

Data-Governance-Report

**mFUND-Begleitforschung
vom Think Tank iRights.Lab**

**Mai 2019
Ausgabe 03**



Neuigkeiten aus der mFUND-Begleitforschung des iRights.Lab

Der bisherige Verlauf der vom iRights.Lab unternommenen mFUND-Begleitforschung zu Data-Governance in datengetriebenen Innovationsprozessen hat gezeigt: Datenschutz und Datensicherheit sind auch für vom mFUND geförderte Einrichtungen sehr wichtige Themen, da sie gerade für kleinere Organisationen mit viel Aufwand verbunden und im Detail schwierig umzusetzen sind. Das von uns durchgeführte Forum während der mFUND-Konferenz im Oktober 2018 in Berlin, in dem wir das Projekt, unsere Vorgehensweise sowie erste Erkenntnisse einem größeren Publikum mit Vertreter*innen aus teilnehmenden Organisationen vorstellen konnten, brachte aber auch zutage: Beide Teilbereiche sind im Rahmen einer umfassenden Data-Governance Aspekte eines viel größeren Komplexes, in dem es noch einige andere Punkte zu beachten gilt.

Und gerade in der alltäglichen Arbeit mit (selbst erhobenen oder sonst erworbenen) Daten spielt oft ein anderer Faktor eine noch viel entscheidendere Rolle: die der Qualität der Daten. Mehr als jede*r dritte Befragte hatte in unserer ersten Online-Umfrage im Februar und März 2018 angegeben, dass der Aspekt „Datenqualität/Datenintegrität“ als große Herausforderung bei der Verwendung von Datensätzen im vom mFUND geförderten Projekt anzusehen ist. Noch vor dem Datenschutz landete er damit auf dem ersten Platz.

Datenqualität zeichnet sich durch Faktoren wie Vollständigkeit, Integrität, Konsistenz, Korrektheit, Genauigkeit oder Relevanz aus. Es geht bei diesem Aspekt also darum, ob die vorliegenden Daten auch tatsächlich für den Zweck des jeweiligen Projektes geeignet sind. Ist die Datenstruktur zu heterogen oder variiert die Qualität der einzelnen Daten in einem Datensatz zu stark, kann schnell der Erfolg der Unternehmung gefährdet sein. Die schwierige Herausforderung, eine gleichbleibend hohe Datenqualität zu gewährleisten, spiegelte sich auch in den Aussagen der Umfrageteilnehmer*innen wider.

Deshalb ist der Aspekt der Datenqualität der Schwerpunkt des vorliegenden dritten mFUND Governance Report. Zunächst präsentieren wir Ihnen eine thematisch bezogene Auswertung unserer Online-Umfrage aus diesem Frühjahr. In einem ausführlichen Interview erläutert der Data-Governance-Experte Marco Geuer, wie Datenqualität verbessert werden kann und worauf es diesbezüglich bei der Umsetzung einer Data-Governance innerhalb einer Organisation zu achten gilt. Denn, so Geuer unmissverständlich: „Ohne Datenqualitätsmanagement keine Data-Governance“. Darüber hinaus geben wir Ihnen einen ersten Einblick in die „Checkliste zur Optimierung von Datenqualität und Datenqualitätsmanagement“. Sie wird den teilnehmenden Organisationen als Handreichung bei der Erarbeitung einer an die eigenen Bedürfnisse angepassten Strategie hoffentlich wertvolle Dienste leisten.

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 u.a. in den Bereichen Datenschutz und Datensicherheit in Einklang gebracht werden, um Innovationen zu fördern.

Executive Summary

Erkenntnisse der Online-Umfrage mit mFUND-geförderten Projekten:

- Über 60 Prozent der Befragten beschrieben den Aufwand zur Datenqualitätssicherung als mehr oder viel mehr als erwartet.
- Ein Großteil der Projekte standardisiert Prozesse zur Datenqualitätssicherung.
- Hinsichtlich der Nutzung externer Daten wird insbesondere der Zugang zu den Datensätzen und die vorhandene Dokumentation kritisch bewertet.

Erkenntnisse aus dem Interview mit Herrn Geuer:

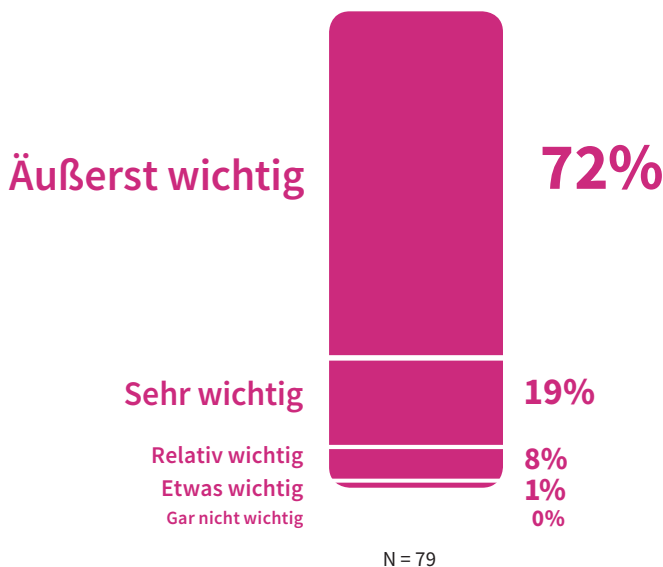
- Eine Data-Governance kommt nicht ohne Maßnahmen zur Sicherung von Datenqualität aus.
- Prozesse zu Datenqualität sollten schrittweise umgesetzt werden.
- Die angelegten Prozesse sollten technisch implementiert und unterstützt werden.

Ergebnisse aus unserer Umfrage

Im Februar und März 2019 haben wir, wie schon im Vorjahr, eine Online-Umfrage zum Umgang mit Daten durchgeführt. Befragt wurden mFUND-Projekte, welche im Vorjahr nicht an der Umfrage teilgenommen haben. Auf Grundlage der 2018 im Rahmen der Begleitforschung durchgeführten qualitativen Interviews haben wir einige Veränderungen am Fragebogen vorgenommen. Weitere Informationen zur Methode der Umfrage können Sie der Infobox entnehmen.

Für einen Großteil der vom mFUND geförderten Projekte ist der Umgang mit Daten sehr wichtig oder äußerst wichtig (91 Prozent). Projekte, die auf der Nutzung von Daten basieren, benötigen ein hohes Maß an Datenqualität. Deshalb steht in dieser Auswertung das Thema Datenqualität im Vordergrund.

Wie relevant ist die Arbeit mit Datensätzen für Ihr Projekt?

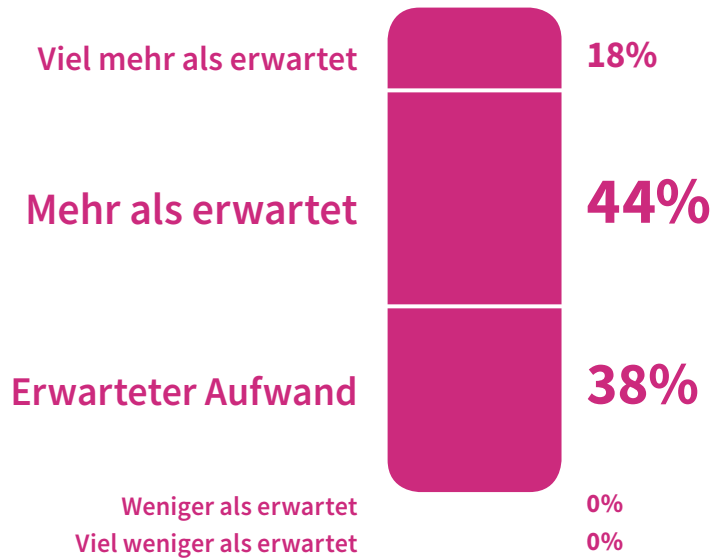


Im Februar und März 2018 starteten wir die erste Umfragephase unter den vom mFUND geförderten Projekten. Nun fand im Februar und März 2019 die zweite Umfragephase statt. In der Umfrage geht es um die Innovationskultur in Deutschland und um die Voraussetzungen für datenbasierte Innovationen. Ein zweiter Schwerpunkt liegt auf dem internen Datenmanagement und dem Umgang mit Datenschutz, Datensicherheit und Datenqualität. An der diesjährigen Umfragephase nahmen insgesamt 98 Vertreter*innen von mFUND-Projekten teil (2018: 127), 79 Teilnehmer*innen beantworteten die Umfrage vollständig (2018: 86). Für die Umfrage wurden ausschließlich Vertreter*innen von mFUND-Projekten eingeladen, welche nicht an der Umfrage im Vorjahr teilgenommen haben. Die Umfrage wird kontinuierlich fortgeführt, sodass Projekte aus neuen Förderrunden ebenfalls teilnehmen können und die Umfrageergebnisse auch neue Entwicklungen berücksichtigen.

Datenqualität

In den qualitativen Interviews, welche im Verlauf des Vorjahres durchgeführt wurden, erwähnten viele unserer Gesprächspartner*innen das Thema Datenqualität. Sie schilderten, dass sie nicht auf den Aufwand vorbereitet waren, den sie für eine zufriedenstellende Datenqualität aufbringen mussten. Deshalb haben wir in dieser Umfragephase einige Fragen zu diesem Themenkomplex gestellt. Den Aufwand, der benötigt wird um die Datenqualität sicherzustellen, bezeichneten 62 Prozent als mehr oder viel mehr als erwartet. Niemand der Befragten gab an, dass sich der Aufwand geringer als erwartet herausstellte. Deshalb sollten Projekte zukünftig schon bei der Projektplanung mögliche zusätzliche Arbeit beim Datenmanagement berücksichtigen.

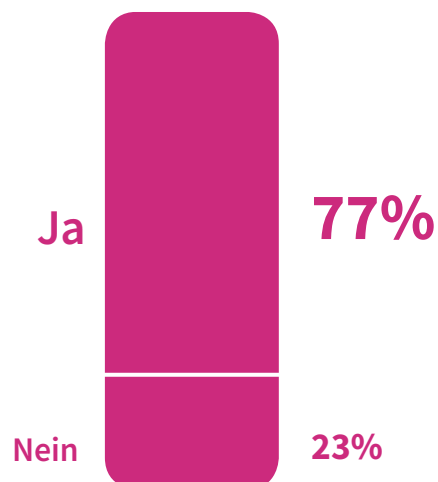
Wie beschreiben Sie den Aufwand, um die Datenqualität von internen und/oder externen Datenquellen sicherzustellen?



N = 73, exkl. „weiß nicht“.

Standardisierung ist ein wichtiger Faktor in Datenmanagementprozessen zur Verbesserung der Datenqualität. Dahingehend sind auch die laufenden mFUND-Projekte gut aufgestellt: 3 von 4 der Befragten gaben an, Prozesse für eine höhere Datenqualität zu standardisieren.

Standardisieren Sie die Prozesse, um Datenqualität sicherzustellen?



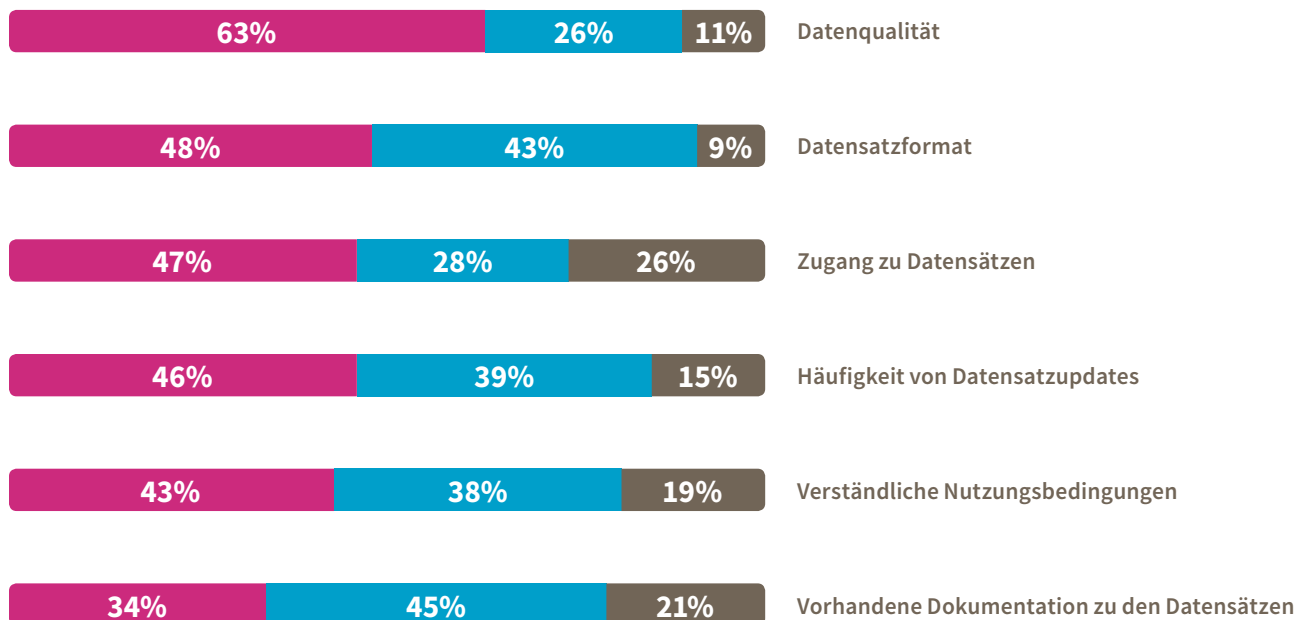
N = 78, exkl. „weiß nicht“.

Externe Datenquellen

Mobilitätsdaten kommen aus sehr unterschiedlichen Quellen. Im vorangegangenen Data-Governance-Report erstellten wir eine Übersicht zu eben diesen. Am häufigsten wurden Landes- oder Bundesbehörden als Datenquellen genannt (39 Prozent, N=79), hierbei handelt es sich um externe Datenquellen. Nicht immer ist es im Voraus möglich einzuschätzen, welche Herausforderungen durch die Nutzung von externen Datenquellen entstehen. Deshalb haben wir die Befragten die Nutzung von externen Datensätzen für ihre Arbeit bewerten lassen. Überraschenderweise schneidet gerade die Datenqualität als bester Aspekt ab – 63 Prozent bewerten sie bei externen Datensätzen als gut bis sehr gut. In allen weiteren Bewertungskategorien wurde die Situation von nicht einmal der Hälfte der Befragten als gut bis sehr gut beschrieben. Während knapp die Hälfte den Zugang zu Datensätzen als gut bis sehr gut bewerten, beschreibt dies ein weiteres Viertel der Befragten als schlecht bis sehr schlecht. Etliche Datensätze scheinen inzwischen gut erhältlich zu sein, andere dafür noch gar nicht. Auch die Dokumentation der Datensätze (z.B. über Metadaten) lässt zu wünschen übrig. Sie wird nur von einem Drittel der Befragten als gut bis sehr gut bewertet. Doch gerade die Dokumentation ist wichtig, wenn Daten verarbeitet und interpretiert werden sollen.

Wie bewerten Sie für sich die Nutzung von externen Datensätzen hinsichtlich folgender Prozesse?

sehr gut / gut mittel schlecht / sehr schlecht



N = 73, exkl. „weiß nicht“.

Interview mit Marco Geuer

iRights.Lab: Seit wann und warum widmen Sie sich dem Thema Datenqualität? Marco Geuer: Ich bin seit 20 Jahren im Bereich Business Analytics tätig. Immer wieder stellte sich heraus, dass schlechte Datenqualität (DQ) entweder Projekte erheblich verzögerte, der Projekterfolg geringer ausfiel als geplant und Vertrauen in die Aussagekraft von Kennzahlen aber auch bei neuen Softwarelösungen verloren gehen. Wie heißt es so schön: „garbage in, garbage out“! Daher bildet sich immer stärker die Erkenntnis heraus: Ohne ein Data Quality Management (DQM) geht es nicht. Dabei trieb und treibt mich immer noch die Frage um, wie man ein DQM ökonomisch sinnvoll implementiert.

Sie sind bei Unitymedia Director für Data Governance & Quality. Wie sind die Themen Data Governance und Datenqualität miteinander verknüpft?

Ganz einfach, ohne ein DQM gibt es keine Data Governance. Ich würde sogar jeder Organisation empfehlen, im Rahmen einer Einführung einer Data-Governance-Organisation mit dem Thema DQM zu beginnen. Man lernt sehr viel über seine Datenlogistik und die Verarbeitungsregeln die man für eine gute Data-Governance-Strategie kennen muss. Darüber hinaus sind frühzeitig bei den ersten Datenqualitäts-Verbesserungen sichtbare „Quick wins“ zu erzielen. Bei Unitymedia haben wir mittlerweile unser Angebotspektrum erweitert. Wir sprechen z.B. nicht mehr nur von Data Profiling im Zusammenhang mit Datenqualität, sondern wir machen eine „Leistungsbewertung von Daten, abhängig vom Zweck, die die Daten erfüllen sollen“. Wir schauen uns nicht nur das Datenqualitätsniveau an, sondern bewerten auch, welche Aussagekraft die Informationen haben, die aus diesen Daten gewonnen werden können. Dies führen wir möglichst zu Beginn von datengetriebenen Projekten im Prototyping-Verfahren durch, um die Investitionen in Projekte zu schützen. Lieber zehn Prototypen für kleinere Summen umsetzen als ein Millionen-Projekt in den Sand setzen. Mal abgesehen davon, dass die Lernkurve viel schneller steigt. Darüber hinaus führen wir einen Datenqualitäts-Index (DQ-Index) ein, der das Datenqualitätsniveau unserer wichtigsten Daten entlang unseres Kernprozesses abbildet. Der DQ-Index fließt unter anderem als strategische Kennzahl in unser unternehmensweites Strategy Dashboard für das Top-Management ein. Auch hier steigt stetig das Bewusstsein, dass zur Erlangung einer Datenintelligenz eine gute Datenqualität gehört.

In Ihrem Blog stellen Sie 11 Indikatoren vor, mit denen Organisationen Datenqualität quantifizieren können. Warum ist die Quantifizierung so relevant?

Hier zitiere ich gerne Peter F. Drucker (US-amerikanischer Ökonom, Anm. d. Red.): „Was man nicht messen kann, kann man nicht steuern.“ Das gilt bei Datenqualität genauso. Wenn ich mein gewünschtes Datenqualitätsniveau nicht definiere, messe und bewerte, werde ich auch nicht in der Lage sein, die richtigen Maßnahmen abzuleiten, um besser zu werden. Anders ausgedrückt, wer sich nicht um die Datenqualität kümmert, wird in einer digitalisierten Welt zunehmend an Wettbewerbsfähigkeit einbüßen, je mehr sich die Datenqualität verschlechtert.

Wie sind Sie auf die Kriterien gekommen?

Über die Kriterien findet man einiges in der Literatur bzw. aus der Forschung von Richard Wang (MIT). Die 11 Kriterien sind die, die sich in der Praxis aus meiner Erfahrung heraus bewährt haben und messbar sind. Wer sich grundsätzlich mit Qualitätsmanagementsystemen beschäftigt, wie z.B. der DIN ISO 9001 ff. (eine Norm für Mindestanforderungen an ein Qualitätsmanagementsystem, Anm. d. Red.), dem sind diese Kriterien nicht unbekannt.

Marco Geuer

ist Director Data Governance & Quality im Business Intelligence Competence Center (BICC) bei der Unitymedia NRW GmbH. Hier ist er verantwortlich für den Aufbau und die stetige Weiterentwicklung von nachhaltigen Data Governance Strukturen sowie eines nachhaltigen Data- und Information Quality Managements.

Marco Geuer betreibt das Blog „Business Information Excellence | Transparenz schafft Handlungsalternativen“, das unter www.business-information-excellence.de zu erreichen ist.

Welches der Kriterien wird am meisten unterschätzt?

Aus meiner Sicht wird am häufigsten das Kriterium der Vollständigkeit missachtet und unterschätzt. Die meisten gehen davon aus, dass die Daten, die sie nutzen grundsätzlich vollständig sind. Dies ist häufig nicht so. So werden heutzutage viele Daten über Sensoren erhoben. Gerade da können aber auch Lücken in der Datenerhebung entstehen, denen man nachgehen muss. Interessanterweise versucht man immer mehr Algorithmen einzusetzen, um Lücken in Daten zu füllen z.B. für eine Vorhersage.

Gibt es noch weitere Methoden, um Datenqualität zu erfassen und zu verbessern?

Zusätzlich zu den 11 Kriterien nutzen wir methodisch Logikbäume, um das Ursache-Wirkungsgefüge von Datenqualität auf die Geschäftsprozesse und die Zielerreichung zu bewerten. Diese Bewertung ist dann die Basis, um die sinnvollsten Maßnahmen mit dem größten positiven Hebel abzuleiten. Darüber hinaus schaffen wir permanent Bewusstsein für das Thema z.B. durch Vorträge, Blogbeiträge im Intranet und monatlich stattfindende halbtägige Datenqualitäts-Seminare.

Sollen Organisationen schon heute bestimmen, welches Datenqualitätsniveau sie für zukünftige Aufgaben und Fragestellungen benötigen? Oder raten Sie, die Datenqualität erst bei konkreten Anlässen zu verbessern?

Auf jeden Fall ist es wichtig, eine Datenqualitätsstrategie zu haben in der die Verfahren, Methoden und Werkzeuge festgelegt und standardisiert sind. Es wäre ökonomisch nicht sinnvoll, das immer wieder neu aufsetzen zu müssen. Darüber hinaus ist es durchaus sinnvoll, bedarfsgerecht und iterativ zu agieren, abhängig von der Unternehmensstrategie und Wichtigkeit der Daten.

Wie lässt sich das benötigte Datenqualitätsniveau für noch unbekannte Aufgaben bestimmen?

Wichtig ist zu wissen, welche Informationen in welcher Qualität für zukünftige Aufgaben benötigt werden. Davon kann abgeleitet werden, welcher Datenumfang wichtig ist und welches Datenqualitätsniveau nicht unterschritten werden darf. Damit wären wir wieder bei den DQ-Kriterien und Logikbäumen.

Viele mFUND-Projekte arbeiten nicht mit selbst erhobenen Daten, sondern nutzen externe Datenquellen. Wie können sie bei Daten, die nicht in der eigenen Organisation entstanden sind, eine zufriedenstellende Datenqualität sichern?

Das kommt darauf an, welchen Einfluss man auf den Datenlieferanten nehmen kann. Im direkten Verhältnis Kunde – Lieferant ist das eher möglich als z.B. bei Open Data. Bei Open Data würde ich mir wünschen, dass es einen DQ-Mindeststandard und ein DQ-Index gibt, der dementsprechend das Datenqualitätsniveau angibt und mitveröffentlicht wird. Zurzeit kann ich nur empfehlen, auch die externen Daten von Beginn an einer Datenqualitätsprüfung zu unterziehen und auf der Basis für sich zu entscheiden, ob und mit welchen Schwächen diese verwendet oder weiterverarbeitet werden können.

Worin unterscheiden sich kleine und große Unternehmen beim Umgang mit internen und externen Daten?

Dies ist meiner Meinung nach keine Frage der Organisationsgröße, sondern der Datenvielfalt, Datenmenge und Wichtigkeit der Daten für den Geschäftserfolg. Je mehr der Geschäftserfolg von Daten und deren Qualität abhängig ist, umso mehr sollten Organisationen bereit sein, in die Datenqualität zu investieren.



Marco Geuer

Welche Rolle spielen Softwarelösungen bei der Umsetzung einer Data Governance?

Die richtige Software-Lösung für Data Governance und das Data Quality Management ist elementar. Aber auch hier gilt, diese muss dem konkreten Bedarf entsprechen. An einer bewussten Software-Evaluierung kommt man nicht vorbei, was übrigens unter Qualitätsgesichtspunkten immer gelten sollte. Entscheidend bei der Evaluierung kann sein, wie viele Datenqualitätsregeln im Programm enthalten sind und welche Regeln zusätzlich aufgestellt werden müssen.

Wie muss sich die Organisationsstruktur oder -kultur wandeln, um eine Data Governance tatsächlich umsetzen zu können?

Data Governance und Quality muss in die DNA einer Organisation übergehen. Es muss klar sein, wie wichtig Daten und der Umgang mit diesen für eine Organisation sein sollen. Dementsprechend muss die Strategie und Taktik ausgerichtet werden. An dem Thema kommt keine Organisation mehr vorbei. Das zeigt z.B. die Einführung der DSGVO. Damit bin ich schon gezwungen eine Data-Governance-Organisation zu betreiben. Übrigens war bei diesem Thema gut zu erkennen, wer schon dementsprechend aufgestellt ist. Wer da schon weiter war, hat sich mit der DSGVO-Konformität leichter getan als andere. Das gilt übrigens auch für die Digitalisierung. Hier sind die Organisationen führend, die sich schon länger um Data Governance und Quality kümmern und diese in ihre Strategie fest verankert haben.

Kriterien für Datenqualität

Marco Geuer hat 2017 in seinem Blog „Business Information Excellence“ 11 Kriterien für die Evaluation von Datenqualität vorgestellt:

- 1 **Vollständigkeit**
- 2 **Eindeutigkeit**
- 3 **Korrektheit**
- 4 **Aktualität**
- 5 **Genauigkeit**
- 6 **Konsistenz**
- 7 **Redundanzfreiheit**
- 8 **Relevanz**
- 9 **Einheitlichkeit**
- 10 **Zuverlässigkeit**
- 11 **Verständlichkeit**

Auf Basis der Kriterien entwickelt das iRights.Lab eine entsprechende Empfehlung in Gestalt einer Checkliste/Handreichung speziell für die im mFund geförderten Projekte. Dabei werden auch die Erkenntnisse unserer bisherigen Begleitforschung einfließen. Die Checkliste wird gemeinsam mit Vertreter*innen von mFUND-Projekten bei unserem Local Meet-Up im Mai diskutiert und weiterentwickelt. Ziel ist es, möglichst praxisnah eine Hilfestellung zu geben. Mit der Checkliste möchten wir Sie bei folgenden Schritten unterstützen:

- **Systematisch bei Datenqualitätsproblemen vorgehen**
- **Häufige Datenqualitätsprobleme vorausschauend erkennen**
- **Schritte für eine Datenqualitätsstrategie definieren**
- **Strategisches Vorgehen bei Datenqualitätsproblemen als Teil einer Data Governance implementieren**

Lesen Sie dazu mehr im nächsten Data-Governance-Report.

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.

Impressum

iRights.Lab GmbH
Schützenstraße 8, D-10117 Berlin
Telefon: +49 (0)30 89 37 01 03
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)

Redaktion: Annika Albert, Wiebke Glässer, Henning Lahmann, Philipp Otto, Ulrike Thalheim

Gestaltung: beworx.de, Titelfoto: Gorodenkoff

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

www.mfund.de

Das iRights.Lab-Team



Foto: Andi Weiland | www.andiweiland.de

Der mFUND

Im 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

