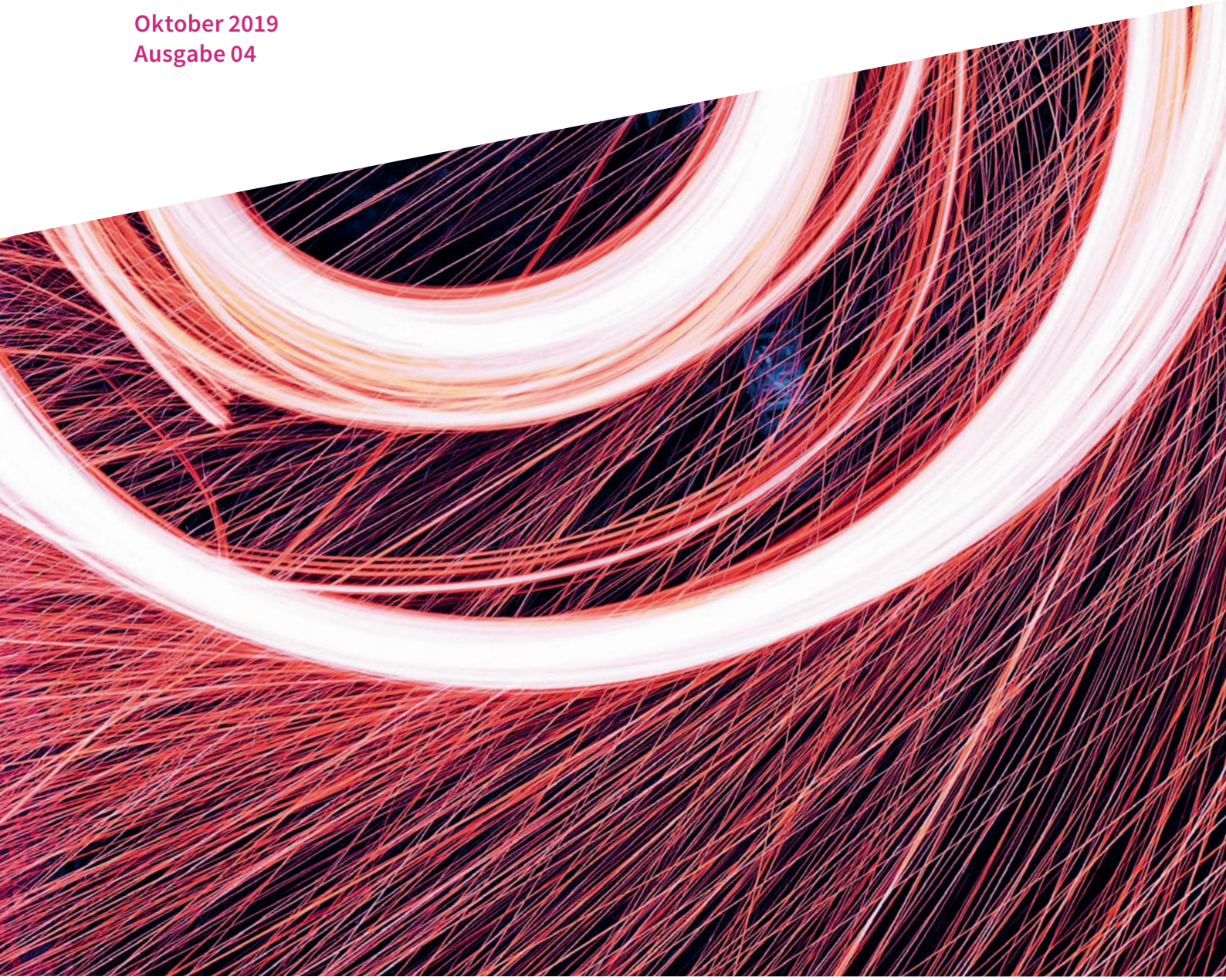
iRights.Lab
Think Tank für die digitale Welt

Data-Governance-Report

mFUND-Begleitforschung vom Think Tank iRights.Lab

Oktober 2019
Ausgabe 04



Bundesministerium für Verkehr und digitale Infrastruktur

mFUND
Das Startkapital für die Mobilität 4.0

Data-Governance-Report

Es handelt sich hierbei um den vierten Data-Governance-Report, der im Rahmen der mFUND-Begleitforschung vom iRights.Lab für alle geförderten Projekte zur Verfügung gestellt wird.

Der vorliegende Report hat zwei zentrale Themen: Zum einen schließt er an den Report vom Mai 2019 an, indem er das Thema Datenqualität erneut aufgreift. Es wird – wie im letzten Report angekündigt – eine hilfreiche Checkliste als Handreichung für die mFUND-Projekte veröffentlicht. Zum anderen beschäftigt er sich inhaltlich mit der ethisch-gesellschaftlichen Ausstrahlungskraft von Data-Governance. Mithilfe der Lektüre, sowie den praktischen Modellen und Checklisten können Projekte und Organisationen eine Sensibilisierung in Bezug auf ethisch-gesellschaftliche Aspekte implementieren.

Darüber hinaus informieren wir immer über Neuigkeiten aus unserer Begleitforschung und geben einen Ausblick auf die künftigen Entwicklungen in unserem Projekt.

DATENQUALITÄT

ETHISCH-GESELLSCHAFTLICHE AUSWIRKUNGEN VON DATA-GOVERNANCE

NEWS

Woran arbeitet das iRights.Lab im Projekt „Data-Governance“ der mFUND-Begleitforschung?

Im Projekt „Data-Governance im Innovationsprozess“ entwickeln wir ein Self-Governance-Konzept im Bereich datenbasierter Innovationen. Auf dem Weg dahin berücksichtigen wir rechtliche, ethische, technologische, gesellschaftliche und politische Parameter, sodass passende Lösungen für verschiedene Organisationstypen entworfen werden. Wir wollen Behörden und Unternehmen dazu befähigen, eigene Regeln und Strukturen für einen verantwortungsbewussten Umgang mit den von ihnen verwalteten und verwendeten Daten aufzusetzen. Dabei sollen die Interessen von Unternehmen, Behörden und Gesellschaft im Bereich Datenschutz und Datensicherheit in Einklang gebracht werden, um Innovationen zu fördern.

Neuigkeiten aus der mFUND-Begleitforschung des iRights.Lab

Die mFUND-Konferenz

Am 26. und 27. September 2019 fand die dritte mFUND-Konferenz im Allianz Forum in Berlin statt. Auf der Bühne führte Dr. Wiebke Glässer vom iRights.Lab die fast 400 angemeldeten Personen durch ein kurzweiliges Programm. Zudem war das iRights.Lab gleich an beiden Tagen mit Meetups in Form von Workshops vertreten. Das Thema des letzten Data-Governance-Reports wurde in dem Workshop „Datenqualität in der Praxis“ moderiert von Michael Puntschuh vom iRights.Lab und Lina Bruns vom Projekt NQDM (Normenentwurf Qualitativ hochwertiger Daten und Metadaten) bei Fraunhofer Fokus bearbeitet. Michael Puntschuh und Levke Burfeind, ebenfalls vom iRights.Lab, leiteten einen weiteren Workshop. Passend zum vorliegenden Report wurden mit den Teilnehmenden Fragen zur ethischen und gesellschaftlichen Relevanz von Mobilitätsprojekten anhand von konkreten Handreichungen diskutiert.

mFUND-Konferenz 2019, Fotos: © Gerhard Kassner

Das Local Meetup zur Datenqualität

Im letzten Report wurden neben einem Interview mit Marco Geuer das Thema Datenqualität bereits vertieft. Auf dieser Basis und anhand ergänzender Recherche und Analysen entwickelte das iRights.Lab eine entsprechende Empfehlung im Sinne einer Checkliste speziell für die im mFUND geförderten Projekte. Dabei sind die Erkenntnisse der bisherigen Begleitforschung mit eingeflossen.

Auf dem Local Meetup im Mai 2019 in Mannheim präsentierten wir diese Checkliste für Datenqualität. Neben unseren Arbeitsergebnissen stellte Lina Bruns vom Fraunhofer Fokus ihre Arbeit zu Empfehlungen und Standards für Datenqualität vor. Nach dem Austausch mit den Expert*innen konnten die Teilnehmenden die Checkliste erproben und evaluieren. Abschließend gab es die Möglichkeit, sich in lockerer Atmosphär auszutauschen und Erfahrungen zu teilen.

Wir haben uns über die Mitarbeit und den Input aus den vertretenden mFUND-Projekten auf dem Local Meetup sehr gefreut und nutzen die Gelegenheit, Ihnen die Liste auf diesem Weg zur Verfügung zu stellen.

Local Meetup Mai 2019
Foto: Michael Puntschuh

Datenqualität-Checkliste

Wir veröffentlichen diese Checkliste nun für alle auch im Data-Governance-Report. Sie besteht aus einer Übersicht von 16 Dimensionen, von Zugänglichkeit über Aktualität bis hin zu Verständlichkeit. Die unter CC-Lizenz veröffentlichte Checkliste soll nun auch Projekte des mFUND bei ihrer Data-Governance unterstützen. Wichtig ist zu betonen, dass es sich nicht um projektspezifische Handlungsempfehlungen handelt. Ihr Wert liegt darin, einen Überblick über relevante Punkte zu vermitteln. Sie ist dazu geeignet, ein umfassendes Bild der verschiedenen Aspekten von Datenqualität zu liefern. Aber sie muss gleichzeitig immer aus der Perspektive des jeweiligen Projektes gelesen und umgesetzt werden. Diese Umsetzung gilt es zudem regelmäßig zu überprüfen und ggf. zu aktualisieren.

Checkliste für eine bessere Datenqualität

WAS KÖNNEN SIE MIT DER CHECKLISTE ERREICHEN?

- Häufige Datenqualitätsprobleme vorausschauend erkennen
- Anregung für eine Datenqualitätsstrategie erhalten
- Strategisches Vorgehen bei Datenqualitätsproblemen als Teil einer Data-Governance implementieren

16 DIMENSIONEN ZUR MESSUNG VON DATEN- BZW. INFORMATIONSQUALITÄT

Diese 16 Dimensionen können Sie dabei unterstützen, Datenqualität systematisch zu beurteilen.

1 ZUGÄNGLICHKEIT (ACCESSIBILITY)
Kann ich die Daten auf einfache Weise abrufen?

2 BEARBEITBARKEIT (EASE OF MANIPULATION)
Ist es möglich, dass ich die Daten ändern/ transformieren kann?

SYSTEM

3 GLAUBWÜRDIGKEIT (BELIEVABILITY)
Sind den Daten weitere Informationen (z.B. Zertifikate oder ausführliche Berichte über Datengewinnung und -aufbereitung) beigefügt, die Vertrauen in die Datenqualität erwecken?

4 HOHES ANSEHEN (REPUTATION)
Steht die Datenquelle für kontinuierlich qualitativ hochwertige Daten? (ähnlich Glaubwürdigkeit)

5 OBJEKTIVITÄT (OBJECTIVITY)
Enthalten die Daten keine Wertungen und sind sachlich?

6 FEHLERFREIHEIT (FREE OF ERROR)
Entsprechen die Daten dem, was sie abbilden sollen?

INHALT

7 AKTUALITÄT (TIMELINESS)
Sind die Daten so aktuell, wie es der Zweck erfordert?

8 GRANULARITÄT (GRANULARITY)
Sind die Daten so detailliert und ausdifferenziert, wie es der Zweck erfordert?

9 WERTSCHÖPFUNG (VALUE-ADDED)
Kann ich aus den Daten einen gesellschaftlichen oder wirtschaftlichen Wert schöpfen?

10 VOLLSTÄNDIGKEIT (COMPLETENESS)
Sind alle notwendigen Datenfelder vollständig gefüllt?

11 ANGEMESSENER UMFANG (APPROPRIATE AMOUNT OF DATA)
Reicht mir der Umfang der Daten aus, um meine Frage zu beantworten?

12 RELEVANZ (RELEVANCY)
Sind die Informationen für meinen Zweck notwendig?

NUTZUNG

13 ÜBERSICHTLICHKEIT (CONCISE REPRESENTATION)
Kann ich auf die benötigten Informationen innerhalb eines Datensatzes einfach zugreifen? (Zugänglichkeit innerhalb eines Datensatzes)

14 EINHEITLICHE DARSTELLUNG (CONSISTENT REPRESENTATION)
Entsprechen die Informationen fortlaufend einem festgelegten Format?

15 VERSTÄNDLICHKEIT (UNDERSTANDABILITY)
Sind die Informationen inhaltlich einfach zugänglich?

16 EINDEUTIGE AUSLEGBARKEIT (INTERPRETABILITY)
Wenn verschiedene Personen mit den Daten unabhängig voneinander arbeiten, werden sie die Daten in der gleichen (korrekten) Weise interpretieren?

DARSTELLUNG

Anmerkungen zu den 16 Dimensionen

- Die vorgestellten 16 Dimensionen basieren auf einer Studie der MIT Wissenschaftler*innen Wang und Strong von 1996[1]. Sie sind eine gängige Form, Datenqualität zu bestimmen. Natürlich existieren auch noch andere Herangehensweisen.
- In den seltensten Fällen sind alle Dimensionen relevant. Vielmehr sollten Sie eine Auswahl aus den 16 Dimensionen entsprechend der Sinnhaftigkeit, Ihrer Rolle als Anwender*in oder Bereitsteller*in und dem Zweck treffen.

Von 16 Dimensionen zu einer Strategie

- Sie können die 16 Dimensionen zum losen Reflektieren von vorliegender Datenqualität oder möglichen Datenqualitätsproblemen bei noch unbekannten Datenquellen nutzen. Sie können die Dimensionen auch noch näher definieren und daraufhin Datenqualität messen und eine Bestandsaufnahme durchführen.
- Eine Messung ist empfehlenswert, weil Sie damit Veränderungen nachverfolgen können und entsprechende Maßnahmen treffen können.
- Je mehr Sie von ad-hoc-Maßnahmen zu einer ganzheitlichen Strategie wechseln, desto mehr müssen Sie Prozesse zur Datenqualitätssicherung verstetigen. Damit sind Sie auf dem besten Wege, eine Data-Governance zu implementieren.
- Eine Data-Governance bezeichnet ein klar definiertes Konzept, dass den Umgang mit Daten innerhalb einer Organisation regelt. Das können zum einen Prozesse zur Datenqualität, aber eben auch Datenschutz und Datensicherheit sein.
- Folgende Schritte empfehlen wir zu einer Data-Governance für eine höhere Datenqualität:
  - Ein innerhalb der Organisation einheitliches Verständnis des Kernprozesses bei Datenqualitätsproblemen festlegen und verstetigen (standardisieren)
  - Vorhandene Datenstandards verwenden und Metadaten pflegen
  - Alle betroffenen Stellen in der Organisation einbeziehen und Verantwortlichkeiten klar benennen
  - Prozesse transparent und nachvollziehbar machen (z.B. durch eine Versionskontrolle bei Datensatzänderungen)
  - Auswahl der Bewertungsdimensionen stetig hinterfragen und ggf. anpassen
  - Regelmäßig Datenqualität messen
- Metadaten: Metadaten geben mehr Informationen zum Datensatz, werden jedoch leider häufig vernachlässigt. Metadaten für eigene Datensätze zu pflegen, sollten Sie hoch priorisieren. Damit sichern Sie ab, dass man auch nach einiger Zeit einen qualitativ hochwertigen Datensatz ohne große Mehrarbeit verwenden kann. Bei externen Datensätzen müssen Sie davon ausgehen, dass diese nicht ihren eigenen Qualitätsstandards entsprechen. Das gilt auch für die entsprechenden Metadaten.

1 Richard Y. Wang, Diane M. Strong: Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 12:4, 5–34, 1996.

Data-Governance zieht Kreise

Wenn es um Data-Governance geht, wird vor allem über Effizienz und Optimierung geredet. Data-Governance findet primär auf einer technisch-organisatorischen Ebene statt. Entscheidungen auf dieser Ebene können jedoch weitere Kreise ziehen. Daher ist es wichtig, sich bei der Implementierung von Data-Governance-Strukturen ebenfalls mit gesellschaftlichen Auswirkungen zu beschäftigen. In der Data-Governance einer Organisation manifestieren sich die selbst gesetzten Standards der Organisation. Sie sollte daher konsistent mit den eigenen Wertvorstellungen und Richtlinien sein. Darüber hinaus bietet eine bewusste, gesellschaftliche Aspekte in den Blick nehmende Data-Governance ein großes Potential für die Implementierung von ethischen Strukturen im eigenen Projekt oder in der eigenen Organisation.

Was ist mit der ethisch-gesellschaftliche Dimension gemeint?

Die ethisch-gesellschaftliche Dimension eines Projekts reicht über die technisch-organisatorische oder rein betriebswirtschaftliche Perspektive hinaus. Sie bezieht sich auf die Wirkung eines Projektes (auch über den konkreten Projektzweck hinaus) auf die Gesellschaft.

Warum hat die Art, wie Data-Governance betrieben wird, ethisch-gesellschaftliche Auswirkungen?

Aspekte der Data-Governance wie zum Beispiel Datenschutz und Datenqualität dienen nicht nur dem wirtschaftlichen Erfolg einer Organisation oder eines Unternehmens, sondern haben Auswirkungen auf die Gesellschaft. Umsichtig betrieben und sorgfältig durchdacht ist Data-Governance daher eine Grundlage für gesellschaftlich erwünschtes Handeln.

Wie werden ethisch-gesellschaftliche Auswirkungen beurteilt?

Nicht immer lässt sich ein Ursache-Wirkungs-Zusammenhang umfassend und präzise feststellen. Gerade die weitreichende Dimension der ethisch-gesellschaftlichen Auswirkungen ist nur schwer formal bewertbar und in der Regel nicht quantifizierbar. Ausgangspunkt sollte daher eine grundsätzliche Sensibilisierung für die Thematik sein, ebenso wie die Nutzung von Tools und Konzepten die als Annäherung und Diskursgrundlage in einer Organisation dienen können.

Die verschiedenen Foren auf der mFUND-Konferenz haben erneut verdeutlicht, dass viele Projekte (zumindest auch) einen gesellschaftlichen Mehrwert erzeugen wollen und können. Sei es die Reduktion der Parksuchzeit in der Innenstadt oder die optimierte Auslastung von LKWs auf Frachtfahrten – jeweils wird die Schonung der Umwelt und das menschliche Wohlbefinden als wichtiger Motivationsfaktor der Projektinitiator*innen hervorgehoben. Bundesminister Scheuer betonte bei seiner Keynote die Bedeutung von Innovationen, die unter anderem als Hoffnungsträger dem Klimawandel entgegentreten sollen.

Neben der Euphorie über das große Potential von datengetriebener Innovation sollte aber auch der Blick für mögliche Nebenwirkungen geschärft werden. Die hier aufgeführten Beispiele können helfen, das eigene Projekt ganzheitlicher zu beurteilen und zu diskutieren. Ungeahnte Folgen haben ihre Wurzeln nicht zuletzt auch in unzureichender und unbedachter Arbeit und Verwendung von Daten – auch wenn sich die Folgen nicht primär auf einer technisch-organisatorischen Ebene widerspiegeln.

Achtung, Datenschwund.

Wo liegen die Daten, die ich generiere, verwerte und weitergebe? Erst letzten Monat ist bekannt geworden, dass seit Jahren ca. 16 Millionen Datensätze aus rund 50 Ländern offen im Netz zugänglich sind. Die Daten sind zum Beispiel bei vermeintlich banalen Vorgängen wie dem Erstellen von CDs für Patient*innen über einen Webserver ins Netz gelangt. Sogar während Anonymisierungsprozessen (die dem Schutz von Patient*innen dienen sollten) gelangten Daten ins Netz.

Im Bereich der Mobilität 4.0 sind Gesundheitsdaten eher weniger relevant. Jedoch gehören auch Nutzungs- und Bewegungsdaten zur digitalen Identität von Bürger*innen. Eine ungenügende Sorgfalt im Bereich Datensicherheit könnte unkontrollierbare Folgen haben.

Die Kreise, die meine Arbeit zieht.

Ein eindrückliches Beispiel wie einzelne (auch technische) Entscheidungen in der Summe ungeahnte Folgen haben können, beschreibt die Journalistin Kasmir Hills in einem komplexen Fall[2]: Verschiedene Apps wollten der Problematik von verschwundenen Mobilfunkgeräten Abhilfe schaffen. Sie arbeiteten mit Geodaten und IP-Adressen. Die Datenbestände gingen durch mehrere Hände, bis sie aufbereitet als konkreter Standort in der App der Nutzer*innen erschienen. Dabei führte mangelnde Datenqualität, eine unzureichende Kennzeichnung und die insgesamt sorglose und unregulierte Weitergabe und Verwendung von Datenbeständen zu einem gewaltigen Schaden bei einzelnen Personen. Hill nennt es die Schaffung von einem „House of Horror“: Ahnungslosen Menschen, die das Pech hatten in dem konkreten Haus zu wohnen, auf das die App fast immer verwies, wurden von verschiedensten Stellen als Kriminelle beschuldigt. Bis die Geschehnisse aufgeklärt wurden dauerte es Jahre.

Insbesondere wenn Daten von Dritten verwendet werden, muss darauf geachtet werden, dass die Daten für den vorgesehenen Zweck geeignet sind. Eine eindeutige Kennzeichnung ist dafür zentral. Ein unbedachter Umgang mit datengetriebener Innovation kann unerwartet zu Lasten Unbeteiligter gehen.

2 Mehr Informationen zu diesem Beispiel finden Sie hier: Kasmir Hill: „How Cartographers for the U.S. Militay Inadverently Created a House of Horrors in South Africa”, 09. Januar 2019, http://bit.ly/35qFvUi.

„If you torture the data enough, it will confess.“[3]

Signifikanten Ergebnisse erscheinen für den Erfolg vieler Forschungen und Hypothesen von überragender Bedeutung. Sehr bekannt ist das sogenannte p-Hacking. Der emeritierte Professor und Meteorologe Neville Nicholls beschäftigt sich mit der Schaffung von hochqualitativen Datenbanken, die Vorhersagen über Extremwetterereignisse anhand von historischen Wetterdaten erleichtern sollen. In seinem Artikel „The Insignificance of Significance Testing“[4] adressiert er Probleme bei Tests von Nullhypothesen insbesondere in der Atmosphärenforschung. Er kritisiert, dass viele Forscher*innen an veralteten und inkorrekten statistischen Methoden festhalten um den Erfolg ihres Projektes nicht zu gefährden.

Eindrückliche Ergebnisse zu präsentieren ist das Ziel vieler Förderprojekte des mFUND. Daher stellt sich nicht nur die Frage, mit welchen statistischen Methoden gearbeitet wird, sondern auch, wie ein Projekt damit umgeht, wenn nach langer Konzeptions- und Forschungsphase kein signifikantes Ergebnis vorzuweisen ist. Wie eine Datenanalyse durchgeführt wird, folgt keinem strikten Vorgabenkatalog, sondern obliegt der Entscheidung des jeweiligen Projekts. Diese basiert auf der Abwägung verschiedenster – auch ethischen – Aspekte.

3 Dieser Satz wird Ronald Coase, einem britischen Wirtschaftswissenschaftler, der insbesondere für das Coase-Theorem bekannt wurde, zugeschrieben.

4 Neville Nichols (2001): „The Insignificance of Significance Testing” in Meteorological Society, Vol. 82, S. 981-986.

Ich bin, was ich nutze.

Technische Systeme können Data-Governance unterstützen. Es gibt eine Vielzahl von verschiedenen Programmen, die von Organisationen genutzt werden können. Neben pragmatischen Entscheidungskriterien können Projekte sich damit hervorheben sich über marktrelevante Aspekte Gedanken zu machen: Unterstütze ich mit meiner Entscheidung für eine bestimmte Software ein Monopol oder eine monopolähnliche Stellung? Und wir wirkt sich diese Monopolstellung auf die Gesellschaft aus?

„Who codes matters.“[5]

Joy Buolamwini hat während ihrer Zeit als Graduate Student am MIT feststellen müssen, dass Gesichtserkennungssoftware nicht in der Lage sind ihr Gesicht zu erkennen. Und nicht nur ihr Gesicht: Die Software war systematisch nicht in der Lage, weibliche Gesichter von schwarzen Menschen als solche zu identifizieren. Algorithmen, die mit zu homogenen Trainingsdaten gefüttert werden neigen dazu dem Bias der programmierenden Person zu folgen. Neben der Bedeutung von umfangreichen und diversen Datensets ist in diesem Beispiel die Rolle der Programmierer*innen, sowie der Datenverwalter*innen besonders hervorzuheben: Dass die Gesichtserkennungssoftware nur mit Gesichtern von weißen Menschen angelernt wurde, ist aufgrund einer kognitiven Verzerrung geschehen. Den (weißen) Programmierer*innen ist nicht aufgefallen, dass das Trainingsset nicht die heterogene Realität abbildet.

Übertragbar ist die Problematik zum Beispiel auf Datensets von Verkehrsteilnehmer*innen („Habe ich als Programmiererin alle unterschiedlichen Fortbewegungsarten im Blick: Die SUV-Fahrerin genauso wie die Rollstuhlfahrerin oder konzentriere ich mich vermehrt auf die Gruppe, zu der ich selber gehöre?“). Diverse Teams können dabei ein großer Pluspunkt sein.

5 Siehe zu diesem Beispiel den Ted Talk von Joy Buolamwini vom November 2016, http://bit.ly/2ID4DgE.

Datengetrieben Systeme agieren nicht im luftleeren Raum. Vielmehr stehen sie in Wechselwirkung mit der entsprechenden Organisation, seinen Nutzer*innen sowie der Umwelt. Nicht alle solche Systeme haben jedoch die gleiche Außenwirkung. Ebenso wie die sie einsetzenden Institutionen nicht pauschal bewertet werden können. Es ist daher dienlich das gesellschaftlich-ethische Wirkungspotentials eines jeden Projektes individuell zu betrachten.

Wie identifiziere ich die gesellschaftliche Relevanz meiner Organisation oder meines Projekts?

Um diese Wirkungsdimensionen näher zu bestimmen, haben Kilian Vieth[6] und Dr. Ben Wagner[7] ein Modell zur Berechnung des Teilhabewirkungspotentials entwickelt. Wir haben dieses Modell für die Projekte des mFUNDs abgewandelt, ergänzt und uns auf die verschiedenen Dimensionen beschränkt.

Die Dimensionen gliedern sich in unterschiedlichen Kategorien. Unter Akteure wird der tatsächlichen ökonomischen und politischen Macht hinter den Betreiber*innen der Entscheidungsprozesse Rechnung getragen. Das Prüfkriterien zur sozialen Einbettung behalten die beabsichtigten und unbeabsichtigten gesellschaftlichen Wechselwirkungen im Blick. Daran schließt sich die Frage nach möglichen Konsequenzen und der Einflussnahme auf die Umwelt des jeweiligen Systems an. Die Auswirkungen auf einer Makroebene werden komplementiert durch die Kategorie Menschen, die nach dem Einfluss (auch) auf Individuen fragt. Die letzte Dimension betrachtet die Art der im Projekt verwendenden Daten.

Je relevanter diese Kategorien für ein Projekt sind, desto höhere Anforderungen sind an das jeweilige algorithmische System zu stellen.

Die hinter den Kategorien stehenden Fragen sind exemplarisch. Aufgrund der Diversität der Projekte mag nicht jede Frage für jedes Projekt relevant oder zutreffend sein. Sie dienen vor allem als Orientierung, um die abstrakten Klassifikationen greifbar zu machen.

6 Kilian Vieth ist Projektmanager „Digitale Grundrechte, Überwachung & Demokratie“ für die Stiftung Neue Verantwortung.

7 Dr. Ben Wagner ist Assistant Professor und Direktor des Privacy & Sustainable Computing Lab an der Wirtschaftsuniversität Wien.

DIMENSIONEN GESELLSCHAFTLICHER RELEVANZ VON DATENBASIERTEN ENTSCHEIDUNGSSYSTEMEN[8]

WETTBEWERB
Wer ist der Betreiber des Algorithmus?
Wie viel politische und ökonomische Macht hat er? Besteht effektiver Wettbewerb zwischen mehreren Betreibern?

ABHÄNGIGKEIT
Ist es ein Problem, wenn der Prozess bzw. das produzierte Gut wegfällt oder der Zugang dazu verwehrt wird?
Gibt es alternative Verfahren/Angebote?
Wie hoch sind die Kosten für den Umstieg auf ein alternatives Verfahren? Welche Organisationen oder Personengruppen sind strukturell abhängig von dem Verfahren?

AKTEURE

SELBSTBESTIMMUNG
Inwiefern dient das Verfahren (nur) zur Vorbereitung einer Entscheidung?
Entscheidet das System (de facto oder offiziell) selbstständig?
Wie viel Freiraum zur Änderung und Kontrolle einer algorithmischen Entscheidung hat die Anwenderin?

ANPASSUNG
Wie passen sich Menschen an den Prozess an?
Welche effektive, soziale Autorität hat das algorithmische Verfahren?
Welche (soziale) Bedeutung wird dem Prozess zugeschrieben? Hat er z.B. Diskurshoheit?
Welche Wechselwirkungen zwischen Menschen und Computer verändern unter Umständen das Ergebnis?

SOZIALE EINBETTUNG

REICHWEITE
Wie viele Personen sind von dem Entscheidungsprozess betroffen? Ist die Reichweite des Verfahrens bekannt bzw. begrenzbar? Wie groß ist sie?

SYSTEMVERÄNDERUNG
Hebelt der Entscheidungsprozess solidarische Prinzipien aus? Ist eine Individualisierung systemverändernd? Hat das Verfahren eine potenzielle, eventuell nicht beabsichtigte Transformation des (sozialen) Wirkungsumfelds zur Folge?

KONSEQUENZEN

BEZIEHUNG ZU MENSCHEN
Werden Menschen bewertet? Betrifft es Menschen direkt oder indirekt? Trifft das System teilhaberelevante Entscheidungen?

DISKRIMINIERUNG
Wie stark könnten Menschen durch das Entscheidungsverfahren benachteiligt werden?
Lässt sich für den Output ein diskriminierendes Muster, z.B. aufgrund von Herkunft, Geschlecht, körperlicher Beeinträchtigung, Alter, etc. feststellen bzw. erwarten?

MENSCHEN

Sind die Daten personenbezogen?
Wo kommen die Daten her? Handelt es sich um besonders sensible Daten?
Sind die Daten anonymisiert bzw. pseudonymisiert?
Lassen sich von den Daten Rückschlüsse auf Personen oder Personengruppen ziehen?

DATEN

8 Die Kriterien basieren zum Großteil auf Vieth/Wagner: Teilhabe, ausgerechnet. Wie algorithmische Prozesse Teilhabechancen beeinflussen können, Bertelsmann Stiftung (Hg.), 2017. Dimensionen gesellschaftlicher Relevanz von datenbasierten Entscheidungssystemen

Die Algo.Rules

An die Einschätzung der eigenen gesellschaftlichen Relevanz mithilfe der vorgestellten Dimensionen schließen sich die Algo.Rules als eine sinnvolle Handreichung an, um eine Diskussion über mögliche Maßnahmen und Problembereiche anzustoßen. Die Algo.Rules wurden von der Bertelsmann-Stiftung in Zusammenarbeit mit dem iRights.Lab in einem interaktiven Beteiligungsprozess entwickelt. Es handelt sich bei den Algo.Rules um einen Katalog an formalen Kriterien, die beachtet werden müssen, um eine gesellschaftlich förderliche Gestaltung und Überprüfung von algorithmischen Systemen zu ermöglichen und zu erleichtern. Sie legen eine Grundlage für ethische Erwägungen und für die Um- und Durchsetzung rechtlicher Rahmenbedingungen. Im Fokus der Algo.Rules stehen jene algorithmischen Systeme, die einen direkten oder mittelbaren, immer jedoch signifikanten Einfluss auf das Leben der Menschen oder die Gesellschaft haben. Es gilt: Je stärker der mögliche Einfluss eines algorithmischen Systems auf das Leben der Menschen oder die Gesellschaft ist, desto genauer sollte die Umsetzung der Algo.Rules geprüft werden.

Bei datengetriebenen Innovationen ist die Verwendung von algorithmischen Systemen zentral. Es ist daher unabdingbar, dass sich die Projekte des mFUND mit der Gestaltung dieser Systeme auseinandersetzen.

Die Algo.Rules wurden von Michael Puntschuh und Levke Burfeind bei dem Workshop zur ethischen und gesellschaftlichen Relevanz von Mobilitätsprojekten auf der mFUND-Konferenz als eine mögliche Handreichung vorgestellt. Die Teilnehmenden waren von den Algo.Rules sehr angetan und konnte sich vorstellen, sie als Gesprächsgrundlage mit in ihr Projekt einfließen zu lassen. Aus diesem Grund haben wir uns entschieden, die Algo.Rules auch für alle anderen mFUND-Projekte in diesem Report zur Verfügung zu stellen.

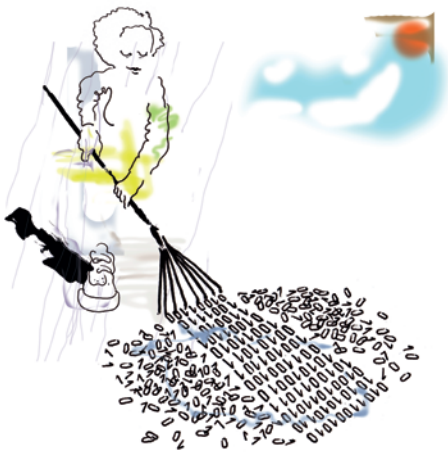
Es handelt sich hier um eine verkürzte Version der Algo.Rules. Umfassende Informationen erhalten Sie auf algorules.org.

Was ist ein Algorithmisches System?

Ein Algorithmus ist eine eindeutige Handlungsvorschrift zur Lösung eines vorab definierten Problems. Ein algorithmisches System ist ein System einer oder mehrerer Algorithmen, die in Software implementiert wurden, um Daten zu erfassen, zu analysieren und Schlüsse zu ziehen, die zur Lösung eines vorher definierten Problems beitragen sollen. Das System kann dabei selbstlernend sein oder vorprogrammierten Entscheidungsregeln folgen. Die Bewertung eines algorithmischen Systems im Rahmen der Algo.Rules umfasst auch den soziotechnischen Gesamtkontext, in den die Software eingebettet ist, wie zum Beispiel die Deutung und Interpretation des Ergebnisses und die Ableitung einer Entscheidung durch Anwender*innen des Systems. Der Geltungsanspruch der Algo.Rules umfasst den gesamten Prozess der Entwicklung und Einbettung algorithmischer Systeme in den jeweiligen sozialen Kontext.

DIE 9 ALGO.RULES

1



KOMPETENZ AUFBAUEN

Die Funktionsweise und die möglichen Auswirkungen eines algorithmischen Systems müssen verstanden werden.

2

VERANTWORTUNG DEFINIEREN

Für die Auswirkungen des Einsatzes eines algorithmischen Systems muss stets eine natürliche oder juristische Person verantwortlich sein.

3

ZIELE UND ERWARTETE WIRKUNG DOKUMENTIEREN

Die Ziele und die erwartete Wirkung des Einsatzes eines algorithmischen Systems müssen vor dessen Einsatz dokumentiert und abgewogen werden.

DIE 9 ALGO.RULES

4

SICHERHEIT GEWÄHRLEISTEN

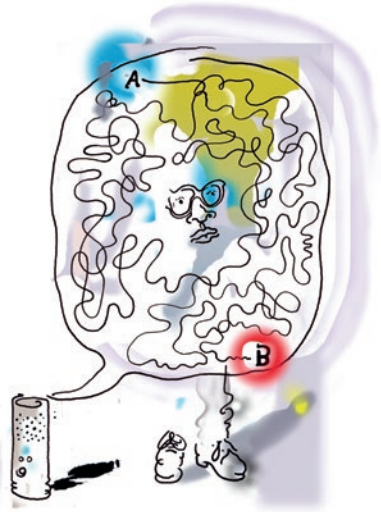
Die Sicherheit eines algorithmischen Systems muss vor dessen Einsatz getestet und fortlaufend gewährleistet werden.

5

KENNZEICHNUNG DURCHFÜHREN

Der Einsatz eines algorithmischen Systems muss gekennzeichnet sein.

6



NACHVOLLZIEHBARKEIT SICHERSTELLEN

Entscheidungsfindung eines algorithmischen Systems muss stets nachvollziehbar sein.

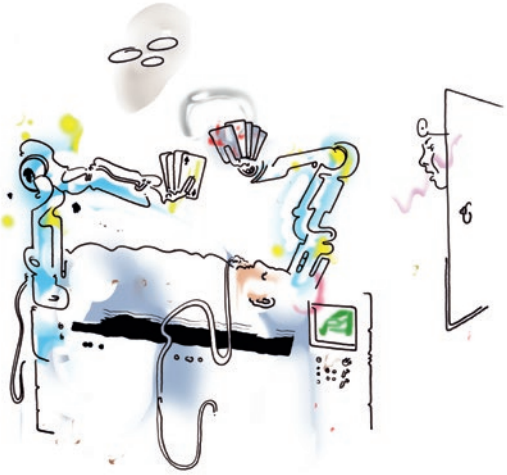
DIE 9 ALGO.RULES

7

BEHERRSCHBARKEIT ABSICHERN

Ein algorithmisches System muss während seines gesamten Einsatzes gestaltbar sein und bleiben.

8



WIRKUNG ÜBERPRÜFEN

Die Auswirkungen eines algorithmischen Systems müssen regelmäßig überprüft werden.

9

BESCHWERDEN ERMÖGLICHEN

Fragwürdige oder die Rechte einer betroffenen Person beeinträchtigende Entscheidungen eines algorithmischen Systems müssen erklärt und gemeldet werden können.

Wie geht es weiter

Im Rahmen der vom iRights.Lab durchgeführten Interviews sind einige mFUND Projekte besonders hinsichtlich ihrer fortgeschrittenen Implementierung von Data-Governance Strukturen aufgefallen. Daher werden wir in den nächsten Monaten Case Studies mit ausgewählten Projekten durchführen. So können insbesondere neue Projekte in den mFUND Förderlinien, aber auch bestehende Organisationen, die ihre Data-Governance verbessern wollen, von der Expertise und den Erfahrungen anderer mFUND-Projekte profitieren.

Das iRights.Lab

Das iRights.Lab ist ein unabhängiger Think Tank zur Entwicklung von Strategien und praktischen Lösungen, um die Veränderungen in der digitalen Welt vorteilhaft zu gestalten. Wir unterstützen öffentliche Einrichtungen, Stiftungen, Unternehmen, Wissenschaft und Politik dabei, die Herausforderungen der Digitalisierung zu meistern und die vielschichtigen Potenziale effektiv und positiv zu nutzen. Dazu verknüpfen wir in einem interdisziplinären Team rechtliche, technische, ökonomische, betriebswirtschaftliche und gesellschaftspolitische Expertise. Wir gestalten und begleiten Strategieprozesse, führen interdisziplinäre Analysen durch und entwickeln Handlungsempfehlungen. Wir möchten die inhaltliche Auseinandersetzung zu zentralen Fragestellungen der Digitalisierung fördern.

Die Begleitforschung des iRights.Lab zu Data-Governance bei datengetriebenen Innovationen läuft bis Ende 2020. Neben dem iRights.Lab beschäftigt sich auch die Begleitforschung des Wissenschaftlichen Instituts für Infrastruktur und Kommunikationsdienste (WIK) im Rahmen eines Arbeitsforums mit dem Thema Datenschutz und Compliance. Ziel der gesamten Begleitforschung des mFUND ist es, datenbasierte Innovationsprojekte zu unterstützen.

Das iRights.Lab-Team

Foto: Andi Weiland | www.andiweiland.de

Der mFUND

IIm Rahmen der Forschungsinitiative mFUND fördert das Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) seit 2016 Forschungs- und Entwicklungsprojekte rund um digitale datenbasierte Anwendungen für die Mobilität 4.0. Neben der finanziellen Förderung unterstützt der mFUND mit verschiedenen Veranstaltungsformaten die Vernetzung zwischen Akteuren aus Politik, Wirtschaft und Forschung sowie den Zugang zum Datenportal mCLOUD.

Siehe auch: www.mfund.de

Impressum

iRights.Lab GmbH, Schützenstraße 8, D-10117 Berlin

Telefon: +49 (0)30 40 36 77 230
Fax: +49 (0)30 91 68 18 49

E-Mail: kontakt@irights-lab.de
www.irights-lab.de

facebook: facebook.com/irights.lab
twitter: @irightslab
instagram: @irights.lab

Geschäftsführer: Philipp Otto

Registergericht: Amtsgericht Berlin-Charlottenburg, Registernummer: HRB 185640
Finanzamt Friedrichshain/Prenzlauer Berg | Steuer-Nr. 30/359/50503 | USt-IdNr.: DE311181302

Inhaltlich Verantwortlicher i.S.d. § 55 Abs. 2 RStV: Philipp Otto (Anschrift siehe oben)

Autorin: Levke Burfeind
Mitarbeit: Michael Puntschuh, Ulrike Thalheim
Redaktion: Annika Albert, Dr. Wiebke Glässer, Philipp Otto

Illustrationen der Algo.Rules: Jan Robert Dünnweller, © Bertelsmann Stiftung
Gestaltung: beworx.de, Titelfoto: Todd Quackenbush / Unsplash

Das Projekt „Data-Governance im Innovationsprozess“ wird gefördert vom Bundesministerium für Verkehr und digitale Infrastruktur im Rahmen des mFUND.
www.mfund.de

iRights.Lab
Think Tank für die digitale Welt