

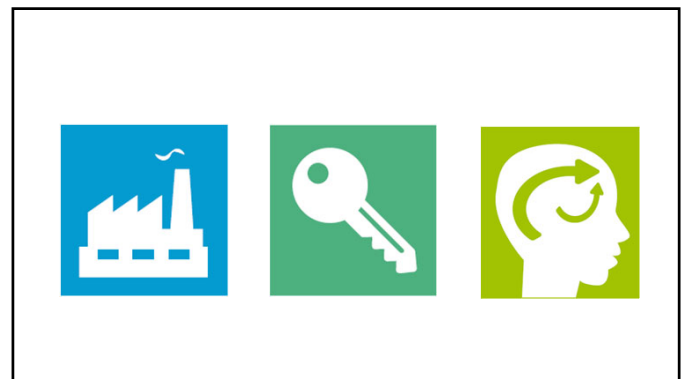


Registerdaten - Big Data - Forschungslücken

Frauke Kreuter
University of Maryland – IAB – Universität Mannheim
Coleridge Initiative – BERD@NFDI – digdeep.de

3.3.2020

1



2



Effizientere Erfassung

Neue Messung

Skalierung



Kombination ist "key"

Kombination ist ein Risiko

Zugang + Weiterbildung



Algorithmen Transparenz

Fehlerkultur

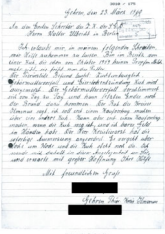
Intervention = Verantwortung

3

Datengenerierung

- Elektronische Erfassung von (administrativen) Prozessen erlaubt Reduktion von mühsamen biografischen Fragen

4

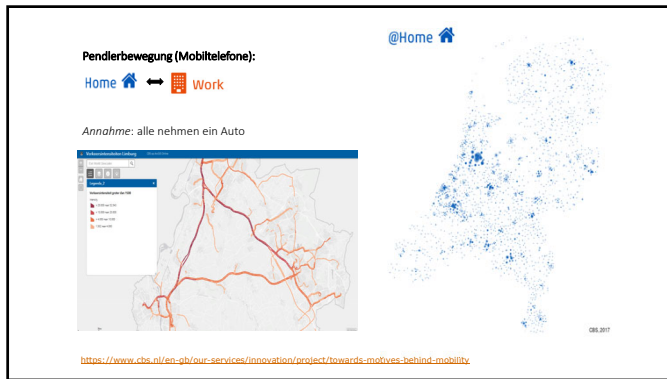

Kohler (2018) – Potsdam Grievance Statistics File (PGSF) DDR-Eingabestatistiken für den Zeitraum 1970-1989.

5

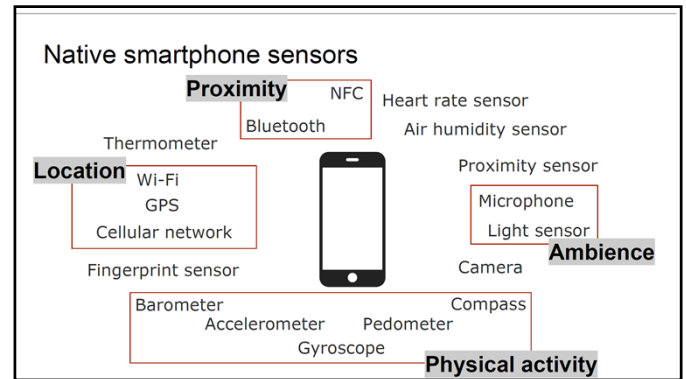


- Einsatz von Sensorik
Messung des bisher nicht Messbaren

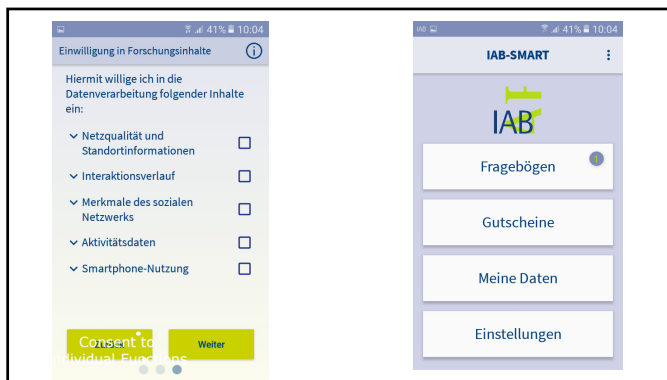
6



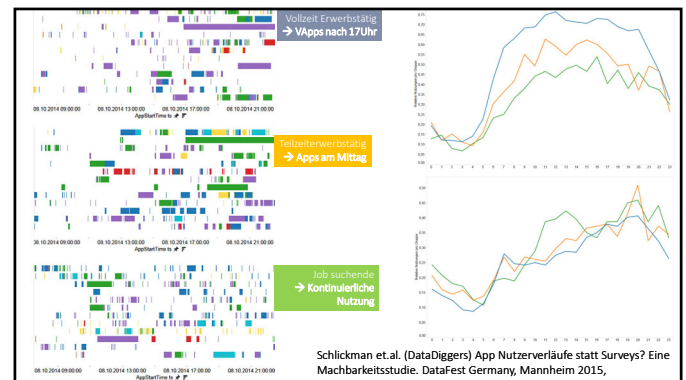
7



8



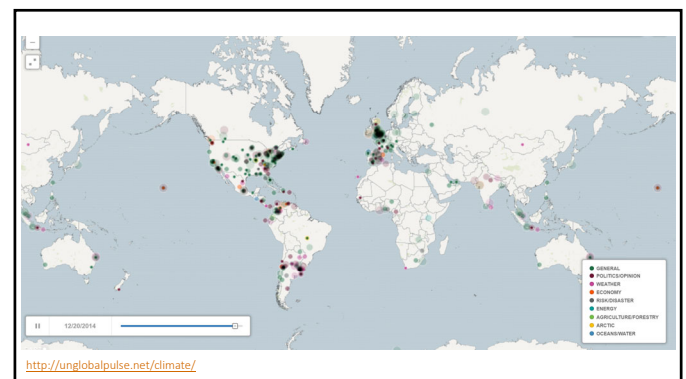
9



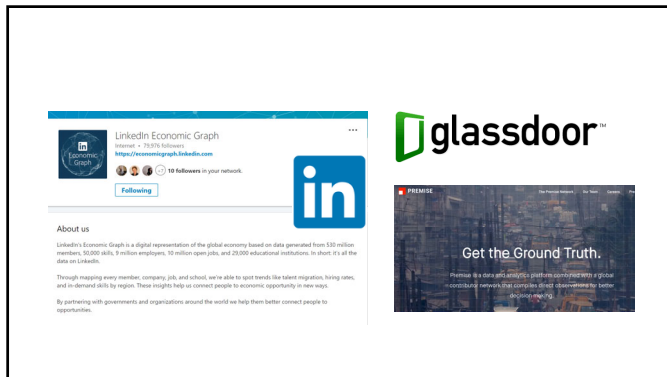
10



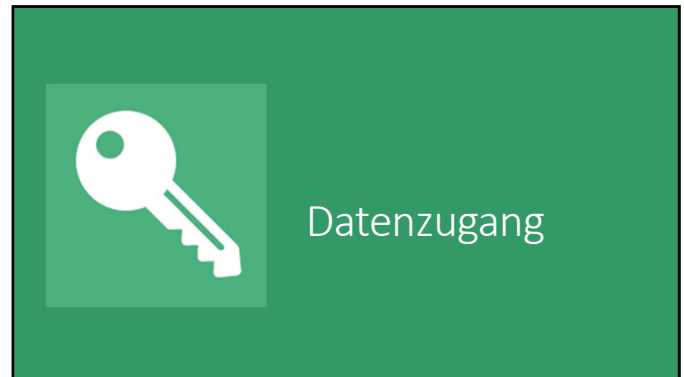
11



12



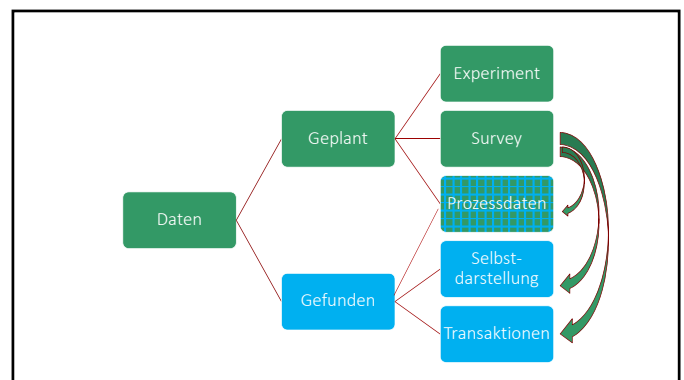
13



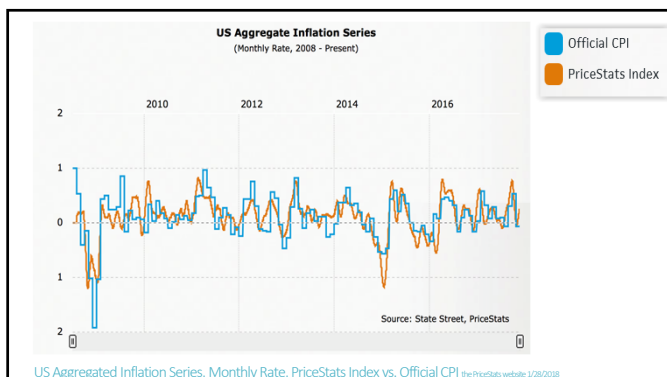
14



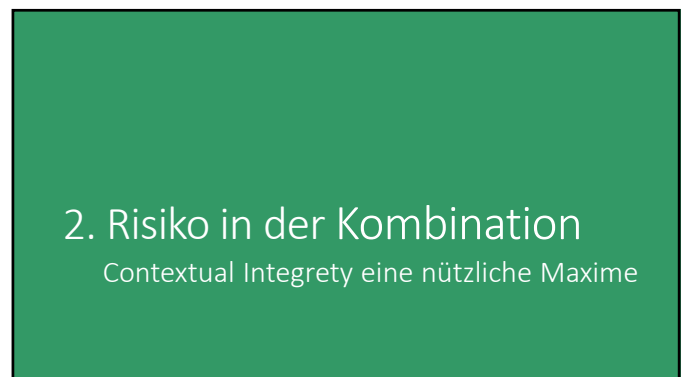
15



16



17



18

A Face Is Exposed for AOL Searcher No. 4417749
By MICHAEL BARBARO and TOM ZELLER JR.
PUBLISHED: AUGUST 15, 2018

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "mush fingers" to "Go single now" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lithium, Ga.," several people with the last name Arnold and "homes sold in shadow lake subdivisions gwinnett county georgia."

It did not take much investigating to follow that data trail to Thomas Arnold, a 64-year-old widow who lives in Lithium, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.

19

H1B Salary Database Companies Job Titles Cities Highest Paid Refinance Student Loan & Save! Subscribe

EMPLOYER	JOB TITLE	BASE SALARY	LOCATION	SUBMIT DATE	START DATE
FEDERAL RESERVE BANK OF SAN FRANCISCO	APPLICATION DEVELOPER IV	160,000	SAN FRANCISCO, CA	05/27/2016	07/05/2016
FEDERAL RESERVE BANK OF SAN FRANCISCO	APPLICATION DEVELOPER IV	160,000	SAN FRANCISCO, CA	06/08/2016	07/18/2016
FEDERAL RESERVE BANK OF SAN FRANCISCO	APPLICATION DEVELOPER IV	162,500	SAN FRANCISCO, CA	05/27/2016	07/05/2016
FEDERAL RESERVE BANK OF SAN FRANCISCO	APPLICATIONS INTEGRATION ENGINEER	123,000	SAN FRANCISCO, CA	12/21/2017	03/15/2018
FEDERAL RESERVE BANK OF SAN FRANCISCO	AUTOMATION TEST & INTEGRATION ENGINEER	110,000	RICHMOND, VA	12/13/2017	12/13/2017
FEDERAL RESERVE BANK OF SAN FRANCISCO	AUTOMATION TEST & INTEGRATION ENGINEER	110,000	RICHMOND, VA	12/13/2017	12/13/2017
FEDERAL RESERVE BANK OF SAN FRANCISCO	DOCUMENT MANAGEMENT SENIOR ANALYST	110,860	SAN FRANCISCO, CA	02/07/2018	03/01/2018
FEDERAL RESERVE BANK OF SAN FRANCISCO	DOCUMENT MANAGEMENT SENIOR ANALYST	110,860	SAN FRANCISCO, CA	03/02/2018	07/13/2018
FEDERAL RESERVE BANK OF CLEVELAND	ECONOMIC ANALYST INTERN	56,000	CLEVELAND, OH	09/04/2015	10/01/2015
FEDERAL RESERVE BANK OF SAN FRANCISCO	ECONOMIC RESEARCH ADVISOR	205,000	SAN FRANCISCO, CA	04/21/2016	08/15/2016

20

Protecting the Confidentiality of America's Statistics: Adopting Modern Disclosure Avoidance Methods at the Census Bureau

PR AUG 17 2018
WRITTEN BY: DR. JOHN M. ABOU, CHIEF SCIENTIST AND ASSOCIATE DIRECTOR FOR RESEARCH AND METHODOLOGY

SUBSCRIBE RSS DIAL EMAIL

The U.S. Census Bureau's commitment to data stewardship—protecting respondent privacy and confidentiality at every stage of the data lifecycle—is grounded in law that is straightforward, robust, and strong. From the time we collect the data, through processing, publication and storage, we are bound by Title 13 of the United States Code to ensure that information about any specific individual, household, or business is never revealed, even indirectly through our published statistics.

21

Back

We value your privacy

You can set consent preferences for individual third-party partners we work with below. Expand each company list item to see what purposes they use data for to help make your choices. In some cases, companies may use your data without asking for your consent, based on their legitimate interests. You can click on their privacy policy links for more information and to object to such processing.

Consent uses your "I Vendors" your

Analytics Inc.	☐	☐
admedia	☐	☐
Adacore Technologies Inc. (DBA Adacore)	☐	☐
adality GmbH	☐	☐
ADARBA MEDIA UNLIMITED	☐	☐
AdClear GmbH	☐	☐
AdCompany, INC.	☐	☐
AdMayer GmbH	☐	☐

SAVE & EXIT

Helen Nissenbaums Forderung nach "Contextual Integrity" ist eher nicht erfüllt ...
.. wenn Research Gate Daten mit über 2000 (Werbe-)Partner teilt.

22

Contextual Integrity

PRIVACY IN CONTEXT
Technology, Policy, and the Integrity of Social Life
HELEN NISSENBAUM

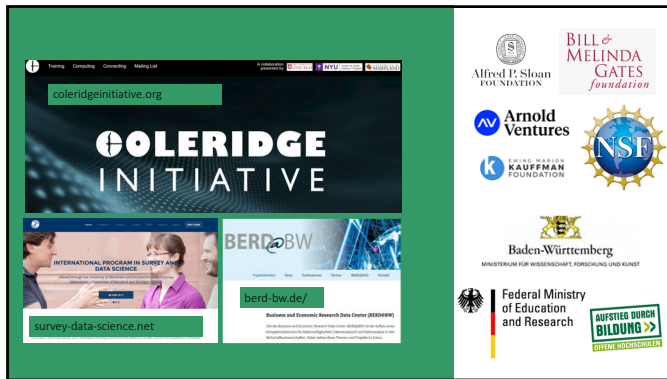
Privacy, Big Data, and the Public Good
Frameworks for Engagement
edited by Julia Lane, Victoria Robinson, Brian Brainer, Helen Nissenbaum

DARK DATA
BY DANIEL MITCHELL
DAVID J. HAND

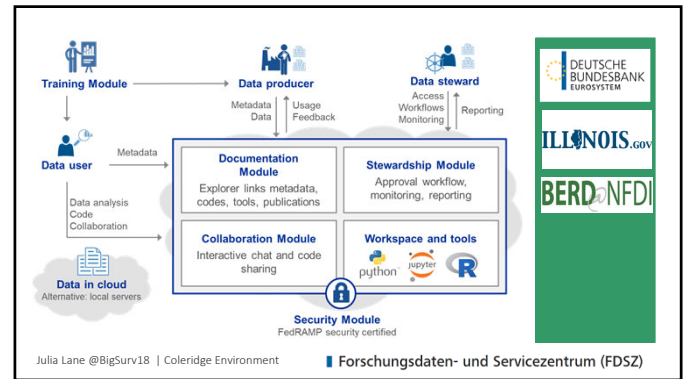
23

3. Zugang + Weiterbildung
muß kombiniert werden

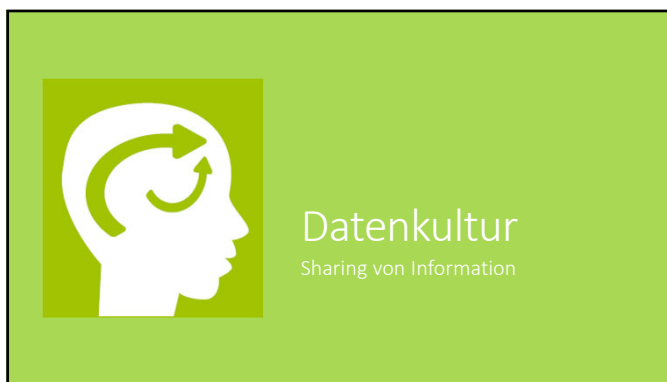
24



25



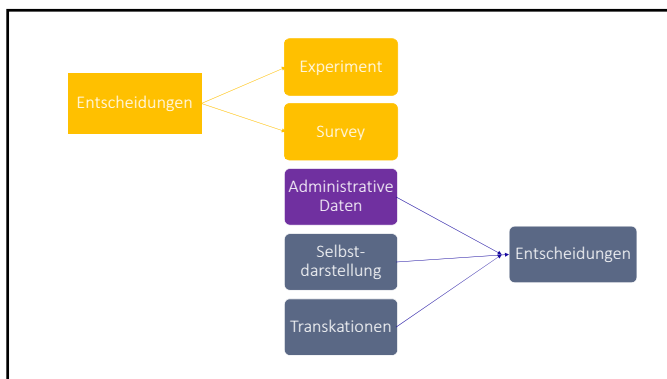
26



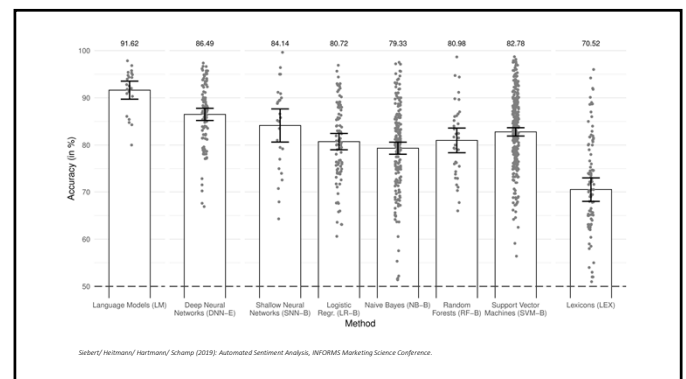
27



28



29



30

2. Fehler sind unabdingbar

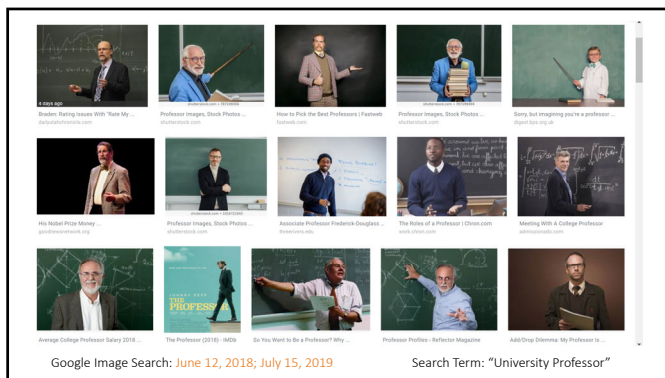
Wir brauchen eine Fehlerkultur als Teil der neuen Datenkultur

31

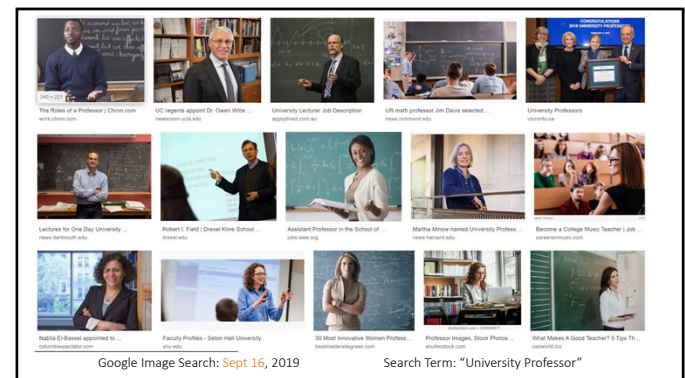
By far the worst approach is to wait for data quality problems to surface on their own.

T. Herzog, F. Scheuren, W. Winkler, 2007

32



33



34

Google drops gender labels from image recognition to reduce bias

Every individual will now be classified as a "person" in Google Cloud's Vision API



Most popular



AI Now: Predictive policing systems are racist because corrupt cops produce dirty data

35

3. Evidenzbasierte Politik

Ist gepaart mit Bedürfnis nach Effizienz. Prediktion bringt neue Ansprüche an Datenqualität und Modellgüte

36

BIG DATA & THE SDGs

How data science and analytics can contribute to sustainable development

- 1. NO POVERTY**: Spending patterns on mobile phone services can provide proxy indicators of income levels.
- 2. ZERO HUNGER**: Crowdsourcing or tracking of food prices listed online can help monitor food security in near real-time.
- 3. GOOD HEALTH AND WELL-BEING**: Mapping the movement of mobile phone users can help predict the spread of infectious diseases.
- 4. QUALITY EDUCATION**: Citizens reporting can reveal reasons for school drop-out rates.
- 5. GENDER EQUALITY**: Analysis of financial transactions can reveal the spending patterns and different impacts of economic shocks on men and women.
- 6. CLEAN WATER AND SANITATION**: Sensors connected to water pumps can track access to clean water.
- 7. AFFORDABLE AND CLEAN ENERGY**: Smart metering allows utility companies to increase or restrict the flow of electricity, gas or water to reduce waste and ensure adequate supply at peak periods.
- 8. DECENT WORK AND ECONOMIC GROWTH**: Patterns in global postal traffic can provide indicators such as economic growth, remittances, trade and GDP.
- 9. INDUSTRY, INNOVATION AND INFRASTRUCTURE**: Data from GPS devices can be used for traffic control and to improve public transport.
- 10. REDUCED INEQUALITY**: Speech-to-text analytics on local radio content can reveal discrimination concerns and support policy response.
- 11. SUSTAINABLE CITIES AND COMMUNITIES**: Satellite remote sensing can track encroachment on public land or spaces such as parks and forests.
- 12. RESPONSIBLE CONSUMPTION AND PRODUCTION**: Online search patterns or e-commerce transactions can reveal the pace of transition to energy efficient products.
- 13. CLIMATE ACTION**: Combining satellite imagery, crowd-sourced witness accounts and open data can help track deforestation.
- 14. LIFE BELOW WATER**: Maritime vessel tracking data can reveal illegal, unregulated and unreported fishing activities.
- 15. LIFE ON LAND**: Social media monitoring can support disaster management with real-time information on victim location, effects and strength of forest fires or haze.
- 16. PEACE, JUSTICE AND STRONG INSTITUTIONS**: Sentiment analysis of social media can reveal public opinion on effective governance, public service delivery or human rights.
- 17. PARTNERSHIPS FOR THE GOALS**: Partnerships to enable the combining of statistics, mobile and internet data can provide a better and real-time understanding of today's hyper-connected world.

GLOBAL RISE

37

Copyrighted Material

michael kearns + aaron roth

al+algorithm
the ethical+
algorithm
the et

the science of
socially aware algorithm design

Copyrighted Material

<https://textbook.coderidgeinitiative.org/chap-bias.html>

- Fairness
- Accuracy
- Transparency
- Ethics

38

Collaboration

Scientists
Engineers
Regulators
Philosophers
Social Workers
Concerned Citizens

<https://icon-library.net/icon/collaboration-icon-12.html>

39

Interaktion: Daten – Nutzer - Algorithmen

40

Effizientere Erfassung

Neue Messung

Skalierung

Kombination ist "key"

Contextual Integrity

Zugang + Weiterbildung

Algorithmen Transparenz

Fehlerkultur

Intervention = Verantwortung

41

Vielen Dank

Frauke Kreuter
University of Maryland – IAB – Universität Mannheim
digdeep.de

3.3.2020
@fraukolos

42