

UCLA

Presentations

Title

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

Permalink

<https://escholarship.org/uc/item/15x0w54k>

Author

Borgman, Christine L.

Publication Date

2017-11-16

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

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

Christine L. Borgman

Distinguished Professor & Presidential Chair in Information Studies

Director, Center for Knowledge Infrastructures

<https://knowledgeinfrastructures.gseis.ucla.edu>

University of California, Los Angeles

<http://christineborgman.info>

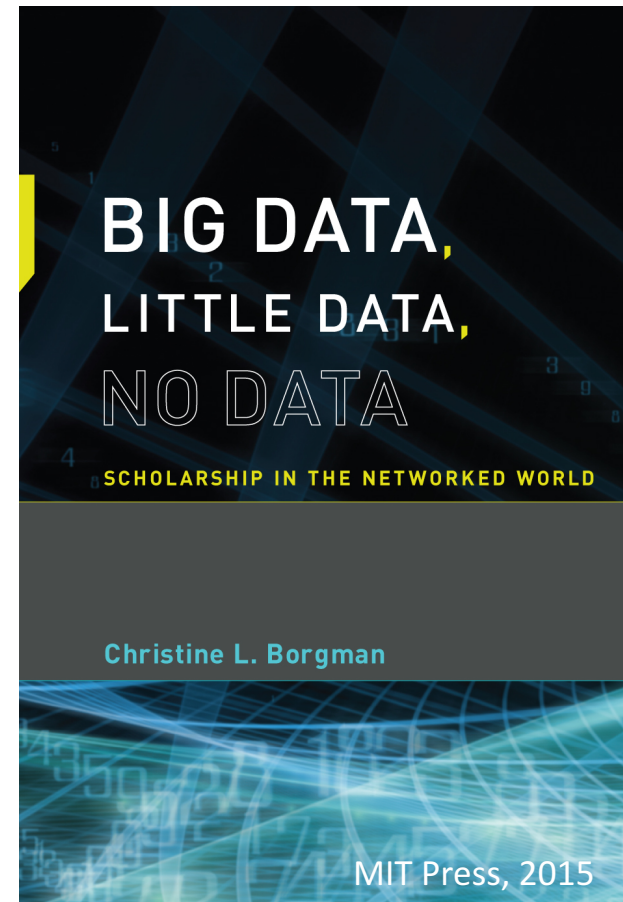
@scitechprof

Berkeley Center for Law and Technology

Tenth Annual Privacy Lecture

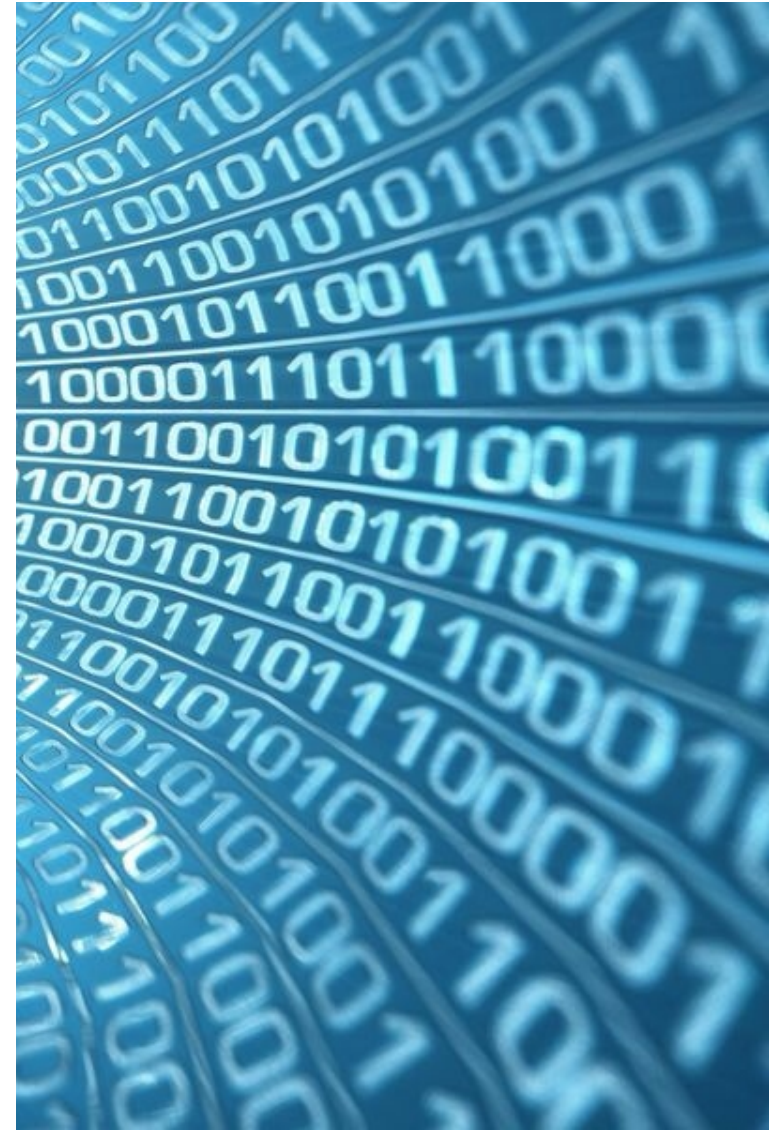
<https://www.law.berkeley.edu/research/bclt>

November 16, 2017



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



Open access to data

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



Supported by
wellcometrust



Australian Government
National Health and Medical Research Council



National Science Foundation
WHERE DISCOVERIES BEGIN

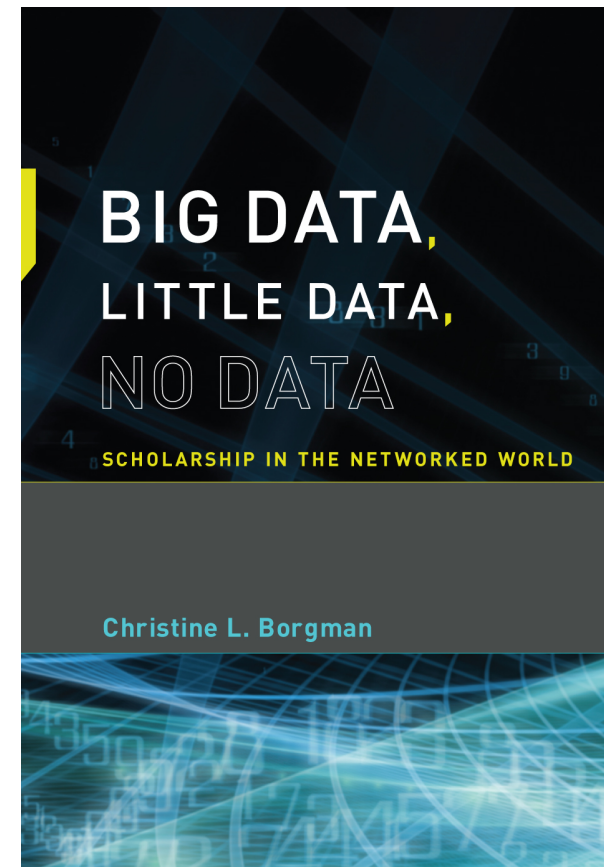


Policy RECommendations for Open Access to Research Data in Europe



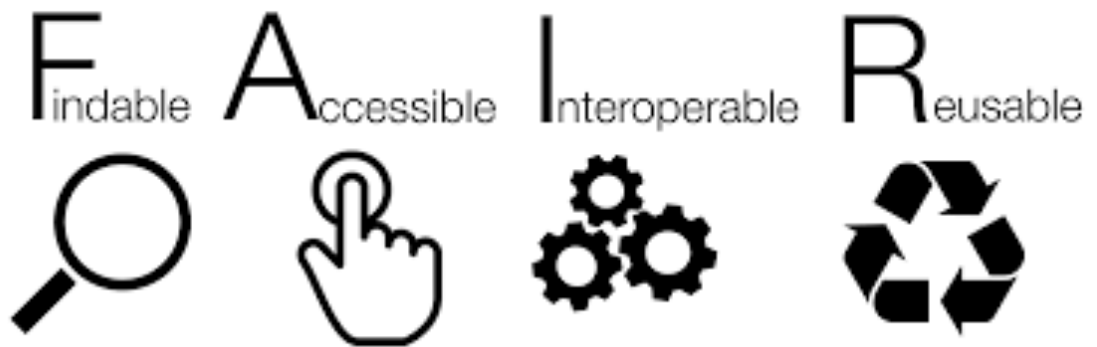
Why Open Access to Data?

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



How to Open Data

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

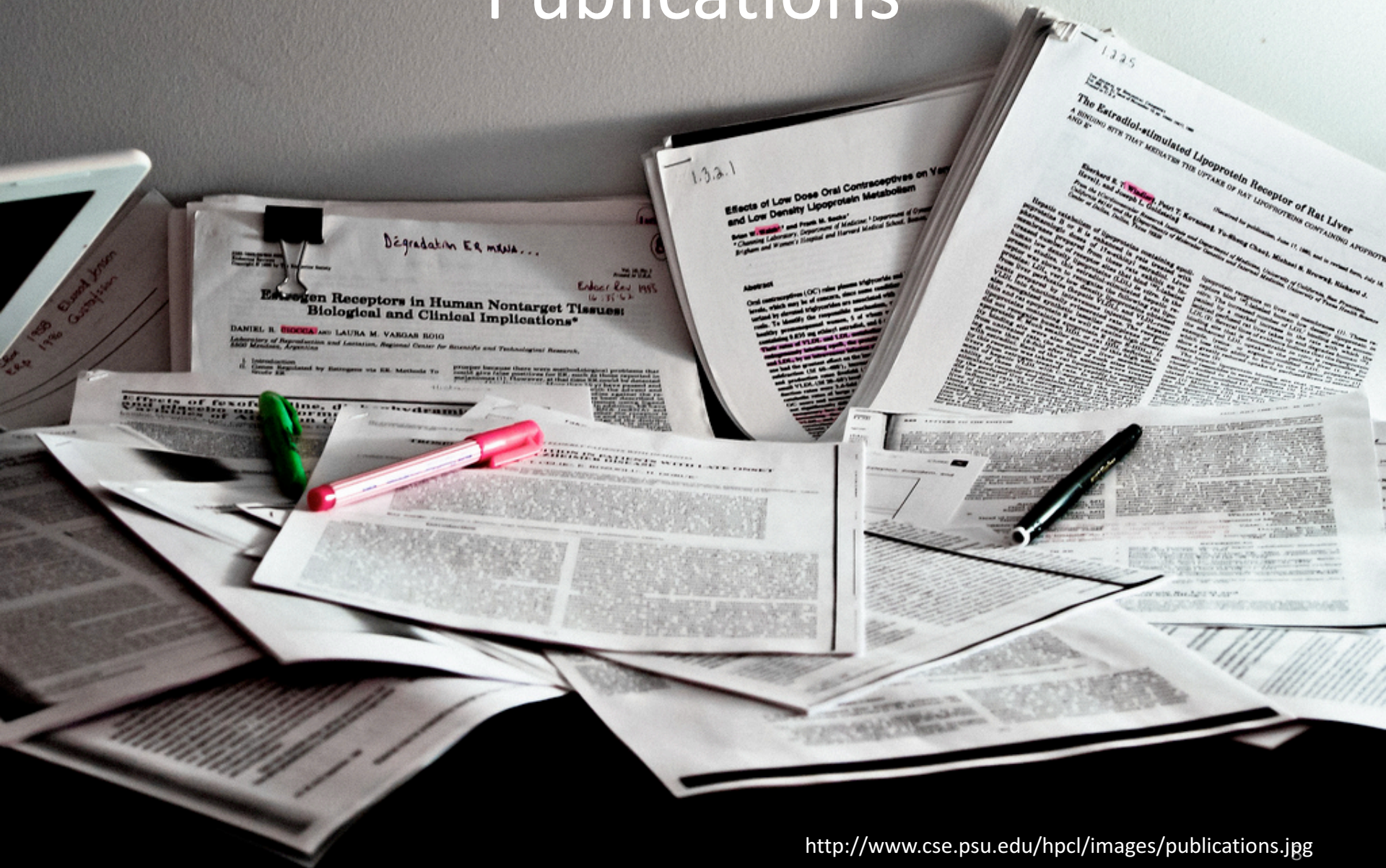




Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

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

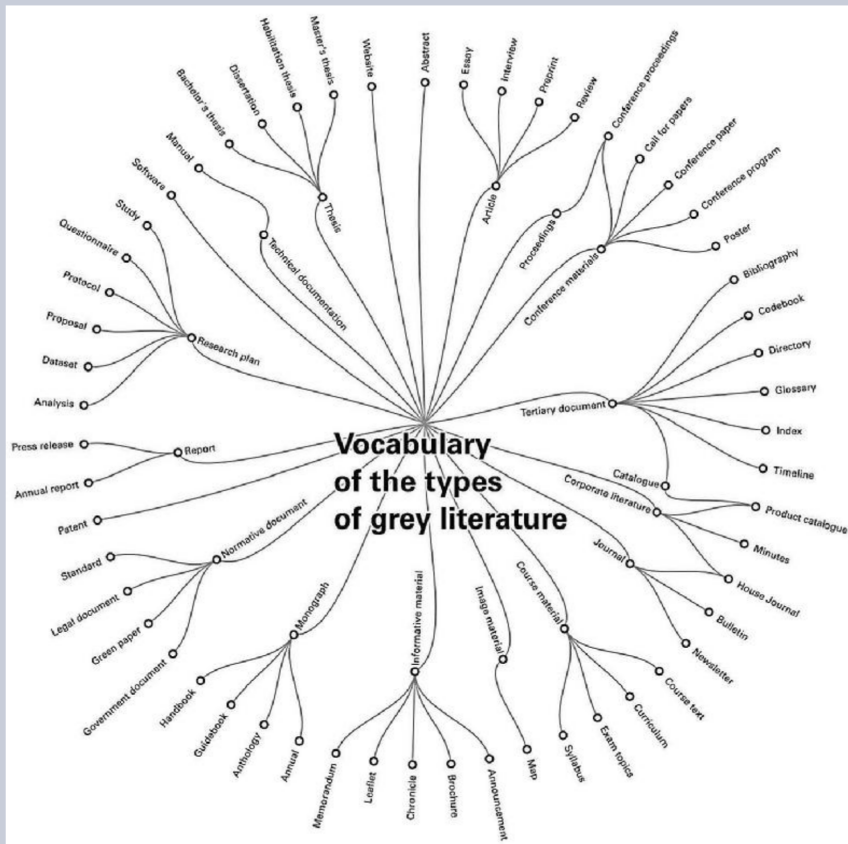
Publications



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.nu.nl/cz/record/81435?ln=en>.



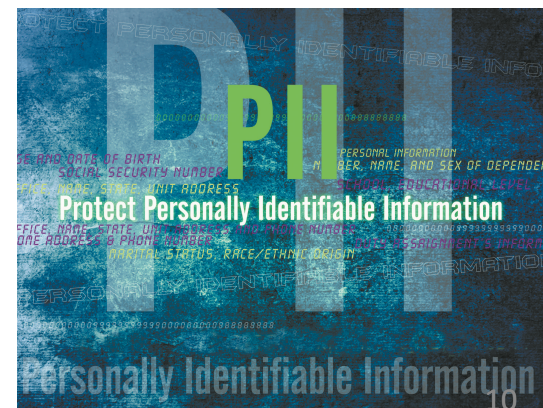
Grey Literature Typology

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

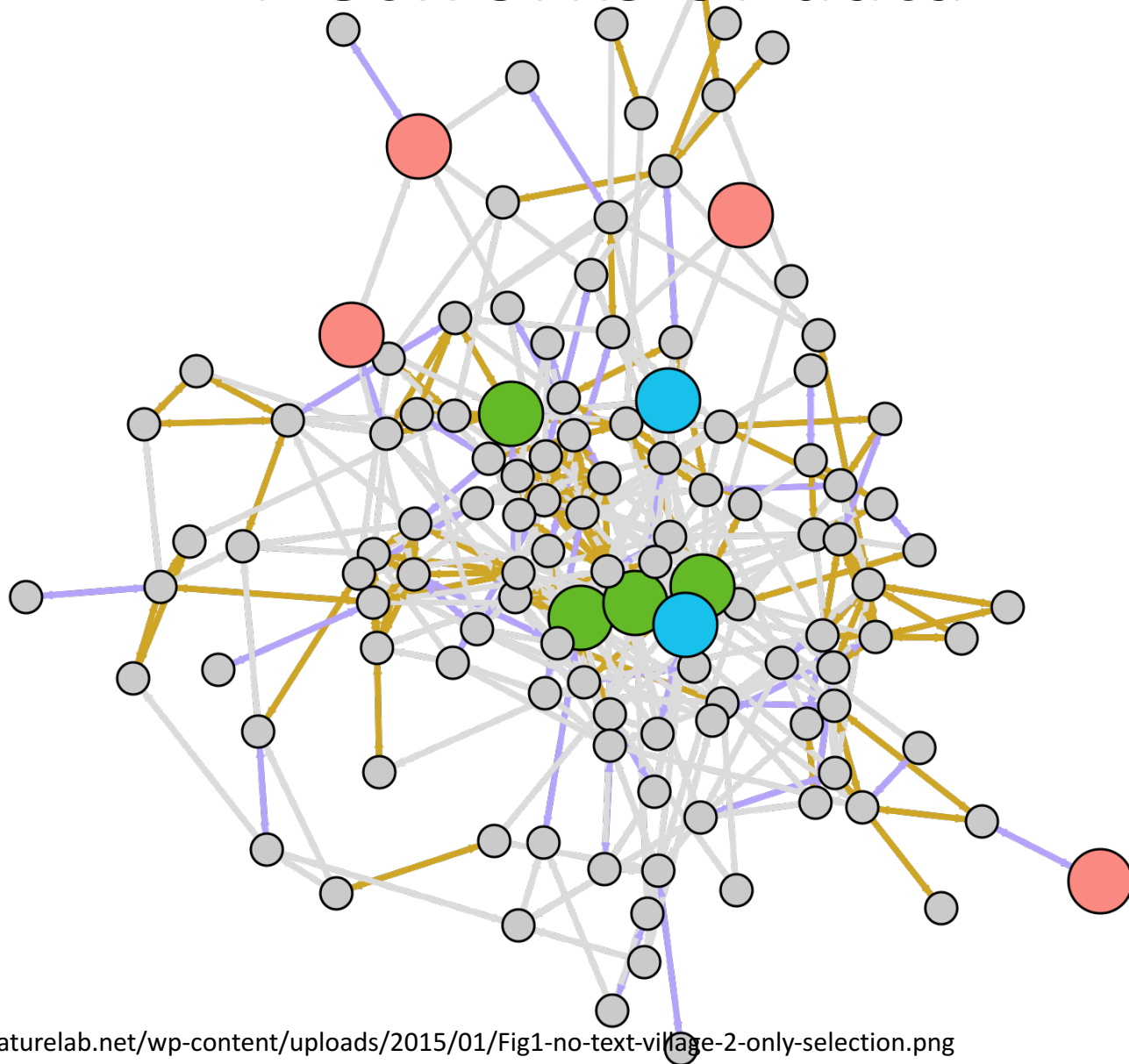


Grey Data

- 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...



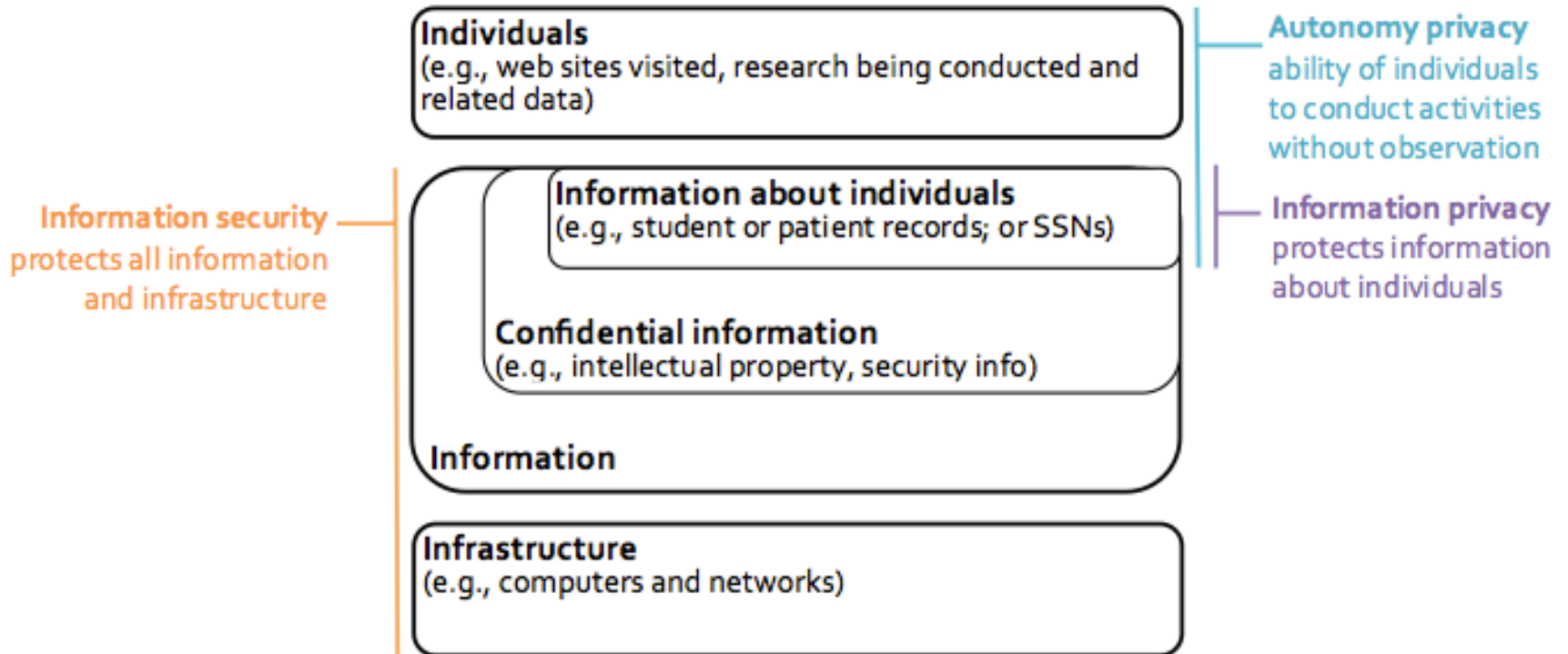
Networks of data



University Responsibilities for Data

- Privacy
- Academic and intellectual freedom
- Stewardship and governance

Information and Autonomy Privacy



Information Privacy

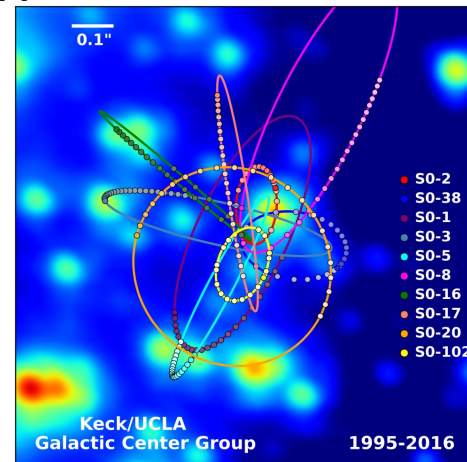


Autonomy Privacy

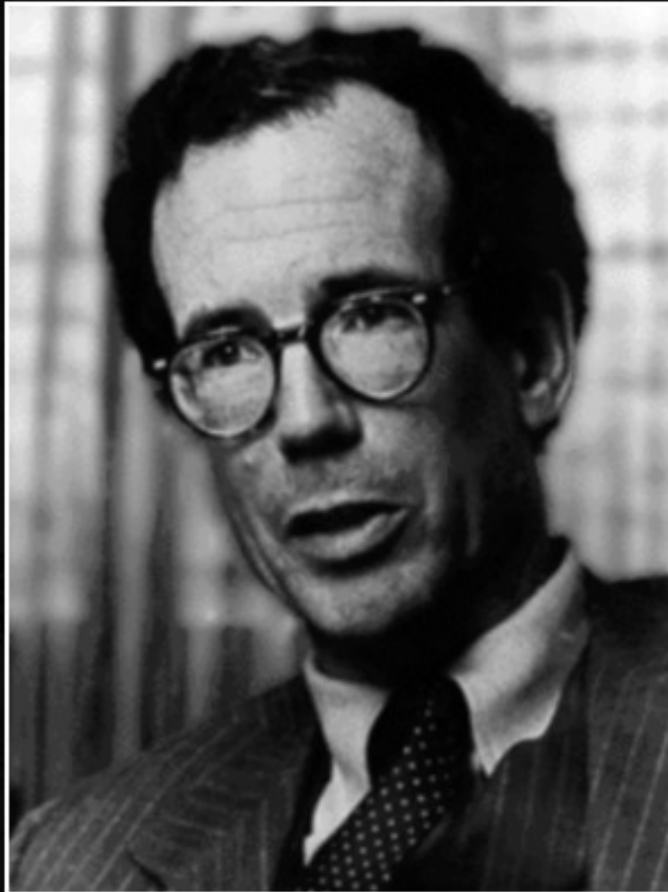
- Ability of individuals to conduct activities without observation
 - Intellectual inquiry
 - Conducting research
 - Classroom discussions
 - Searching for information
 - Email, web browsing
 - Reading



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/>



Academic Freedom



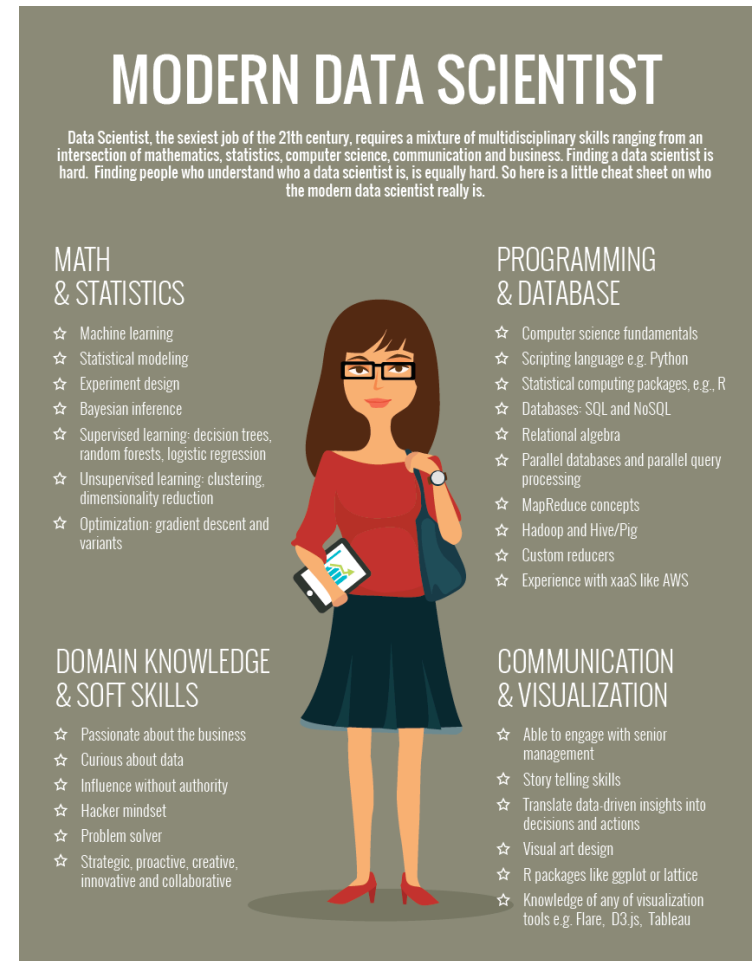
Academic freedom really means freedom of inquiry. To be able to probe according to one's own interest, knowledge and conscience is the most important freedom the scholar has, and part of that process is to state its results.

— *Donald Kennedy* —

AZ QUOTES

Stewardship and Governance

- Protect
 - Privacy: information, autonomy
 - Academic freedom
- Secure infrastructure
- Data management
 - Findable
 - Accessible
 - Interoperable
 - Reusable
- Governance
 - Principles
 - Processes



Privacy Frontier: Open Data

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

Uses and Misuses of Data

Reuse

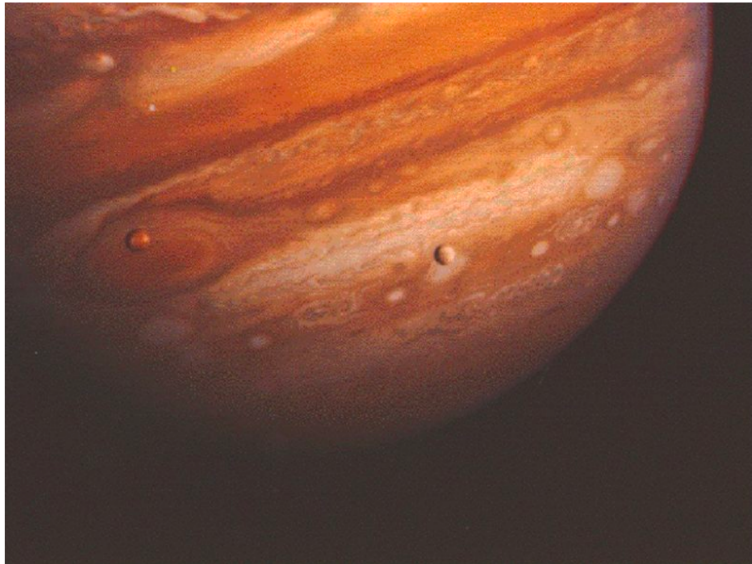
SPACE SCIENCE & SPACE PHYSICS

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

IS THERE A REPRODUCIBILITY CRISIS?



©nature

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

Reproducibility

[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

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

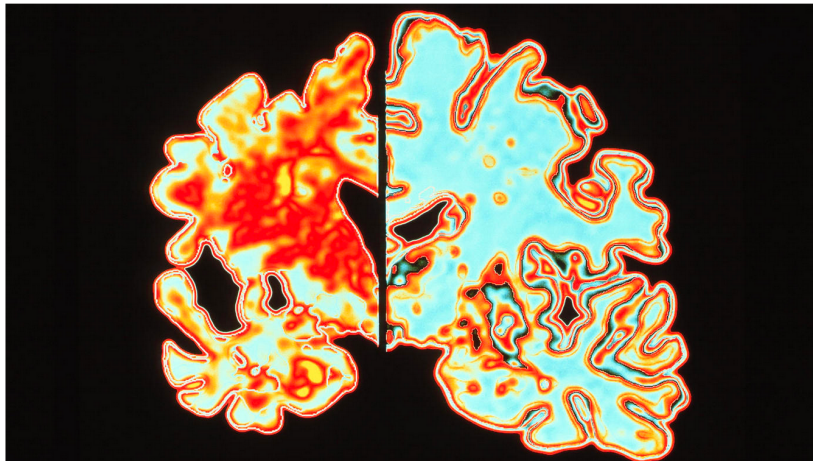


Control of Data



LOCAL / Education

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



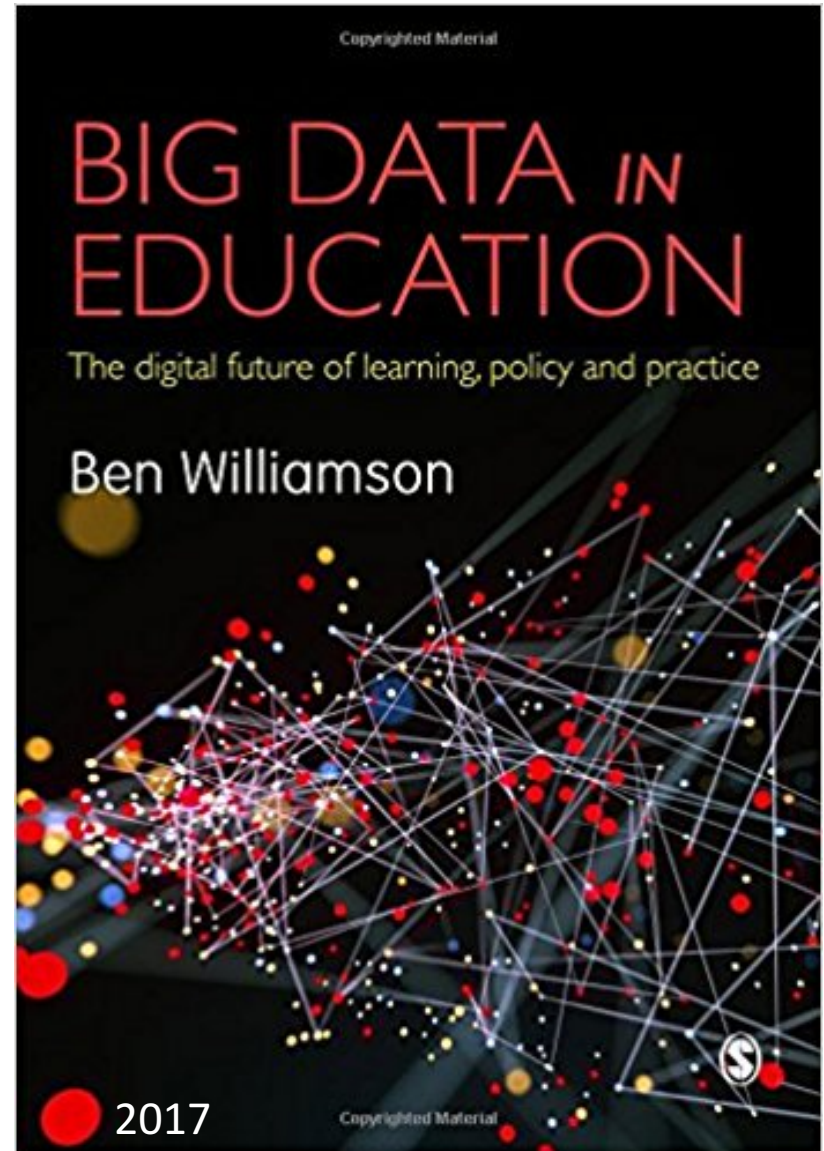
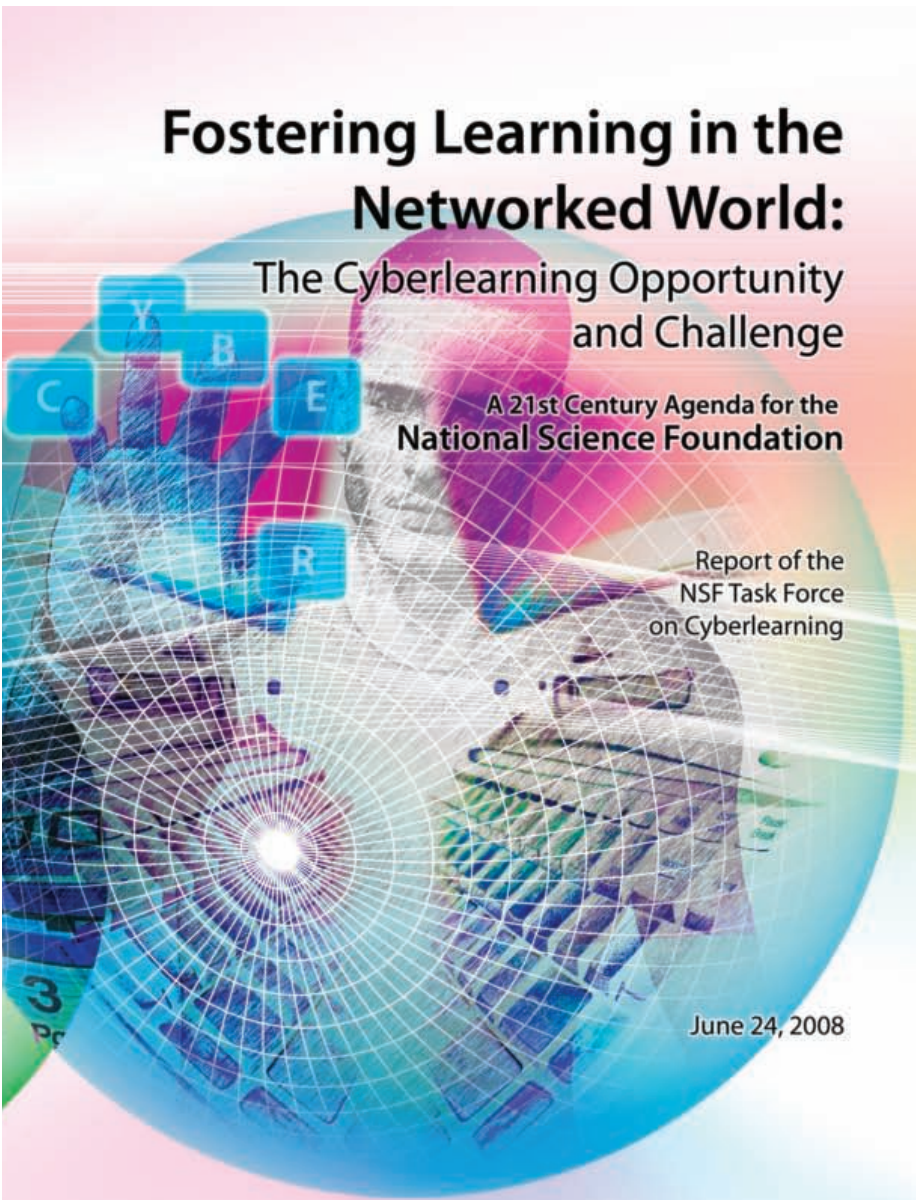
An ultrasound comparison of a brain of a patient with Alzheimer's disease, left, and a normal brain, right. (Pasioka / Getty Images)

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

Uses and Misuses of Data



Uses and Misuses of Data

Harvard secretly photographed students to study attendance

THE CHRONICLE OF HIGHER EDUCATION

[NEWS](#) [OPINION](#) [DATA](#) [ADVICE](#) [JOBS](#)

SECTIONS

FEATURED: [The Future of Work](#) [10 Teaching Innovators](#) [The Daily Briefing](#) [How to Be a Dean](#)

SPECIAL REPORTS

Where Every Student Is a Potential Data Point

Biosensors to monitor U.S. students' attentiveness

Stephanie Simon

7 MIN READ



DENVER (Reuters) - The Bill & Melinda Gates Foundation, which has poured more than \$4 billion into efforts to transform public education in the 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.



Menu

The CHRISTIAN SCIENCE
MONITOR

[WORLD](#) | [PASSCODE](#) | [PASSCODE VOICES](#) | [PRIVACY](#)

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

Uses and Misuses of Data



Freedom to Read Foundation



Free People Read Freely ®



Privacy

An Interpretation of the Library Bill of Rights

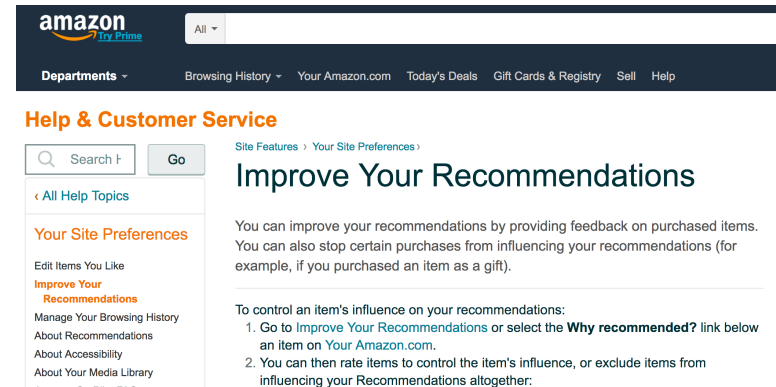
“In a library (physical or virtual), the right to privacy is the right to open inquiry without having the subject of one’s interest examined or scrutinized by others.”

Uses and Misuses of Data

Libraries: Right to read anonymously



Publishers: Tracking and recommendations



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

Clifford Lynch

Uses and Misuses of Data

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

41



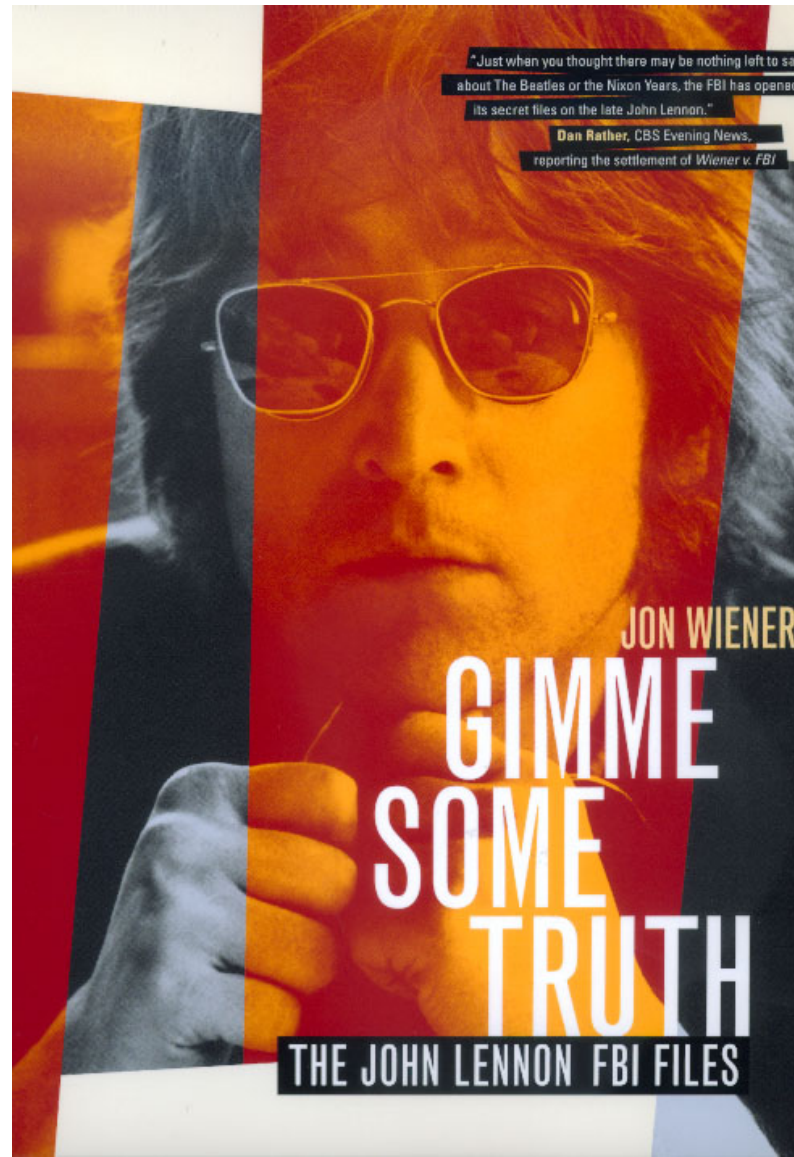
The Massachusetts Group Insurance Commission had a bright idea back in the mid-1990s—it decided to release "anonymized" data on state employees that showed every single hospital visit. The goal was to help researchers, and the state spent time removing all obvious identifiers such as name, address, and 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

Public Records Requests



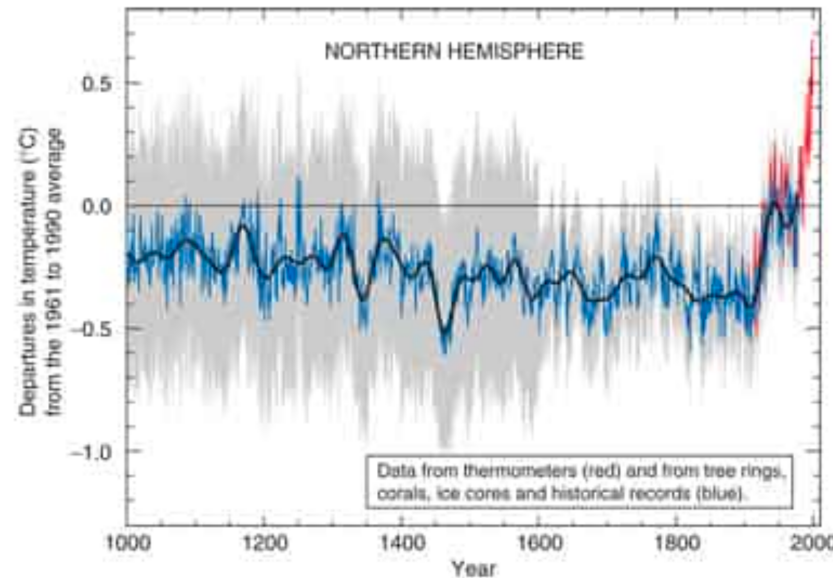
Public Records Requests



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



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

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

Public Records Requests



ACADEMIC FREEDOM

[APM](#)

[THE UCLA CALL](#)

[UCOP](#)

[FACULTY CODE OF CONDUCT](#)

[UNION CONTRACTS](#)

[VISITING SCHOLARS](#)

[SEXUAL HARASSMENT
PREVENTION/TITLE IX OFFICE](#)

[ACADEMIC FREEDOM](#)

[STAFF TRAINING](#)

[DEADLINES](#)

[FORMS](#)

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,

RELATED INFORMATION

[Faculty Guide to Public Records
Requests](#)

Cyber Risk and Data Breaches

[HOME](#)[LEARN](#)[BLOG](#)[DATA BREACHES](#)[POLICY & REPORTS](#)

[Home](#) / [Data Breaches](#)

DATA BREACHES

Breach Subtotal

Breach Type:	CARD, HACK, INSD, PHYS, PORT, STAT, DISC, UNKN
---------------------	--

Organization Type:	EDU
---------------------------	-----

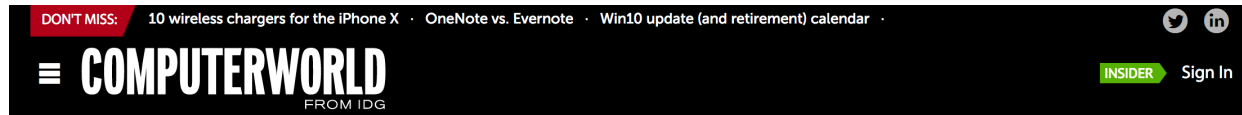
Year(s) of Breach:	2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005
---------------------------	--

Company or Organization:	all
---------------------------------	-----

Breaches made public fitting this criteria:	798
--	-----

Records total:	25,066,333
-----------------------	------------

Cyber Risk and Data Breaches



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 foothold on the company's payment systems.

HVAC systems connect to networks at various retail companies, government buildings and even hospitals, according to the security firm. HVAC vendors and other third parties often have remote access right to

MORE LIKE THIS

Target breach happened because of a basic network segmentation error

Target hackers try new ways to use stolen card data

Breach goes from bad to worse for Target and its customers



VIDEO

Tech Talk: The technology that scares us

Cyber Risk and Data Breaches



TIME

SUBSCRIBE

TECH • INTERNET

How Web Cams Helped Bring Down the Internet, Briefly



Getty Images

By **Haley Sweetland Edwards** October 25, 2016

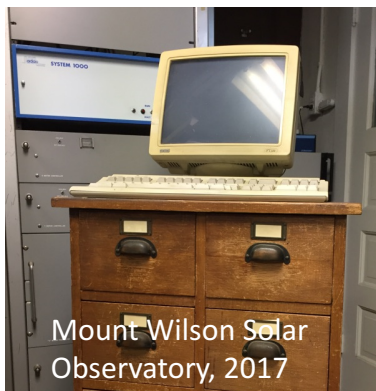
Data management: Private and secure



We just need to migrate the data from these systems to fit into that hole over there.



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



Graduate students



Post-doctoral fellows

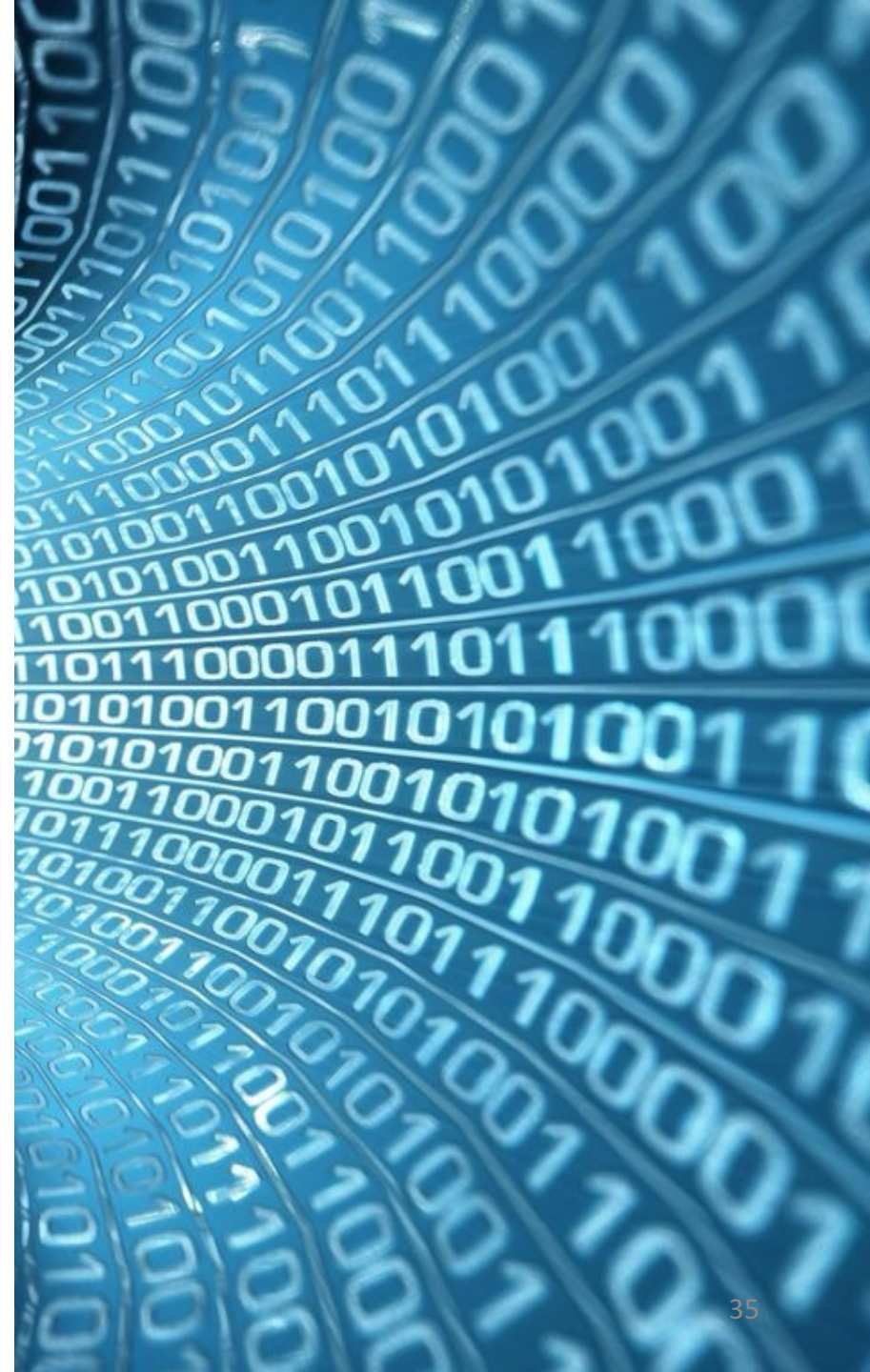
Data

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

(privacy and security aphorism)

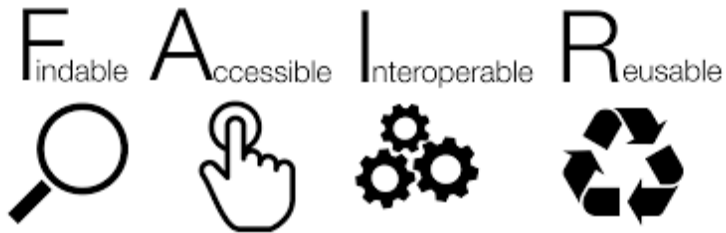
Therefore:

If you collect it, you
must protect it.

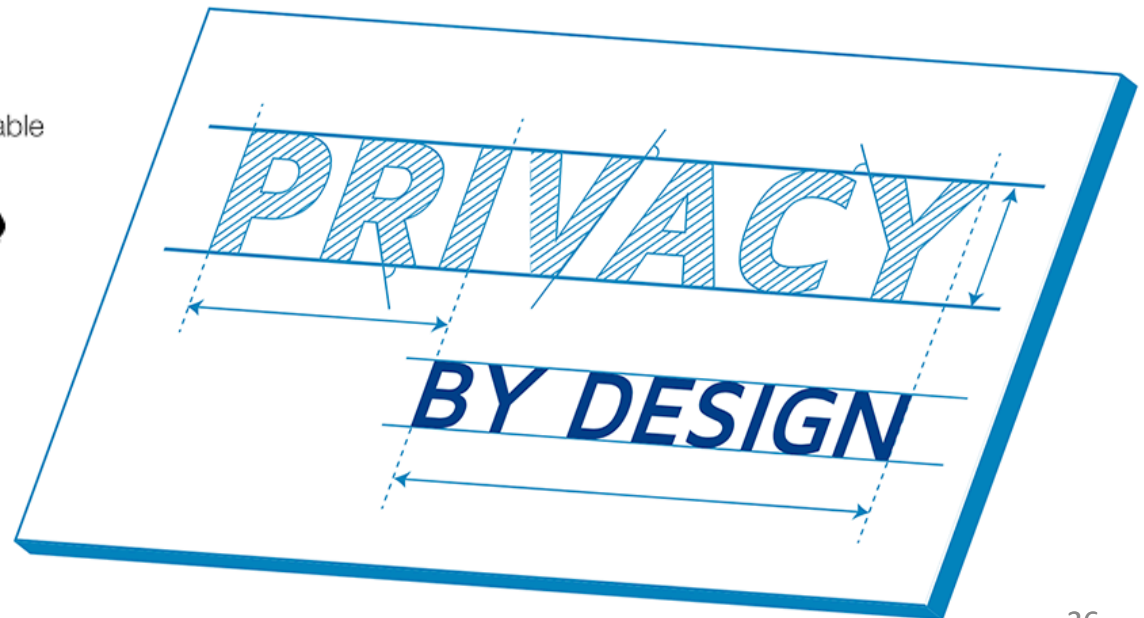


Protect Data and Privacy *open by design*

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



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/>

Protect Data and Privacy

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21st 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, Q3.js, Tableau



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

UCLA Corporate Financial Services

Search this site

BUSINESS & FINANCE SERVICES

CORPORATE ACCOUNTING

PAYROLL

TAX & RECORDS

TREASURY

/ 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 they are off-site or in your office.

LEGAL ISSUES

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

COPIES VS. ORIGINALS

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

ELECTRONIC FILES

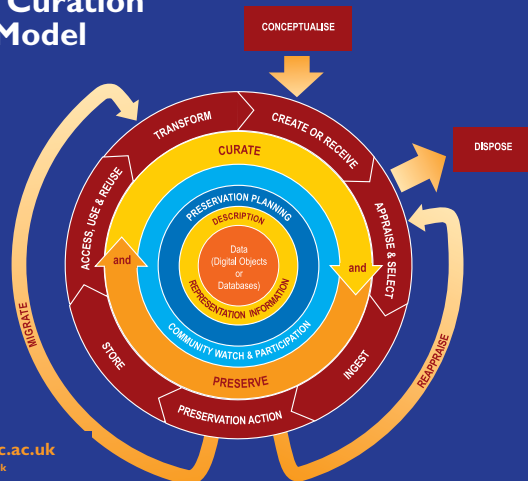
Retention does not apply only to paper records, but to electronic records too. This means it is necessary to erase certain computer files, including emails, over time, or they too will be discoverable.

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 to buy a cross-cut shredder.

The DCC Curation Lifecycle Model



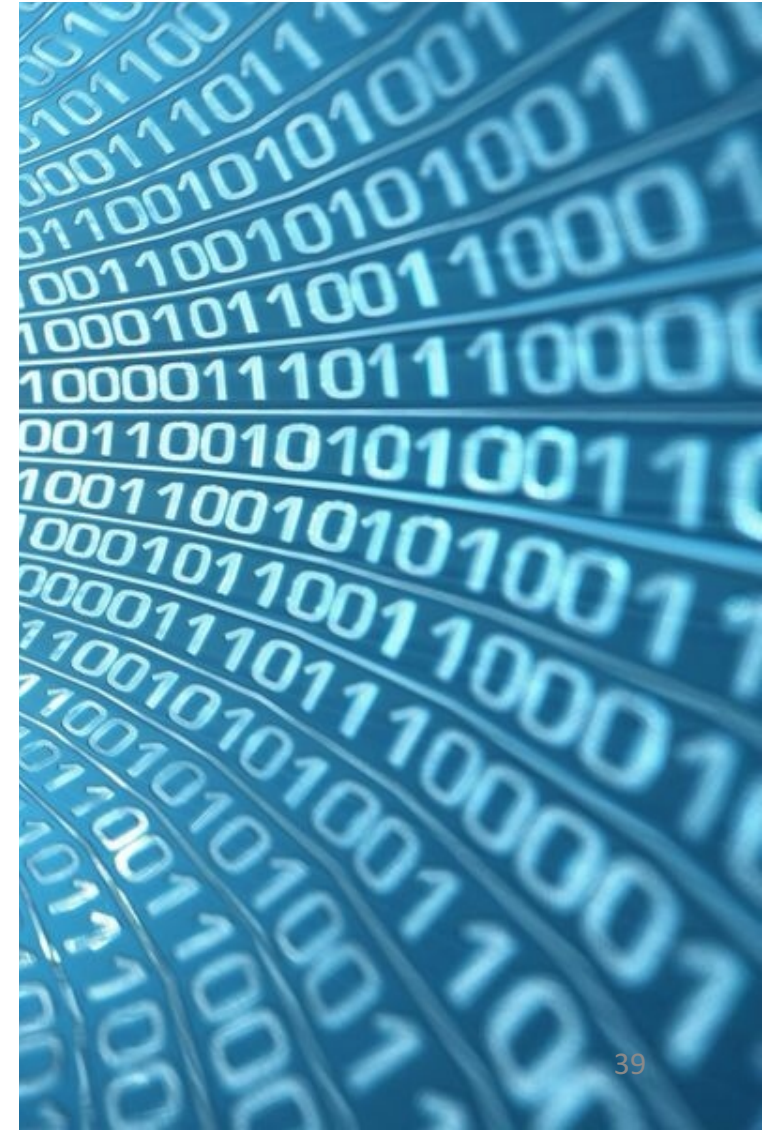
Promote Responsible Data Practices

- 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



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

- Exploit data for missions
 - Research
 - Teaching
 - Services
- Sustain trust of community
 - Privacy: information, autonomy
 - Academic freedom
 - Stewardship and governance
- Privacy Frontier: Open Data
 - Uses and misuses of data
 - Public records requests
 - Cyber risk and data breaches
 - Data management and infrastructure





UC Leadership in Data Policy

- We must maximally enable the **mission** of the University by supporting the values of **academic and intellectual freedom**.
- We must be **good stewards** of the **information entrusted** to the University.
- We must ensure that the University has **access to information** resources for **legitimate business purposes**.
- We must have a University community with **clear expectations of privacy**—both **privileges and obligations** of individuals and of the institution.
- We must make decisions within an **institutional context**.
- We must acknowledge the **distributed nature** of information stewardship at UC, where **responsibility for privacy and information security** resides at every level.

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
- Individuals: 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, Morgan Wofford, Sheryl Vacca, Gene Lucas, Tom Andriola...



Christine Borgman



Peter Darch



Irene Pasquetto



Bernie Boscoe



Michael Scroggins



Milena Golshan