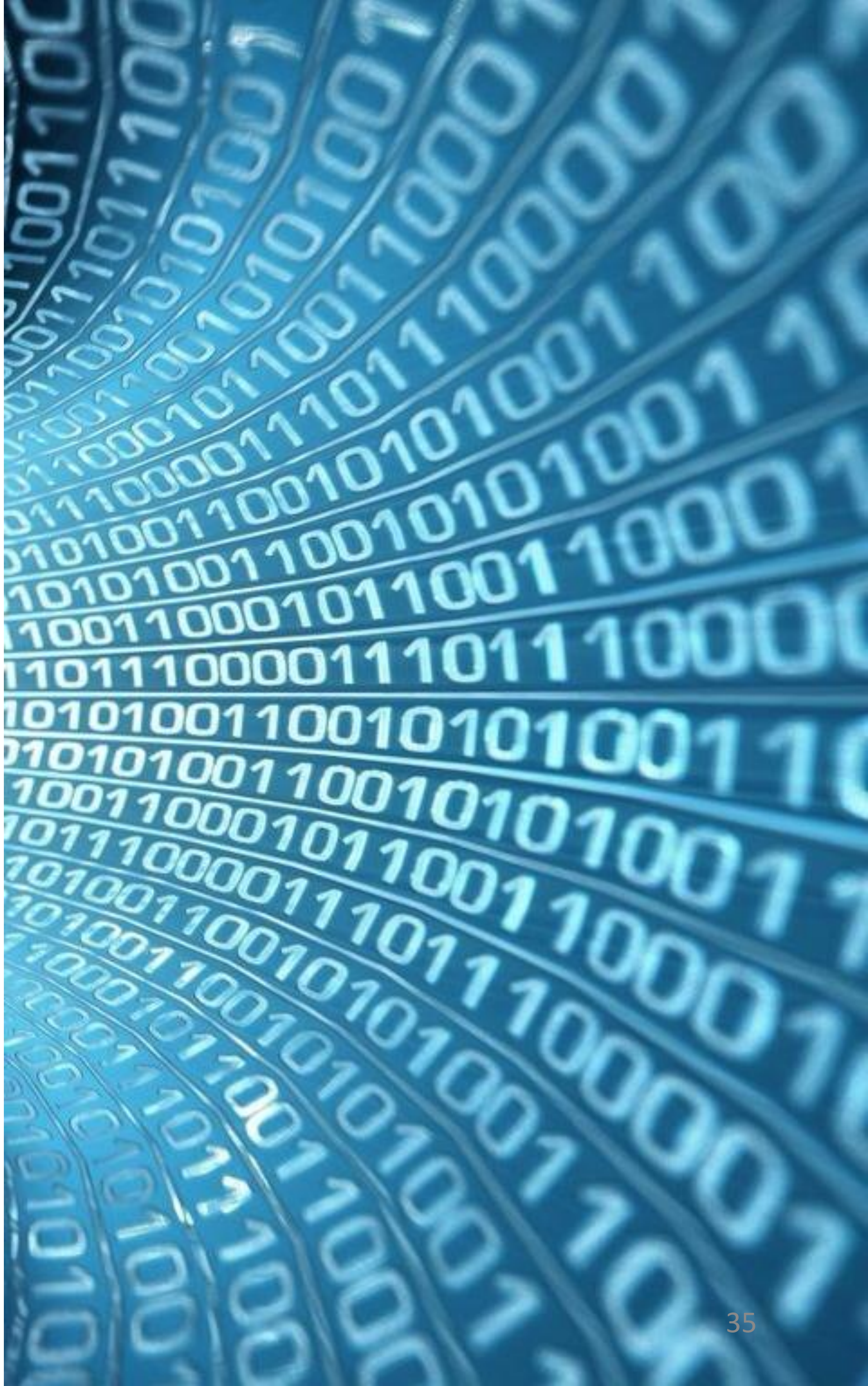
Data

If you can’t protect it, don’t collect it.

(privacy and security aphorism)

Therefore:

If you collect it, you must protect it.

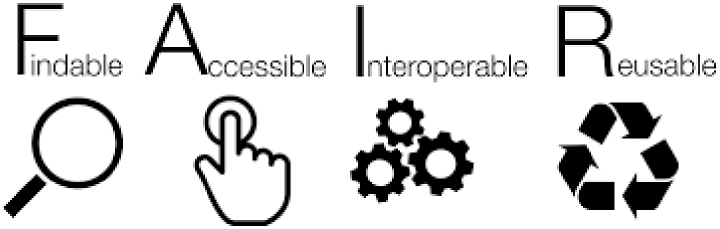


open by design

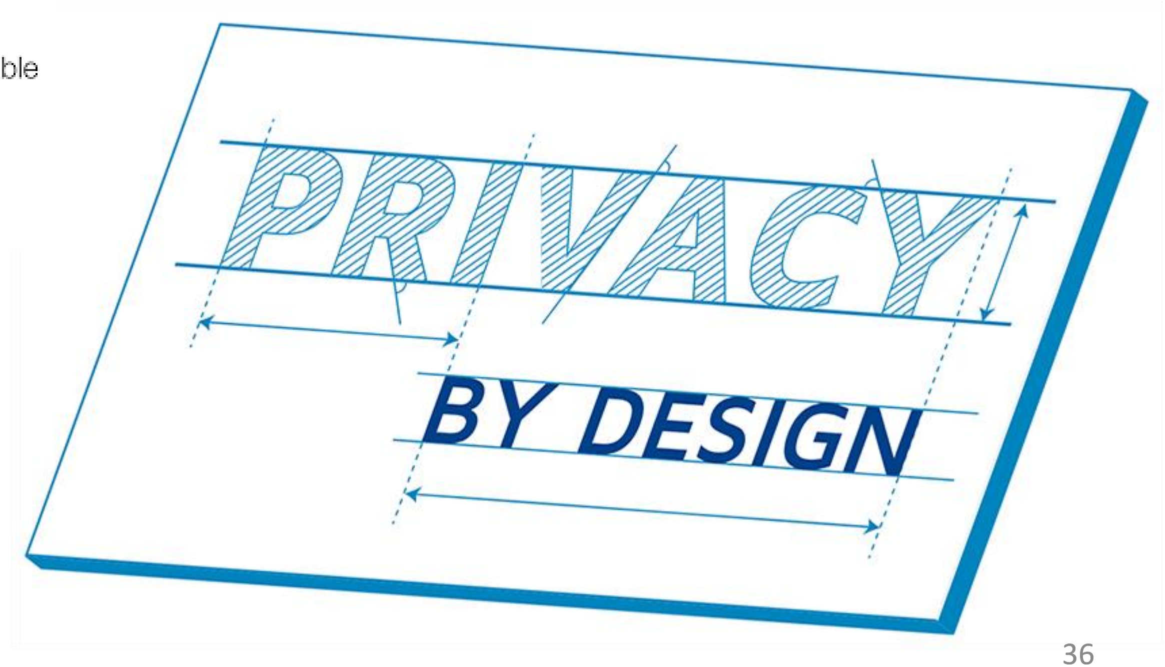
http://democracyos.eu/blog/open-by-design

OPEN DATA

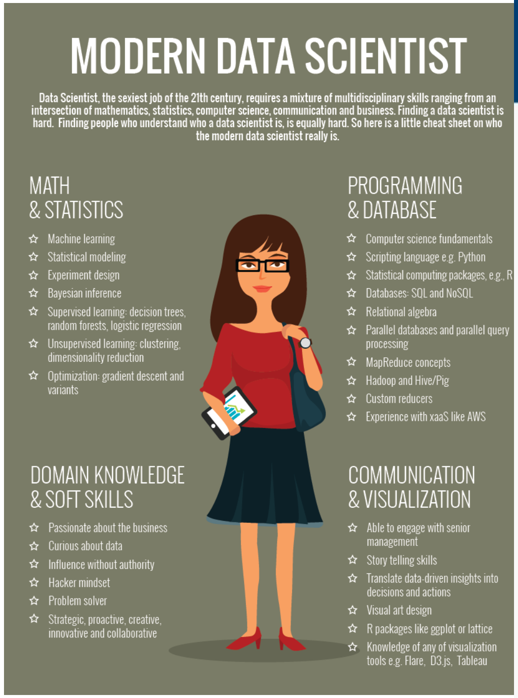
https://wwwdb.inf.tu-dresden.de/opendatasurvey/



Wilkinson, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, http://dx.doi.org/10.1038/sdata.2016.18



https://privacybydesign.foundation/en/



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing packages, e.g., R
- Databases: SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop and Hive/Pig
- Custom reducers
- Experience with xaaS like AWS

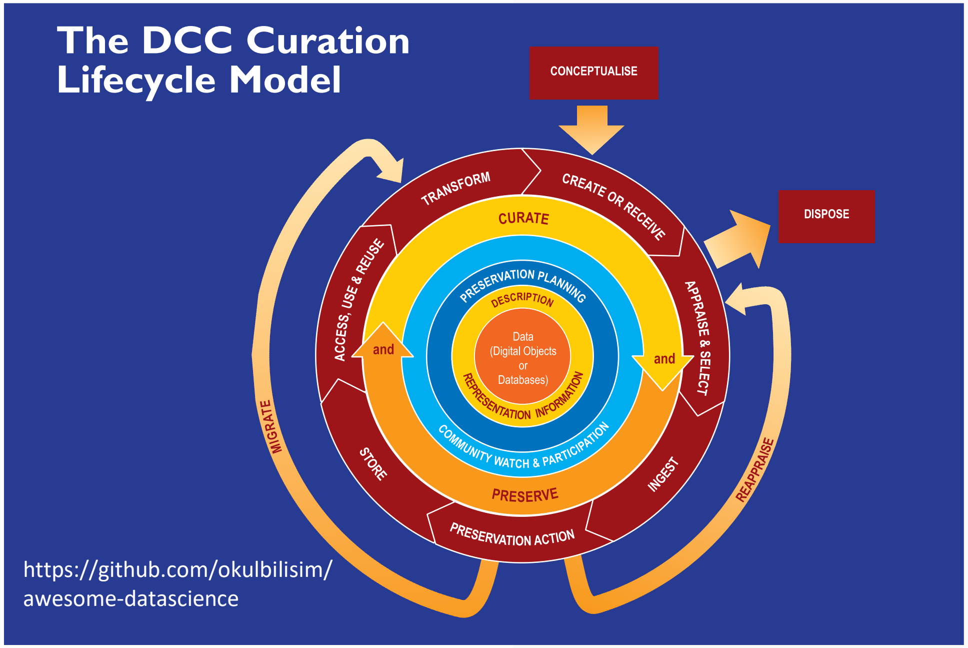
DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

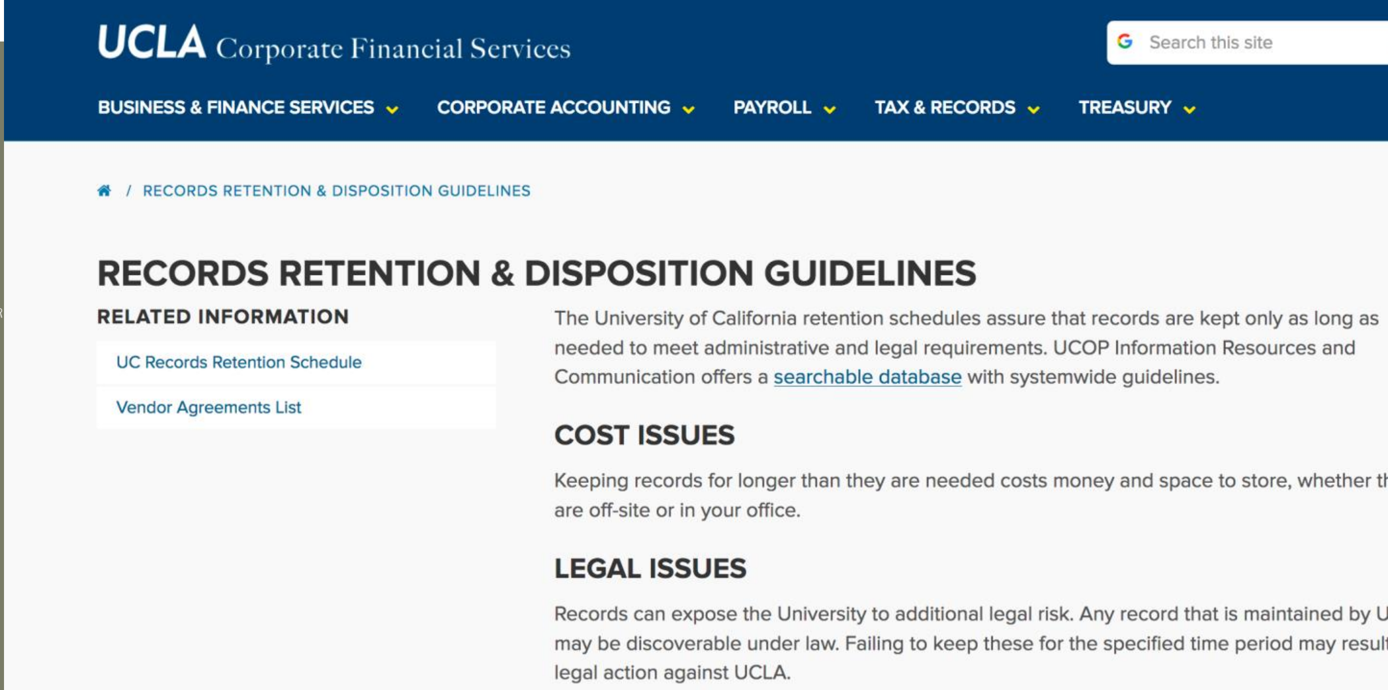
- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

JISC www.dcc.ac.uk info@dcc.ac.uk



The DCC Curation Lifecycle Model

https://github.com/okulbilisim/awesome-datascience



UCLA Corporate Financial Services

/ RECORDS RETENTION & DISPOSITION GUIDELINES

RECORDS RETENTION & DISPOSITION GUIDELINES

RELATED INFORMATION

UC Records Retention Schedule

Vendor Agreements List

The University of California retention schedules assure that records are kept only as long as needed to meet administrative and legal requirements. UCOP Information Resources and Communication offers a searchable database with systemwide guidelines.

COST ISSUES

Keeping records for longer than they are needed costs money and space to store, whether th are off-site or in your office.

LEGAL ISSUES

Records can expose the University to additional legal risk. Any record that is maintained by U may be discoverable under law. Failing to keep these for the specified time period may resul legal action against UCLA.

COPIES VS. ORIGINALS

Records that are held past their retention date are still subject to subpoena as are copies of f known as shadow files. Contact the Office of Record prior to destroying your copies.

ELECTRONIC FILES

Retention does not apply only to paper records, but to electronic records too. This means it is necessary to erase certain computer files, including emails, over time, or they too will be discoverable.

DESTROYING RECORDS

Records must be destroyed in accordance with the University's records retention policies. Documents that contain personal or sensitive information should be shredded.

If you have a lot of records to dispose of, check the Vendor Agreements List to find who has a contract with UCLA for document destruction. For smaller volumes it may be a good option tc buy a cross-cut shredder.

If you would rather use another vendor, contact Campus Purchasing. If a third party shreds yc documents, be sure to obtain a certificate of completion to verify that the items have been destroyed properly.

Remember that confidential records must be protected throughout the entire process.

Consult your IT manager for ways to destroy electronic records safely and securely.

Promote Responsible Data Practices

- Respect information and autonomy privacy
  - Open data: release and reuse
  - Data collection and use
  - Data management
  - Collaborations
  - Publications
- Community
  - Faculty
  - Staff
  - Students
  - External partners
- Joint governance process

https://www.universityofcalifornia.edu/subject/term/technology-engineering

http://www.berkeley.edu/utility/jobs

http://gsa.rice.edu/

https://www.commondreams.org/views/2014/09/20/corporations-your-diet

Summary and Takeaways

- Data are university assets: Exploit and protect
- Privacy in context: Information, autonomy
- Stewardship in context: Preserve or purge
- Open data: Reuse and risk
- Security: More data, bigger targets
- Data aggregation: Power and privacy
- Data governance: Ownership, responsibility

Acknowledgements

- UCOP Privacy and Information Security Initiative
- UCLA Data Governance Task Force
- UCLA Board on Privacy and Data Protection
- UC Academic Computing and Communications Committee
- Electronic Privacy Information Center
- Berkeley Ctr for Law and Technology
- Marc Rotenberg, Kent Wada, Jim Davis, Amy Blum, Scott Waugh, Dana Cuff, Jerry Kang, Leah Lievrouw, David Kay, Maryann Martone, Joanne Miller, Jim Chalfant, Shane White, Clifford Lynch, Anne Washington, Sheryl Vacca, Gene Lucas, Aimee Dorr, Tom Andriola, David Rusting …

UCLA Center for Knowledge Infrastructures

ALFRED P. SLOAN FOUNDATION 1934

Christine Borgman

Peter Darch

Irene Pasquetto

Bernie Boscoe

Michael Scroggins

Milena Golshan

Cheryl Thompson

Morgan Wofford