

MAX VON GRAFENSTEIN, ANNIKA ULICH

Kurzpapier: Data Governance

Ein Framework zur Erfassung „erfolgreicher“ Data Governance-Modelle

ABSTRACT

Daten zugänglich zu machen, zu nutzen und zu teilen wird heute ganz allgemein als wichtiger Baustein für nachhaltiges Wachstum und Wohlstand anerkannt. Entsprechend arbeiten Akteure in verschiedenen gesellschaftlichen Bereichen an ihren je eigenen Digitalisierungs-, Daten- und Smart City-Strategien. Data Governance wird dabei als elementare Voraussetzung erkannt. Geht man ins Detail, findet man sich jedoch schnell mit einer Wolke aus Begriffen wie Data Pools and Data Lakes, Intermediären, Plattformen und Datentreuhändern konfrontiert. Selten bzw. nur unter Schwierigkeiten findet man heraus, wie die dahinter stehenden Data Governance-Strukturen konkret beschaffen sind. Doch erst wenn wir diese Strukturen anhand eines Frameworks konsistent beschreiben können, sind wir in der Lage, unterschiedliche Data Governance-Modelle zu vergleichen und auf die Frage hin bewerten, welche Modelle in welchen Kontexten „erfolgreich“ sind. Hierfür muss man zunächst klären, was Data Governance „erfolgreich“ macht und was die besonderen Herausforderungen dabei sind. Das vorliegende Discussion Paper fasst die Ergebnisse der bisherigen HIIG-Forschungsprojekte “Data Governance: Definition einer gemeinsamen Forschungsgrundlage” sowie “Data Governance: Reconciling conflicting interests in data – a research framework” zusammen.

KEYWORDS

Data governance, information governance, data protection, law, value of data, risks of data, context-dependency, coordination, centralisation, de-centralisation.

CITATION/ ZITATION

Grafenstein, M. v. & Ulich, A. (2021). Kurzpapier: Framework zur Erfassung „erfolgreicher“ Data Governance-Modelle. HIIG Discussion Paper Series 2021-11. 7 pages. <https://doi.org/10.5281/6325396>.

LICENCE/LIZENZ

This work is distributed under the terms of the Creative Commons Attribution 4.0 Licence (International) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (<https://creativecommons.org/licenses/by/4.0/>). Copyright remains with the authors.

AUTHOR INFO / AFFILIATION / FUNDING ETC.

Max von Grafenstein, Einstein Center Digital Future (Universität der Künste), Alexander von Humboldt Institut für Internet und Gesellschaft

Annika Ulich, Alexander von Humboldt Institut für Internet und Gesellschaft

CONTENTS/INHALT

1 WAS IST DATA GOVERNANCE UND WIESO IST ES WICHTIG?	3
2 WAS SIND DIE HERAUSFORDERUNGEN FÜR “ERFOLGREICHE” DATA GOVERNANCE?	3
2.1 Kontextabhängigkeit des Werts und der Risiken von Daten	3
2.2 Einfluss des Rechts auf die Verwendbarkeit der Daten	4
2.3 Erforderliche Koordination auf rechtlicher, organisatorischer und technologischer Ebene	5
2.4 Grad der Zentralisierung bzw. Dezentralisierung, Standardisierung und Partizipation	6
3 WIE KÖNNEN WIR DATA GOVERNANCE-MODELLE BESCHREIBEN UND DAVON LERNEN?	7
4 REFERENCES	8

1 WAS IST DATA GOVERNANCE¹ UND WIESO IST ES WICHTIG?

Daten zugänglich zu machen, zu nutzen und zu teilen wird heute ganz allgemein als wichtiger Baustein für nachhaltiges Wachstum und Wohlstand anerkannt.² Entsprechend arbeiten Akteure in verschiedenen gesellschaftlichen Bereichen an ihren je eigenen Digitalisierungs-, Daten- und Smart City-Strategien.³ Data Governance wird dabei als elementare Voraussetzung erkannt.⁴ Geht man ins Detail, findet man sich jedoch schnell mit einer Wolke aus Begriffen wie Data Pools⁵ and Data Lakes⁶, Intermediären⁷, Plattformen⁸ und Datentreuhändern⁹ konfrontiert. Selten bzw. nur unter Schwierigkeiten findet man heraus, wie die dahinter stehenden Data Governance-Strukturen konkret beschaffen sind:

- Wer darf auf welche Daten für welche Zwecke unter welchen Bedingungen zugreifen?
- Mit welchen Technologien werden die Daten wo und wie erhoben, gespeichert und verarbeitet?
- In welchen organisatorischen Strukturen bzw. mit welchen Verfahren werden diese und weitere Parameter für die Verarbeitung der Daten festgelegt?

Doch erst mit Antworten auf diese Fragen können wir unterschiedliche Data Governance-Modelle vergleichen und auf die Frage hin bewerten, welche Modelle in welchen Kontexten „erfolgreich“ sind. Hierfür muss man zunächst klären, was Data Governance „erfolgreich“ macht und was die besonderen Herausforderungen dabei sind.

2 WAS SIND DIE HERAUSFORDERUNGEN FÜR „ERFOLGREICHE“ DATA GOVERNANCE?

Das Ziel „erfolgreicher“ Data Governance ist, die mit der Verarbeitung von Daten auftretenden Interessenkonflikte so gut es geht aufzulösen. Dafür ist es erforderlich, sowohl den Wert der Daten als auch die mit ihrer Verarbeitung verbundenen Risiken je nach Perspektive der beteiligten Akteure möglichst optimal zu schöpfen bzw. zu kontrollieren.

2.1 Kontextabhängigkeit des Werts und der Risiken von Daten

Der Erfolg von Data Governance-Strukturen hängt dabei heute mehr denn je davon ab, ob sie auf die Kontextabhängigkeit und damit verbundene Dynamik des Werts und der Risiken der Daten reagieren

¹ Dieses Kurzpapier bildet eine Zusammenfassung des in Veröffentlichung begriffenen Beitrags M v Grafenstein, Data Governance: A Research Framework – Reconciling conflicting interests in data by taking the dynamics and contextual dependency of its value and risks into account dar; siehe bereits Wernick, A., Olk, C., & Grafenstein, M. v. (2020). Defining Data Intermediaries. Technology and Regulation, 65–77. DOI: 10.26116/techreg sowie Grafenstein, M. v., Wernick, A., & Olk, C. (2019). Data Governance: Enhancing Innovation and Protecting Against Its Risks. Intereconomics, 54 (4), 228-232. DOI: 10.1007/s10272-019-0829-9.

² Siehe nur die Datenstrategie der Bundesregierung: Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum.

³ Siehe etwa die Digitalisierungs- und Datenstrategien der EU Kommission und der Bundesregierung sowie die Digitalisierungs- und Smart City-Strategie der Stadt Berlin.

⁴ Siehe z.B. Data governance and data policies at the European Commission.

⁵ Siehe z.B. EU Kommission, A European data strategy (COM(2020) 66 final), S. 5, die immerhin davon ausgeht – wenn auch ohne weitere Beschreibung – dass solche Pools sowohl zentral als auch dezentral ausgestaltet sein könnten.

⁶ Siehe etwa R Hai, S Geisler, C Quix, Constance: An Intelligent Data Lake System, SIGMOD '16: Proceedings of the 2016 International Conference on Management of Data.

⁷ Siehe z.B. L Tuukka, K Yki, Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach.' (2017) 4 Big Data & Society 3; siehe auch Verhulst et al (n 22) 11, die den begriff 'Trusted Intermediary' verwenden.

⁸ Siehe etwa A Gawer, 'Bridging differing perspectives on technological platforms: Toward an integrative framework.' 43 Research policy (2014) 1239; European Commission 'Free flow of data' (n 59) 17; Verhulst et al (n 22) 20.

⁹ Vgl. etwa den Vorschlag der EU Kommission für ein Daten-Governance-Gesetz.

können. Der Wert und die Risiken von Daten lassen sich heute nur noch bedingt statisch, also zu einem bestimmten Zeitpunkt für die Zukunft vorher bestimmen. Heute können der Nutzungszweck und -kontext ständig wechseln und damit auch ihr Wert und ihre Risiken. Die Herausforderung dabei ist, dass erst wenn Daten maschinell ausgelesen und in einem bestimmten Kontext zur informationellen Grundlage von Entscheidungen werden, sich Wert und Risiken konkretisieren. Bis zur Nutzung in einem konkreten Kontext bzw. bis zur Bestimmung eines solchen Nutzungszwecks bleiben Wert und Risiken abstrakt. Gleichzeitig lassen sich die Risiken zumindest technisch nur kontrollieren, solange sie maschinell verarbeitet werden. Sobald die Daten ausgelesen und von Personen zur Grundlage von Entscheidungen werden, lässt sich die Information meist nur noch organisatorisch-rechtlich kontrollieren (etwa über Verschwiegenheitsvereinbarungen). Während sich der Wert der Daten also erst mit der Nutzung konkretisiert, verlieren Datenhalter oder Betroffene (zumindest technisch) zunehmend an Kontrolle. Beim Teilen von Daten für dies oft zu einem **Wert-Risiko-Dilemma**, dass der Datenhalter im Teilen der Daten einen konkreten Kontrollverlust und damit meist ein konkretes Risiko sieht, während der Datennutzer dem Datenhalter noch kein konkretes Wertversprechen machen kann. Erst im Laufe seines eigenen Innovationsprozesses bildet sich ein konkreter Nutzwert der Daten heraus, an dem er den Datenhalter beteiligen kann. Aus den Verhaltenswissenschaften wissen wir, dass wir selten ein konkretes Risiko gegen ein abstraktes Wertversprechen eintauschen – so wenig tut dies ein Datenhalter. Data Governance-Strukturen müssen diese Dynamiken bezüglich Wertschöpfung und Risikokontrolle widerspiegeln, indem sie das beschriebene Dilemma bestmöglich auflösen.

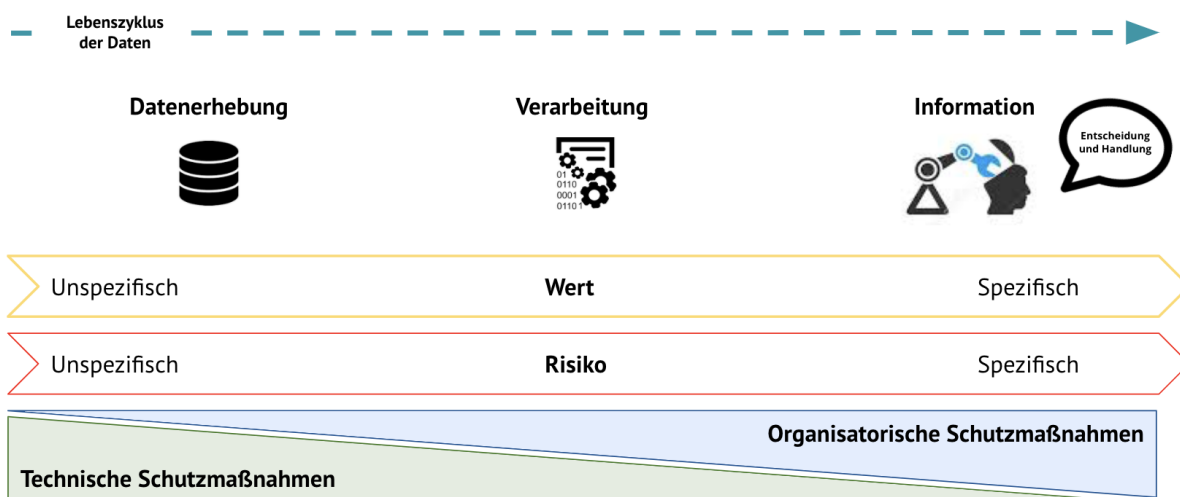


Abb. 1: Das Wert-Risiko-Dilemma

2.2 Einfluss des Rechts auf die Verwendbarkeit der Daten

Hinzu kommt, dass der Gesetzgeber insb. in der EU Risiken immer umfassender reguliert. Das prominenteste Beispiel ist das Datenschutzrecht, allen voran die Datenschutz-Grundverordnung (DSGVO), die mit ihrem extrem weiten Anwendungsbereich Datenverarbeitungen in fast jedem gesellschaftlichen Winkel erfasst. Daneben kommen weitere Regularien in Betracht wie der Schutz von Geschäftsgeheimnissen, des Wettbewerbs und der IT-Sicherheit (sowie evtl. bald auch die Nutzung von Algorithmen mit „Künstlicher Intelligenz“ usw.). Das hat Auswirkungen auf herkömmliche Zielsetzungen von Data Governance. Denn ein wesentliches Ziel von Data Governance ist die Sicherstellung von **Datenqualität**, wonach Daten gebrauchstauglich („fit for use“) gemacht und gehalten werden sollen. In einem stark regulierten Umfeld bedeutet das, dass Daten nur dann eine hohe Qualität aufweisen, wenn sie

unter Anwendung dieser (und weiterer) Gesetze auch verwendet werden dürfen. Die Herstellung bzw. Aufrechterhaltung von Datenqualität ist also keine rein technische Frage mehr, sondern bekommt eine normative Dimension. Datenqualität sicher zu stellen setzt damit ein Zusammendenken verschiedener (analytischer) Data Governance-Dimensionen voraus: Neben der technologischen sowie organisatorischen bzw. verfahrenstechnischen Dimension kommt die rechtliche bzw. normative Dimension hinzu. Diese Dimensionen sind eng verschränkt, was sich wieder am Beispiel des Datenschutzrechts veranschaulichen lässt.

2.3 Erforderliche Koordination auf rechtlicher, organisatorischer und technologischer Ebene

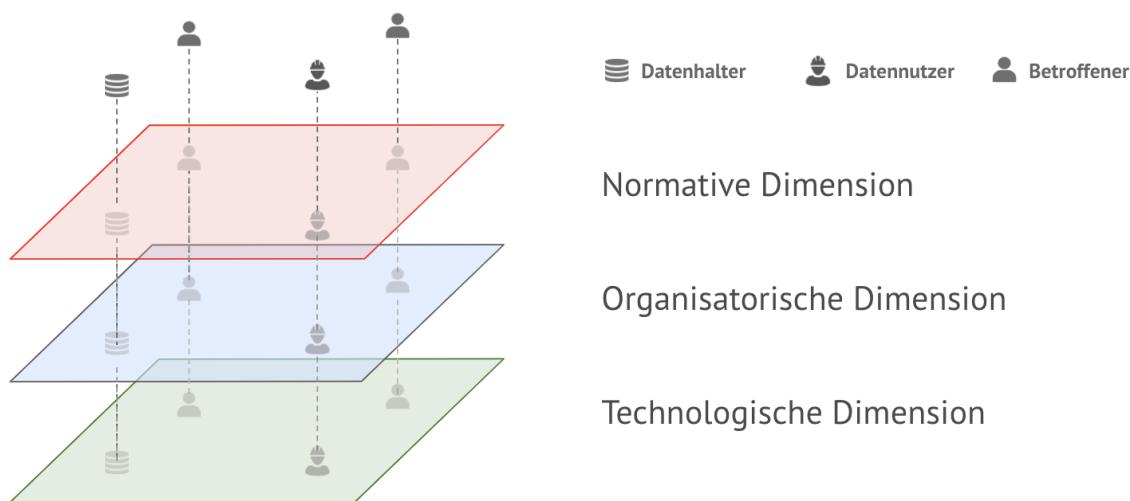


Abb. 2: Die drei Data Governance-Ebenen

Im Datenschutzrecht hat die wissenschaftliche Diskussion und (darauf aufbauend) der Gesetzgeber relativ früh erkannt, dass sich die Risiken der Verarbeitung personenbezogener Daten weniger nach ihrer Art oder ihrem Erhebungskontext bestimmen lassen als vielmehr nach ihrem Verwendungszweck bzw. -kontext. Auch hat man erkannt, dass sich die Risiken vor allem dann wirksam kontrollieren lassen, wenn man die datenschutzrechtlichen Anforderungen direkt in das technische-organisatorische System implementiert, in das die Datenverarbeitung eingebettet ist. Diese Erkenntnis ist heute in der EU Datenschutz-Grundverordnung (DSGVO) in Form des Ansatzes Datenschutz durch Technikgestaltung (Art. 25 Abs. 1 DSGVO) umgesetzt und lässt sich auf die Umsetzung weiterer Gesetze, etwa zum Schutz von Geschäftsgeheimnissen, des Wettbewerbs und natürlich der IT-Sicherheit übertragen.

Nach diesem Ansatz müssen sich die Verarbeiter von Daten (z.B. im Datenschutz der „Controller“ bzw. „Processor“) und weitere Akteure, die in rechtlicher, organisatorischer oder technologischer Hinsicht in eine Datenverarbeitung eingebunden sind, so koordinieren, dass sie die Risiken für den Datenschutz, Geschäftsgeheimnisse, den Wettbewerb und/oder die IT-Sicherheit wirksam kontrollieren. Die von der Datenverarbeitung nachteilig Betroffenen (im Datenschutzrecht also die sog. Data Subjects) sollten dabei in einer möglichst aktiven Rolle i.R.d. verschiedenen Dimensionen eingebunden werden. Diese **Partizipation** erlaubt es, nicht nur die Risiken, sondern auch den Wert der Daten optimal zu schöpfen (siehe etwa die politikwissenschaftliche Debatte zu deliberativen Prozessen bzw. die wirtschaftswissenschaftlichen Ansätze unter dem im Bereich Open Innovation). Die in normativer Hinsicht verantwortlichen Akteure (z.B. Geschäftsführer, Politiker) sollten mit Blick auf die organisatorische und technologische Ebene daher vor allem auf solche Dienstleister zurückgreifen, die ihnen bei der Erfüllung der normativen Erwartungen helfen. Das bedeutet wiederum, dass tendenziell nur solche Dienstleister zum Zuge kommen, die diese

normativen Erwartungen erfüllen können. In der Gesamtschau der dafür erforderlichen Koordination können typischerweise eine Reihe unterschiedlicher Rollen bzw. Vermittlungsinstanzen („Intermediäre“) beschrieben werden.

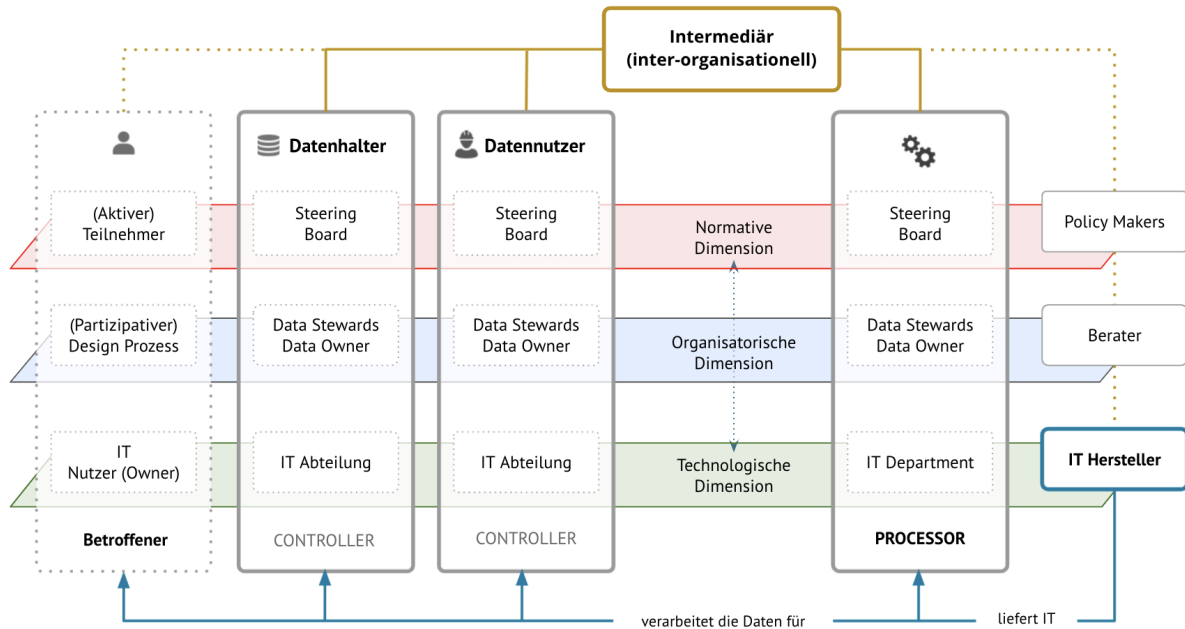


Abb. 3: Überblick über die möglichen Rollen

2.4 Grad der Zentralisierung bzw. Dezentralisierung, Standardisierung und Partizipation

Neben der Kontextabhängigkeit des Werts und der Risiken der Daten sowie der wechselseitigen Abhängigkeit der verschiedenen Data Governance-Dimensionen spielt der Grad der Zentralisierung bzw. Dezentralisierung innerhalb der Data Governance-Dimensionen eine wichtige Rolle. Üblicherweise wird diese Frage vor allem in Hinsicht auf die technologische Dimension diskutiert. Zusammengefasst beobachtet man hier eine Art Zielkonflikt: Auf der einen Seite geht ein hoher Zentralisierungsgrad mit hoher Skalierbarkeit einher. Ein Beispiel sind im Cloud-Bereich die sogenannten Hyperscaler (insb. AWS, Microsoft Azure und Google Cloud). Auf der anderen Seite konzentriert sich damit die faktische Kontrolle über die Technologie – und damit auch der Zugriff auf die mit ihr verarbeiteten Daten – auf wenige Akteure. In dezentralen Systemen verbleibt die Kontrolle dagegen bei all jenen, die ihren je eigenen technologischen Systemteil beisteuern. Aus diesem Grund werden dezentrale Systeme etwa im Datenschutz favorisiert. Ähnliches lässt sich im industriellen Bereich beobachten, wenn die Preisgabe von Geschäftsgeheimnissen befürchtet wird. Der damit einhergehenden Heterogenität der einzelnen Systemteile und der folglich sinkenden Skalierbarkeit versucht man dabei über eine Standardisierung der Schnittstellen zu begegnen. Das ist der Gedanke – um auf den Cloud-Bereich zurück zu kommen – der zum Beispiel hinter der Initiative Gaia-X steht.

Aber nicht nur auf der technologischen Ebene sondern auch in normativer und organisatorischer bzw. verfahrenstechnischer Hinsicht lassen sich unterschiedliche Zentralisierungs- bzw. Dezentralisierungsgrade beschreiben. Zum Beispiel ob bestimmte Datenzugangs- und -verwendungsregeln von einem marktbeherrschenden Plattformbetreiber präzise festgelegt, innerhalb eines Konsortiums vage ausgehandelt oder zwischen einzelnen Marktteilnehmer jedes Mal neu verhandelt werden, beeinflusst wesentlich die Transaktionskosten und Skalierbarkeit der Regelung und ihre Anwendung. Auch hier geht es um Einfluss und Kontrolle, aber eben in normativer Hinsicht. Die „De-/Zentralisierung“ wird hier

typischerweise mit dem Begriff des Grads der Teilnahme beschrieben (siehe bereits oben). Ähnlich verhält es sich bei den Organisationsformen und Verfahren, mit denen Technologien implementiert, Regeln aufgestellt und Daten zur Verfügung gestellt, geteilt und verwendet werden. Hier können Lösungen präzise, aber aufwändig für die jeweiligen Bedürfnisse individualisiert oder für eine Vielzahl von zumindest ähnlichen Fällen standardisiert zur Verfügung gestellt werden. Hier spricht man vor allem vom Standardisierungsgrad.



Abb. 4: Typologie verschiedener Data Governance-Strukturen

3 WIE KÖNNEN WIR DATA GOVERNANCE-MODELLE BESCHREIBEN UND DAVON LERNEN?

Nur wenn wir Data Governance-Modelle präzise beschreiben können, sind wir in der Lage, diese zu vergleichen und in Hinblick auf ihr „erfolgreiches“ Konfliktlösungspotential zu bewerten. In diesem Sinne versucht das vorliegende Kurzpapier, einen ersten, groben Analyserahmen zur Verfügung zu stellen, anhand derer wir von (in Berlin, Deutschland, Europa und weltweit existierenden) Data Governance-Strukturen mit Blick auf die oben beschriebenen Herausforderungen für eigene Data Governance-Projekte lernen können. Nach dem hier vorgeschlagenen Analyserahmen lauten die wesentlichen Fragen:

- Wer darf auf welche Daten für welche Zwecke unter welchen Bedingungen zugreifen?
- Mit welchen Technologien werden die Daten wo und wie erhoben, gespeichert und verarbeitet?
- In welchen organisatorischen Strukturen bzw. mit welchen Verfahren werden diese und weitere Parameter für die Verarbeitung der Daten festgelegt?

Ein besonderes Augenmerk liegt dabei auf den Fragen, inwieweit die jeweiligen Data Governance-Strukturen auf die Kontextabhängigkeit des Werts und der Risiken der Daten reagieren (können), wer auf welche Weise in die Entscheidungsprozesse zur Festlegung der Datenzugangs- und -verwendungsregeln eingebunden ist (zum Beispiel betroffene und/oder interessierte Kreise in der Bürgerschaft) und wer die Technologie rechtlich oder faktisch kontrolliert. Mit dem vorliegenden Analyseraster hoffen wir, Data Governance-Strukturen danach beschreiben und bewerten zu können, wie gut sie die mit der Datenverarbeitung typischerweise auftretenden Interessenkonflikte lösen. Die Erkenntnisse sollen wiederum in die Weiterentwicklung des Analyserasters einfließen, das sich lediglich als ein Anstoß und – vor allem – als „Living Document“ versteht.

4 REFERENCES

- Die Bundesregierung. (2021): Datenstrategie der Bundesregierung. Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum. Abgerufen am 10. November 2021, von <https://www.bundesregierung.de/breg-de/suche/datenstrategie-der-bundesregierung-1845632>.
- Europäische Kommission. (2020): Data governance and data policies at the European Commission. Abgerufen am 10. November 2021, von https://ec.europa.eu/info/publications/data-governance-and-data-policies-european-commission_en.
- Europäische Kommission. (2020): A European strategy for data. Abgerufen am 14. Dezember 2021, von https://ec.europa.eu/info/sites/default/files/communication-european-strategy-data-19feb2020_en.pdf
- Europäische Kommission. (2021): Digitaler Kompass 2030: der europäische Weg in die digitale Dekade. Abgerufen am 14. Dezember 2021, von <https://eur-lex.europa.eu/legal-content/de/TXT/?uri=CELEX%3A52021DC0118>
- Europäischer Rat. (2021): EU will gemeinsame Datennutzung vereinfachen: Rat legt Standpunkt zum Daten-Governance-Gesetz fest. Abgerufen am 10. November 2021, von <https://www.consilium.europa.eu/de/press/press-releases/2021/10/01/eu-looks-to-make-data-sharing-easier-council-agrees-position-on-data-governance-act/>
- Gawer, A. (2014): Bridging differing perspectives on technological platforms: Toward an integrative framework. In: *Research Policy*, 43 (7), 1239-1249. DOI: <https://doi.org/10.1016/j.respol.2014.03.006>.
- Grafenstein, M. v. (forthcoming). Data Governance: A Research Framework - Reconciling conflicting interests in data by taking the dynamics and contextual dependency of its value and risks into account.
- Grafenstein, M. v., Wernick, A., & Olk, C. (2019). Data Governance: Enhancing Innovation and Protecting Against Its Risks. In: *Intereconomics*, 54 (4), 228-232. DOI: <https://doi.org/10.1007/s10272-019-0829-9>.
- Hai, R., Geister, S. & Quix, C. (2016): Constance: An Intelligent Data Lake System. In : *SIGMOD '16: Proceedings of the 2016 International Conference on Management of Data*, 2097-2100. DOI: <https://doi.org/10.1145/2882903.2899389>.
- Lehtiniemi, T., & Kortensniemi, Y. (2017): Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach. In: *Big Data & Society*. DOI: <https://doi.org/10.1177/2053951717721935>.
- Senatsverwaltung für Stadtentwicklung und Umwelt Berlin. (2015): Smart City-Strategie Berlin. Abgerufen am 14. Dezember 2021, von https://smart-city-berlin.de/fileadmin/user_upload/PDFs/Strategie_Smart_City_Berlin.pdf
- Wernick, A., Olk, C., & Grafenstein, M. v. (2020). Defining Data Intermediaries. A Clearer View Through the Lens of Intellectual Property Governance. In: *Technology and Regulation*, 2020, 65–77. DOI: <https://doi.org/10.26116/techreg>.