

From Big Data to Big Brother

Surveillance and Privacy in the Information Age

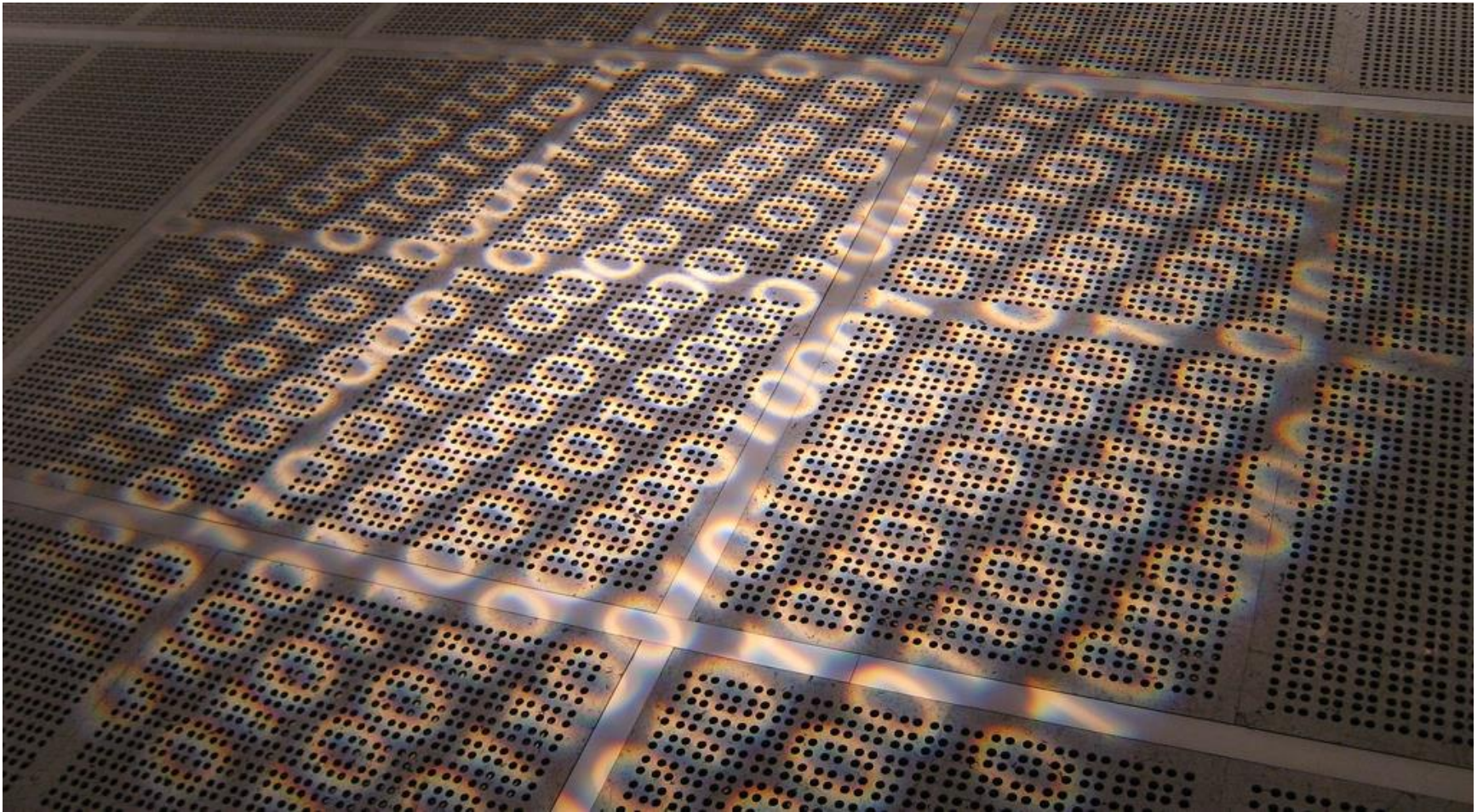


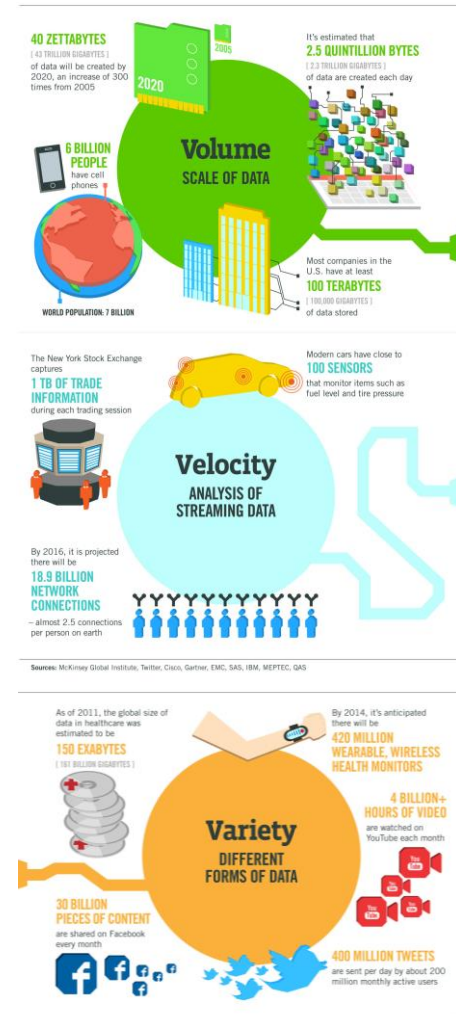
Image courtesy of [Josh Bancroft](#) on flickr. License CC-BY-NC.

What is Big Data?

Basic attributes (Kitchin, 2014)

- High-volume
- High-velocity
- High-variety
- Exhaustivity (n=all)
- Fine resolution
- Relationality
- Flexibility

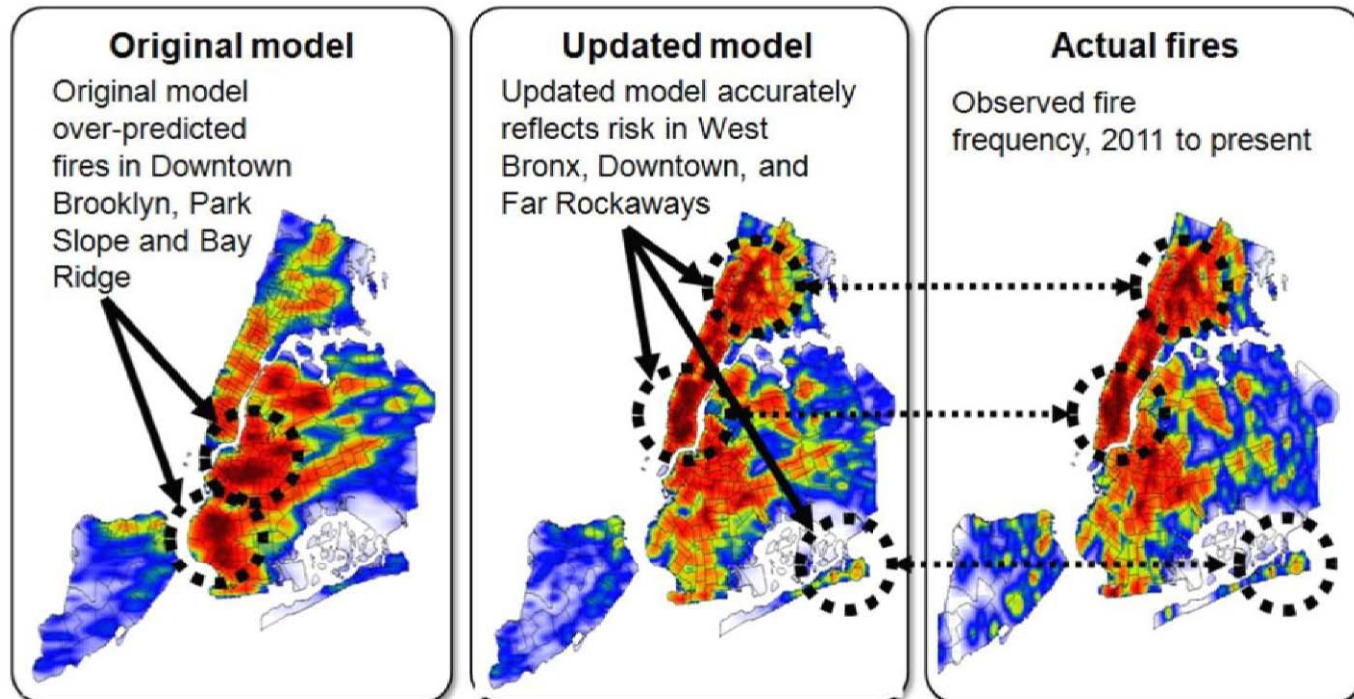
Ref. Bollier, 2010 and Kitchin, 2014



Security

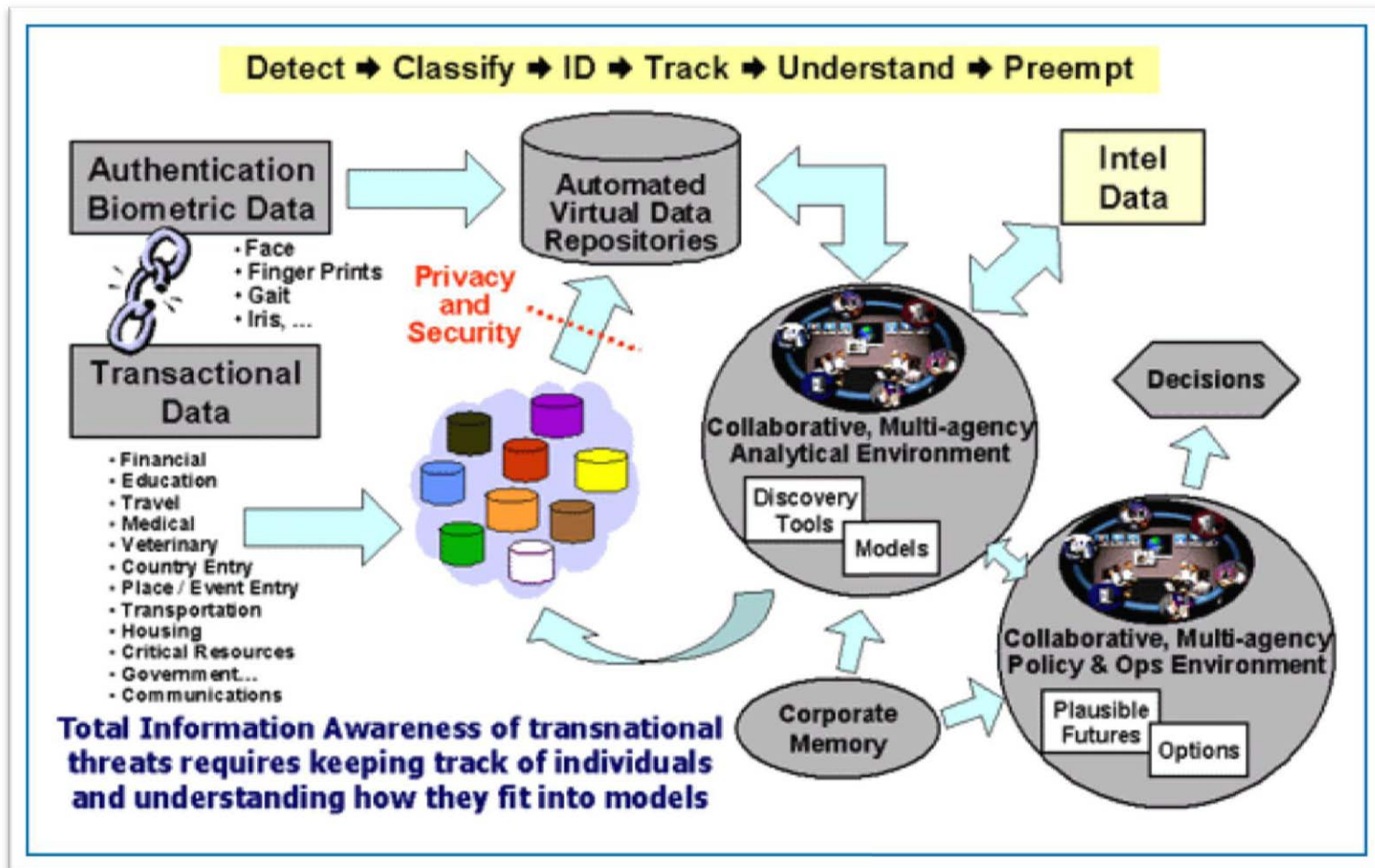
Using Predictive Analytics to Improve City Services

FDNY's Risk Based Inspection System (RBIS)



New York City

Security



Total Information Awareness program

This image is in the public domain.

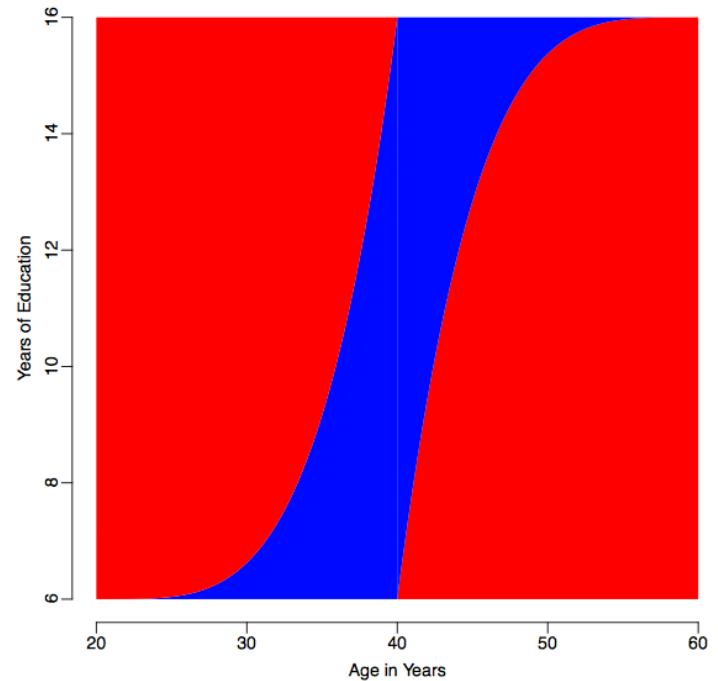
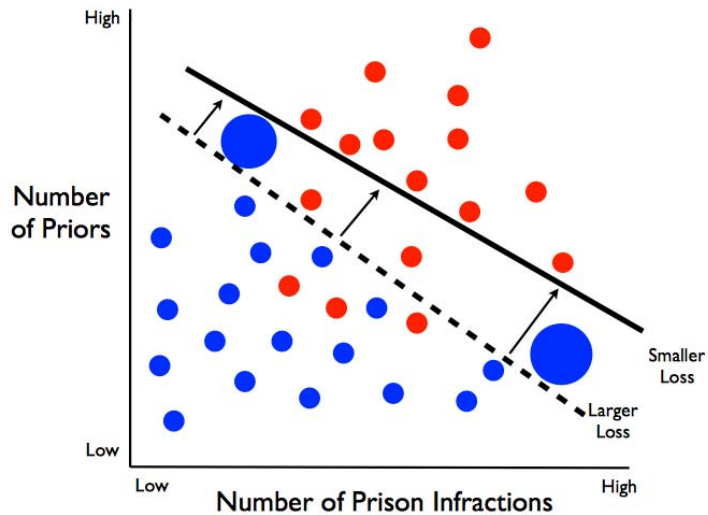
Security



Future Attribute Screening Technology

This image is in the public domain.

Security



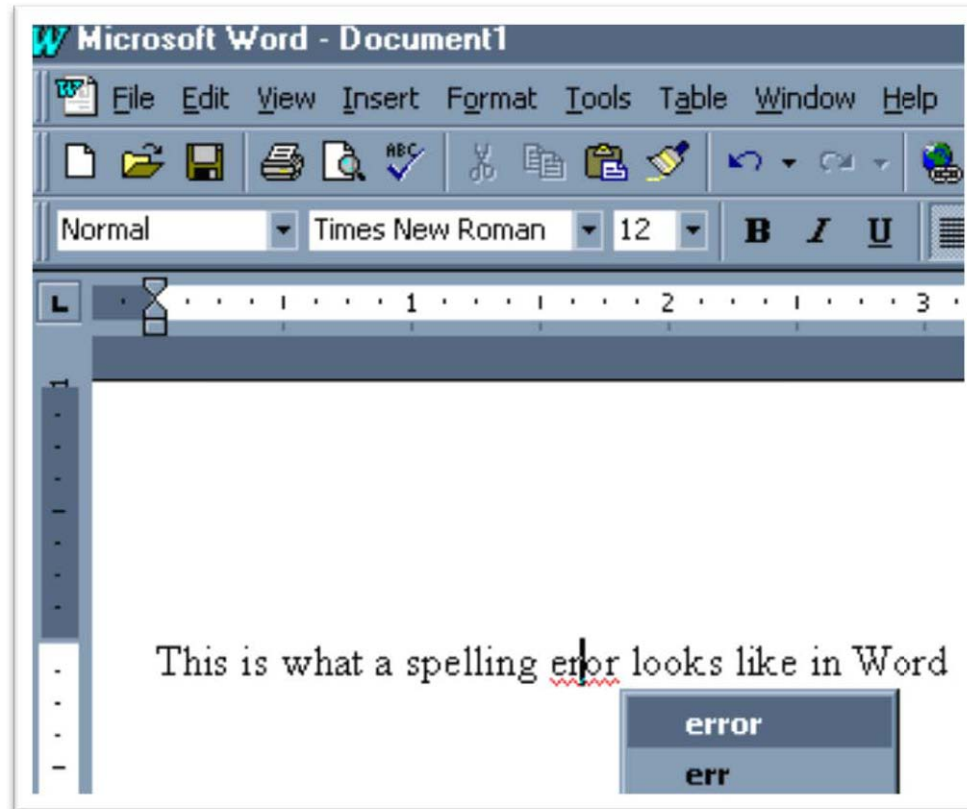
© John Wiley & Sons, Inc. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

Richard Berk's algorithm that predicts if a criminal released from jail will be involved in an homicide.

Services

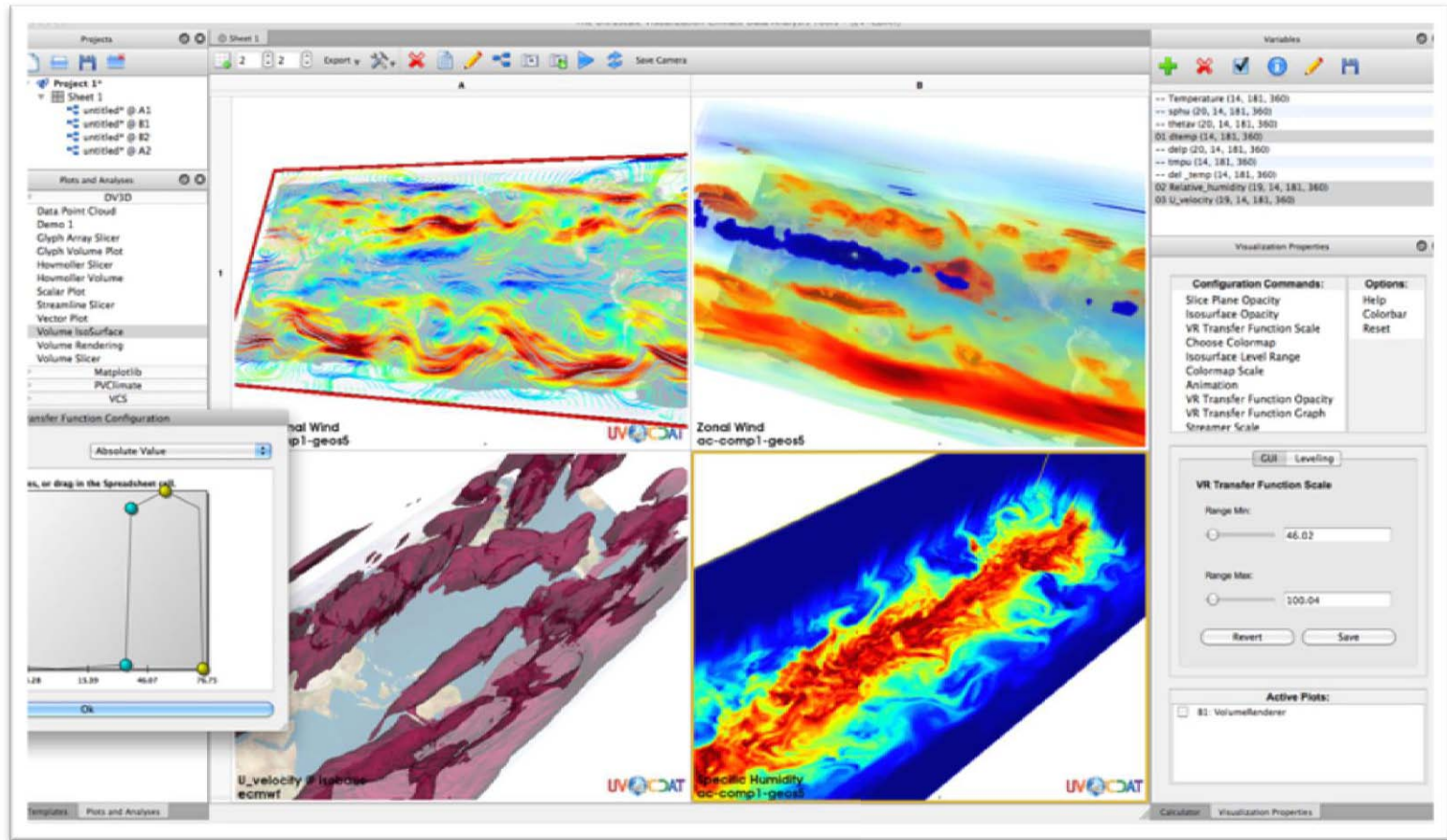
The screenshot shows the Google Translate web interface. At the top, there is a navigation bar with links for '+You', 'Search', 'Images', 'Maps', 'Play', 'YouTube', 'News', 'Gmail', 'Documents', 'Calendar', and 'More'. Below this is the Google logo and a 'Sign in' button. The main section is titled 'Translate' and features a 'From: French - detected' dropdown, a swap button, a 'To: English' dropdown, and a blue 'Translate' button. The input text is 'comme un éléphant dans un magasin de porcelaine' and the output is 'like a bull in a china shop'. Below the translation, there is a 'New!' notification: 'Click the words above to edit and view alternate translations. Dismiss'. At the bottom, there are links for 'Google Translate for Business: Translator Toolkit', 'Website Translator', and 'Global Market Finder'. The footer contains 'Turn off instant translation', 'About Google Translate', 'Mobile', 'Privacy', 'Help', and 'Send feedback'.

Services



© Microsoft. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

Research



© NASA. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

NASA Center for Climate Simulation

Ethical challenges

Potential consequences of Big data misuses and abuses (Bollier, 2010):

- imperil consumer freedom
- imperil civil security
- imperil personal privacy
- imperil of civil liberties

Dataveillance

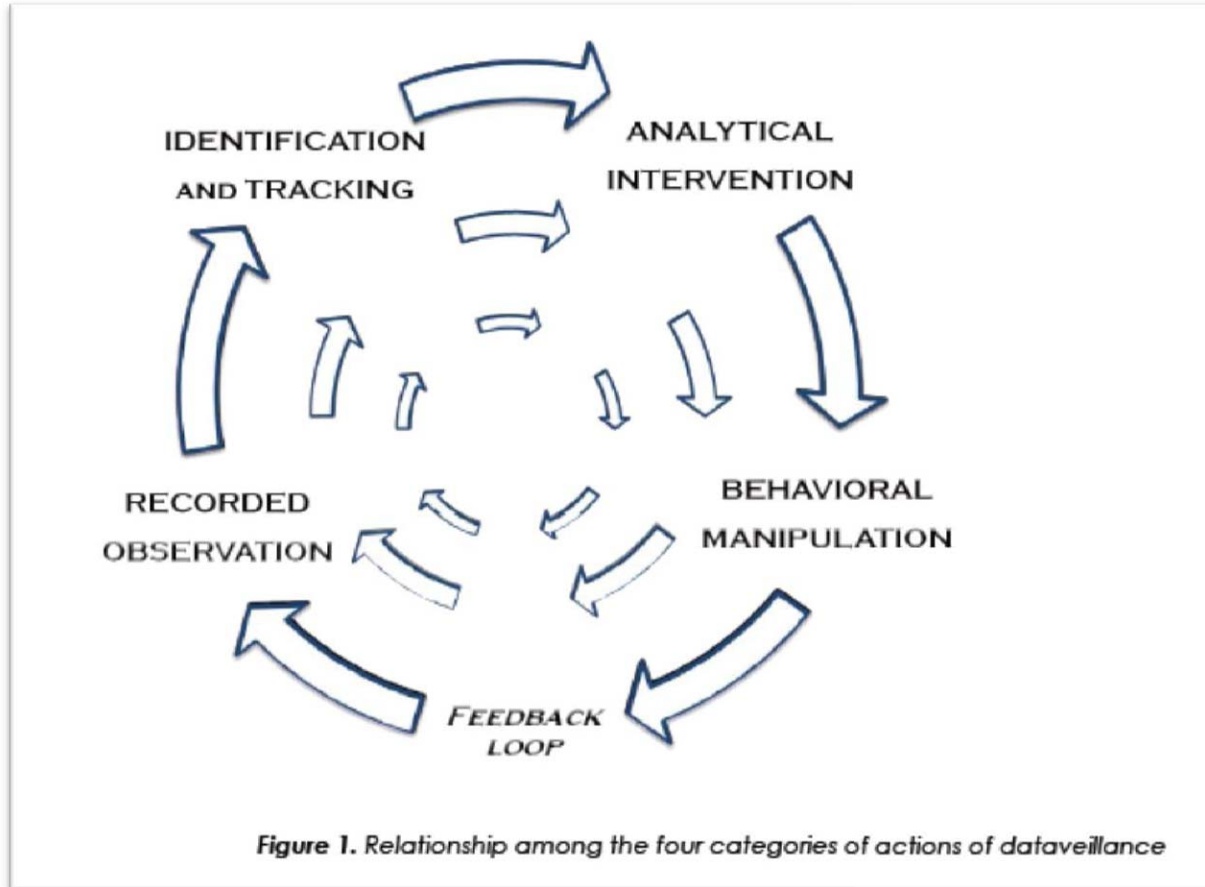
Definition: “systematic monitoring of people or groups, by means of personal data systems, in order to regulate or govern their behavior.”

(Degli Esposti, 2014)

Applications:

- Recruitment and retention
- Customer loyalty
- Supply chain efficiency
- Security and risk prevention

Dataveillance



© Surveillance & Society (Sara Degli Esposti). All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

(Ref. Degli Esposti, 2014)

Dataveillance

Main problems:

- unequal relationship between cie & customers
- happens without our knowledge (panopticon)

(Ref. Degli Esposti, 2014)

Civil liberties

People could eventually:

- pay more for health insurance
- be excluded from health coverage
- be denied a loan
- be subjected to profiling or extra scrutiny
- be kept in jail
- be arrested before committing a crime
- be found guilty by association

(Ref. Mayer-Schönberger and Cukier, 2014)

Civil liberties

Accusing a person of a possible behavior would:

- negate the very foundation of our justice (presumption of innocence and responsibility of people)
- deny human volition and free will
- exonerate us from any responsibility

(Ref. Mayer-Schönberger and Cukier, 2014)

Civil liberties

Should we accept to limit our individual freedom and rely on big data analytics for the common good?

Can big data analytics help us to avoid gender discrimination and racial profiling?

(Ref. Mayer-Schönberger and Cukier, 2014)

Privacy and civil liberties

Strategies currently in place:

- notice and individual consent
- opting-out
- anonymization
- hacking privacy violators

(Ref. Mayer-Schönberger and Cukier, 2014)

Privacy and civil liberties

Potential solutions:

- shifting from ‘privacy by consent’ to ‘privacy by accountability’
- defining new regulations (government or market?)
- redefining the notion of justice to preserve human agency
- opening the access to datasets and algorithms (transparency)
- mandating impartial experts to certify the algorithms and the interpretations
- blurring datasets (differential privacy)
- specifying how people can disprove a prediction about them

(Ref. Mayer-Schönberger and Cukier, 2014)

Privacy and civil liberties

Potential solutions (Bollier, 2010):

- erasing data after a given period of time
- giving to people the right to take possession of their data
- limiting data collection to what is needed
- favouring self-discipline by introducing new social norms
- gaming the system by voluntarily altering our behavior (Nissenbaum)
- being proactive

Methodological implications

Paradigm shift – Big data requires researchers to:

- deal with the abundance and embrace messiness
- conduct research with no precise question in mind
- gain insights from data (induction) instead of testing theories by analyzing data (deduction)
- use hundreds of algorithms instead of selecting one
- stop obsessing with causality and be satisfied with correlations

(Ref. Kitchin, 2014)

Methodological implications

Advantages of big data compared to traditional scientific methods:

- reduces the risk of error and bias associated with sampling (but can also face problems)
- provides the opportunity to create more dynamic and sophisticated models (but does not offer certainties)
- helps finding correlations that no theory can identify (but can lead to serious fallacies)
- allows to consider data once perceived as “noise”

(Ref. Kitchin, 2014)

Methodological implications

TWO SCHOOLS OF THOUGHT

Empiricist (business):

- Big data can speak for themselves without the need of theories, models or hypothesis (fallacious)
- Big data analytics are free of human bias. They can be interpreted by anyone and their meanings transcend contexts (fallacious)

Data-driven science (academia)

- Use of existing theories and concepts to analyze the datasets
- Use of big data to identify good questions and formulate hypothesis
- Use of datasets for answering questions that these datasets were not designed to answer (problem)

(Ref. Kitchin, 2014)

Methodological implications

Alternative path suggested by Kitchin (2014):

- Drawing on critical theory to frame how the research is conducted and how the results are interpreted.
- Admitting that research is never ideologically neutral and objective.
- Complementing big data studies with small data studies.

References

Bollier, David. 2010. *The Promise and Peril of Big Data*. Queenstown: The Aspen Institute.

Degli Esposti, Sara. 2014. “When Big Data Meets Dataveillance: The Hidden Side of Analytics”. *Surveillance & Society*, vol. 12, no 2, pp. 209-225.

Kitchin, Rob. 2014. “Big Data, New Epistemologies and Paradigms Shifts”. *Big Data & Society*, April-June, pp.1-12.

Mayer-Schönberger, Viktor and Kenneth Cukier. 2014. *Big Data*. New York: Mariner Books.

Additional Resources

- Examples of Big Data:
 - Public Health
 - [Flu trends from around the world](#)
 - [Patients Like Me](#)
 - [CareCloud](#)
 - [Mini-Sentinel](#)
 - [23 and me](#)
 - [RateMDs](#)
 - Security
 - [Centrifuge Systems](#)
 - [TransHeat Application](#)
 - [Sensors in car seats](#)
 - Finance & Business
 - [I Know First](#) - Daily Market Forecast
 - [FarmVille 2](#)
 - [The Numbers](#) - Where Data and Movie Business Meet

Additional Resources cont.

- Services
 - [Inrix](#)
 - [TripAdvisor](#)
 - [CitySense](#)
- Research
 - [Marine Explore](#)

MIT OpenCourseWare
<http://ocw.mit.edu>

CMS.701 / CMS.901 Current Debates in Media
Spring 2015

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.