Data-Governance-Framework für das Digital Urban Center for Aging and Health (DUCAH)[1]

1. INTERSEKTORALE, INTERDISZIPLINÄRE UND INTERINDUSTRIELLE INNOVATION: DAS DIGITAL URBAN CENTER FOR AGING AND HEALTH

Das Projekt *Digital Urban Center for Aging and Health (DUCAH)* setzt sich aus unterschiedlichen Stakeholdern u.a. aus Wirtschaft, Sozialwirtschaft, Forschung und Technologie zusammen. Gemeinsames Ziel ist die Schaffung eines Raumes für Innovationen im Gesundheitswesen rund um das Altern in der digitalen Gesellschaft. Wie können Menschen beispielsweise mittels Ambient Assisted Living Systemen in Kombination mit einer digital vernetzten Sorgegemeinschaft länger in ihrem Quartier und zu Hause gepflegt werden? Um dazu beizutragen, werden perspektivisch im Rahmen des DUCAH Daten erzeugt, gesammelt und verarbeitet. Auch besonders schützenswerte Gesundheitsdaten fallen darunter. Um dieser Vielfalt der Datenquellen und potentiellen Nutzungszusammenhängen im Rahmen des DUCAH gerecht zu werden, braucht es eine verlässliche Data Governance für das Projekt.

2. WAS IST DATA GOVERNANCE 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

1 Dieses Kurzpapier (Stand: 22.11.2021, Version 1) bildet eine Konkretisierung des Diskussionspapiers „Framework zur Erfassung ‚erfolgreicher' Data Governance-Modelle" (im Erscheinen), aufbauend auf dem in Veröffentlichung begriffenen Beitrag M. v. Grafenstein, „Reconciling Conflicting Interests through Data Governance: A Research Framework", für den Bereich DUCAH; 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.

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 Pools[5] and Data Lakes[6], Intermediären[7], Plattformen[8] und Datentreuhändern[9] 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?**

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. Auch für eine Entwicklung einer Data-Governance-Strategie sind diese Fragestellungen von zentraler Bedeutung.

3. WAS SIND DIE HERAUSFORDERUNGEN FÜR „ERFOLGREICHE" DATA GOVERNANCE?

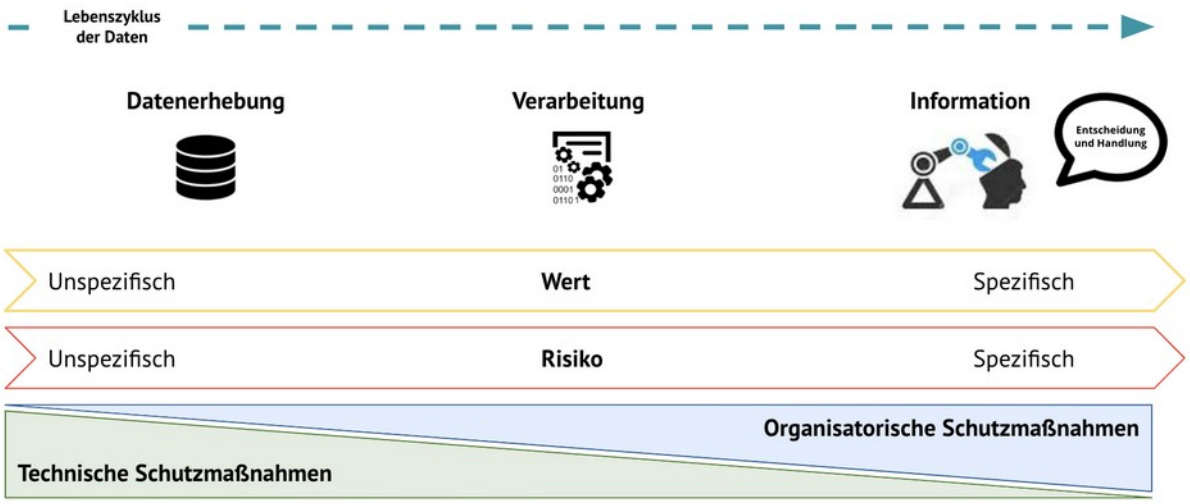
Das Ziel „erfolgreicher" Data Governance ist, die mit der Verarbeitung von Daten auftretenden Interessenkonflikte so gut es geht aufzulösen. Dafür ist es erforderlich, sowohl den Wert der Daten als auch die mit ihrer Verarbeitung verbundenen Risiken je nach Perspektive der beteiligten Akteure möglichst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

a) 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 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

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.

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.

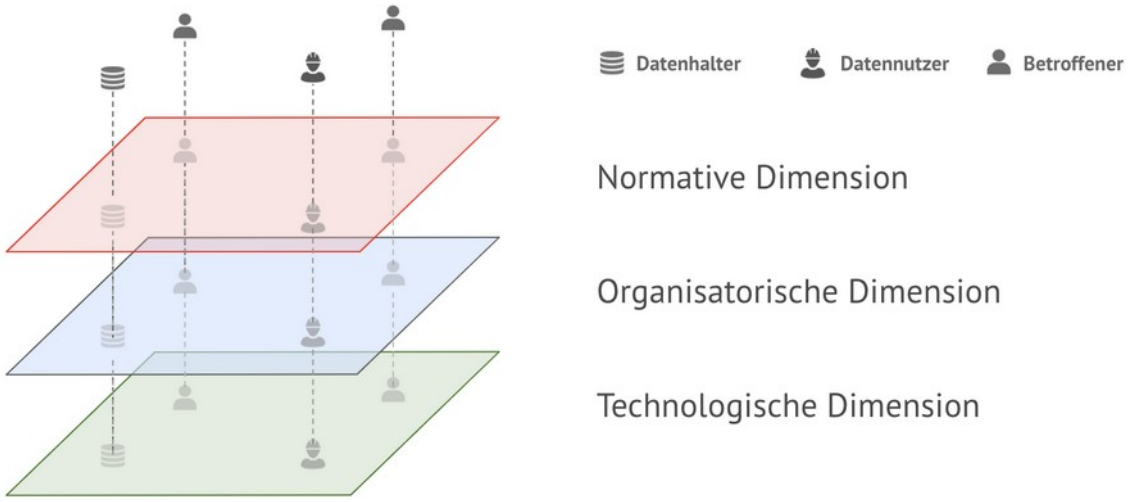


b) Einfluss des Rechts auf die Verwendbarkeit der Daten

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.). Gerade im Gesundheitsbereich gibt es zahlreiche, oftmals sehr strenge Datenschutzvorschriften. 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 sicherzustellen 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.

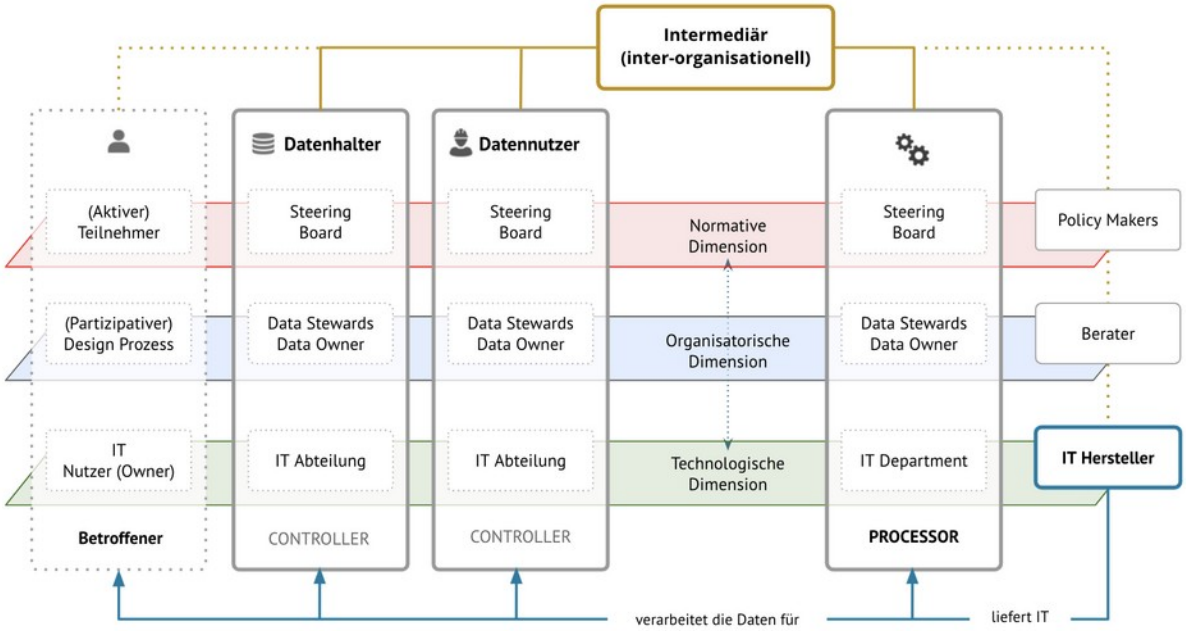
c) Erforderliche Koordination auf rechtlicher, organisatorischer und technologischer Ebene



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 *Verwendungs*zweck bzw. -kontext. Auch wurde 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 Datenschutzrecht 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 im Rahmen der verschiedenen Dimensionen eingebunden werden. Diese **Partizipation** erlaubt es, nicht nur die Risiken, sondern auch den Wert der Daten optimal zu schöpfen (siehe etwa der im Kontext der Corona-Pandemie verstärkte Ruf nach mehr Partizipation im Gesundheitswesen zur Verbesserung kommunaler Gesundheitsversorgung)[10]. 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.



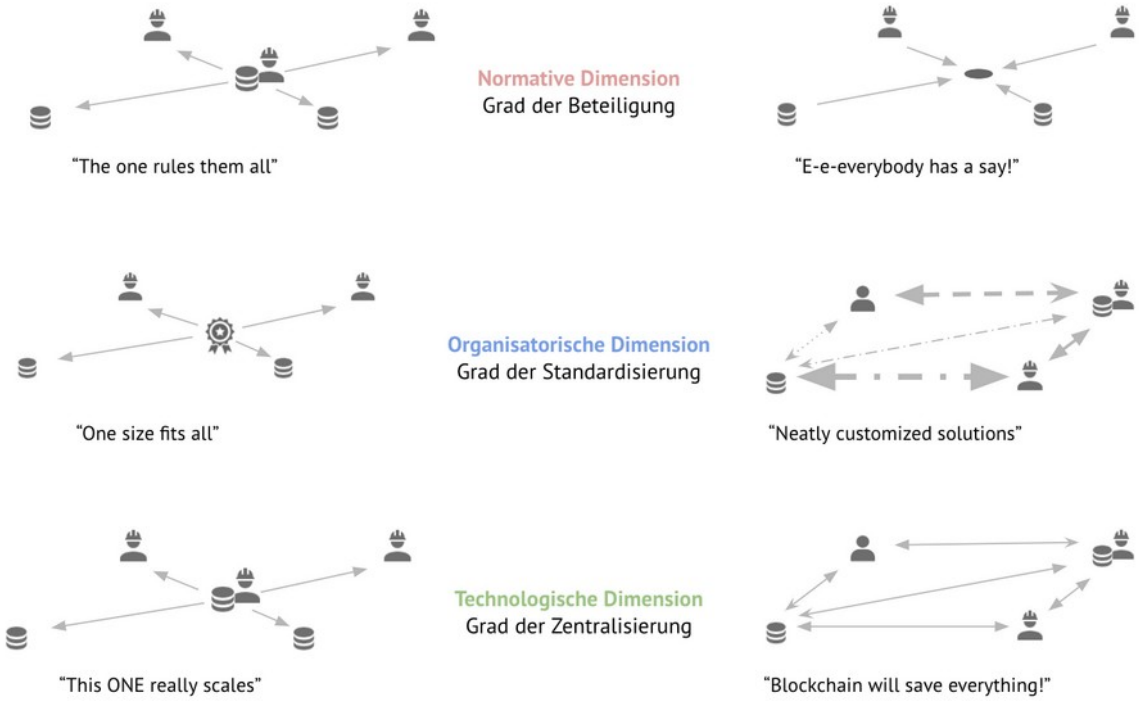
d) Grad der Zentralisierung bzw. Dezentralisierung, Standardisierung und Partizipation

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

10 Siehe etwa https://www.bosch-stiftung.de/de/presse/2021/06/fuer-einen-neustart-im-gesundheitssystem-robert-bosch-stiftung-praesentiert, zuletzt abgerufen am 17. November 2021.

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 Marktteilnehmern jedes Mal neu verhandelt werden, beeinflusst wesentlich die Transaktionskosten und Skalierbarkeit der Regelsetzung 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.



4. WIE KÖNNEN DATA GOVERNANCE-MODELLE BESCHRIEBEN WERDEN UND WAS KANN DUCAH DARAUS LERNEN?

Nur wenn wir Data-Governance-Modelle präzise beschreiben können, sind wir in der Lage, diese zu vergleichen und in Hinsicht auf ihr „erfolgreiches" Konfliktlösungspotential zu bewerten. In diesem Sinne kann das vorliegende Kurzpapier einen ersten, groben Rahmen für das DUCAH aufzuzeigen, anhand derer die beteiligten Stakeholder mit Blick auf die oben beschriebenen Herausforderungen für das eigene Data-Governance-Projekt lernen können.

Für den Anfang lassen sich für DUCAH folgende erste Überlegungen festhalten:

- Die unter Punkt 2 festgehaltenen Fragen des hier vorgeschlagenen Analyserahmens geben einen ersten Anhaltspunkt für den grundlegenden Aufbau der Data-Governance-Struktur im DUCAH und sollten Ausgangspunkt für das Finden geeigneter Benchmarks sowie die Entwicklung einer eigenen Data-Governance-Struktur sein.

- Für die Beantwortung dieser Fragen und die Entwicklung einer Data-Governance-Strategie sind ausreichend Ressourcen (insbesondere personell-finanziell und zeitlich) vonnöten. Die Suche nach Antworten ist eng mit der Konzeptionierung des DUCAH an sich verbunden. So muss z.B. festgelegt werden, wer auf welche Weise in die Entscheidungsprozesse zur Festlegung der Datenzugangs- und -verwendungsregeln eingebunden ist (zum Beispiel betroffene und/oder interessierte Genoss*innen im Pflegequartier) und wer die Technologie rechtlich oder faktisch kontrolliert, sowie welche Entscheidungsbefugnisse sinnhaft in genossenschaftliche Strukturen integriert werden können.

- Die Kernidee des DUCAH ist die Partizipation und Selbstverwaltung, die gerade auch in der Gründung einer Genossenschaft in eine Organisationsform überführt wird. Hier müssen zwischen Zentralisierung bzw. Dezentralisierung, Standardisierung und Partizipation ein ausgewogener Weg beschritten werden. Zwei Herausforderungen seien an dieser Stelle schon genannt:
    - Aufgrund der rasanten Entwicklungen auf dem Gesundheitsmarkt in der Pandemie und im Hinblick auf die rechtlichen Rahmenbedingungen (z.B. DiGa-Gesetzgebungen) sollte die Data-Governance-Struktur von DUCAH nicht starr angelegt werden, sondern auf die Kontextabhängigkeit des Werts und der Risiken der Daten reagieren können.
    - Eine andere zentrale Herausforderung für die Data Governance im DUCAH wird sein, die unter Punkt 3.a) beschriebenen Dynamiken bezüglich Wertschöpfung und Risikokontrolle dynamisch abbilden zu können (insb. in Hinsicht auf sich ständig ändernde rechtliche Anforderungen), da bereits jetzt davon ausgegangen werden muss, dass sie im DUCAH Kontext auftreten werden.

- Die bisherigen Überlegungen zur Integration von DUCAH in GAIA-X sollten weiter verfolgt werden. Der hohe Anspruch des europäischen Vorzeigeprojekts DUCAH an Datenschutz schließen eine Integration in nicht vor dem Zugriff durch außereuropäische Akteure gesicherte Cloud-Systeme aus (zum Beispiel AWS). Der Gefahr der sinkenden Skalierbarkeit der Daten für soziale Innovation aus dem DUCAH-Projekt heraus, könnte die Standardisierung der Schnittstellen mittels GAIA-X entgegenwirken.

- Dienstleister und Partner müssen gefunden werden, die bei der Erfüllung der im Rahmen des DUCAH-Projekts zu definierenden normativen Erwartungen helfen und diese normativen Erwartungen wiederum selber erfüllen können. Hier besteht nach erster Abwägung die Erwartung, dass dies für DUCAH eher kein „One Size Fits All"-Lösung sein wird.

Die Bedeutsamkeit der Entwicklung einer eigenen Data-Governance-Strategie liegt für DUCAH auf der Hand. Menschen können länger in ihren eigenen vier Wänden leben bleiben, wenn sinnhafte datengetriebene Innovationen dazu beitragen. Und diese können nur gewonnen werden durch stakeholderübergreifende Zusammenarbeit und Innovation. Abgesichert durch verlässliche Data-Governance-Strukturen trauen sich Organisationen, ihre „Datensilos" für Forschung und Entwicklung zu öffnen.