

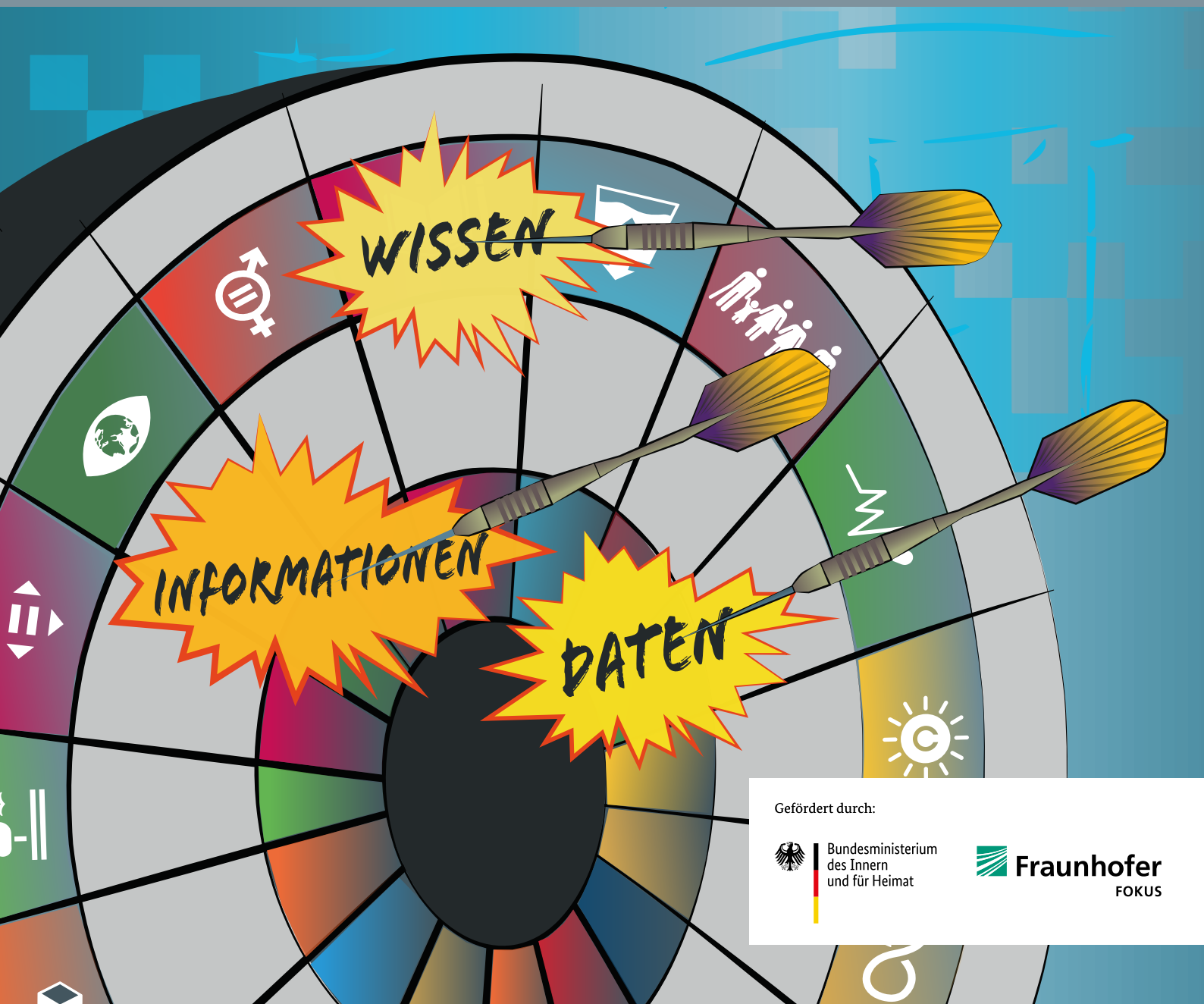


Kompetenzzentrum  
Öffentliche IT

FORSCHUNG FÜR DEN DIGITALEN STAAT

# DIE LOGIK DER DATEN NUTZEN: FORTSCHRITTLICHE DATENSTRATEGIEN ENTWICKELN

Leonard Mack



Gefördert durch:



Bundesministerium  
des Innern  
und für Heimat



Fraunhofer  
FOKUS

# IMPRESSUM

**Autor:**

Leonard Mack

**Gestaltung:**

Nico Duffke, Reiko Kammer

**Illustration:**

Martha Friedrich

**Herausgeber:**

Kompetenzzentrum Öffentliche IT  
Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS  
Kaiserin-Augusta-Allee 31, 10589 Berlin  
Telefon: +49-30-3463-7173  
Telefax: +49-30-3463-99-7173  
info@oeffentliche-it.de  
www.oeffentliche-it.de  
www.fokus.fraunhofer.de

ISBN: 978-3-948582-09-8

1. Auflage September 2022

Dieses Werk steht unter einer Creative Commons Namensnennung 3.0 Deutschland (CC BY 3.0) Lizenz. Es ist erlaubt, das Werk bzw. den Inhalt zu vervielfältigen, zu verbreiten und öffentlich zugänglich zu machen, Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anzufertigen sowie das Werk kommerziell zu nutzen. Bedingung für die Nutzung ist die Angabe der Namen der Autor:innen sowie des Herausgebers.

Logos und vergleichbare Zeichen dürfen nur im Kontext des Werkes genutzt und nicht abgewandelt werden.

Von uns verwendete Zitate unterliegen den für die Quelle geltenden urheberrechtlichen Regelungen.

# VORWORT

Die Corona-Pandemie und der ihretwegen verschobene Zensus zeigen gleichermaßen, wie sich die Einschätzung von Daten und ihre Nutzung im Laufe der Zeit gewandelt haben und weiter wandeln. Ohne die internationale Verfügbarkeit von Daten und ihre zielgerichtete Auswertung hätten keine Impfstoffe in Rekordzeit entwickelt, keine Lageberichte aktuell gehalten und keine Maßnahmen begründet werden können. Gerade dass dabei die eine oder andere Datenlücke offenbar wurde, verweist auf die wachsende Bedeutung von Datenauswertungen für die Politikgestaltung und das Verstehen gesellschaftlicher Dynamiken. Dass sich beim Zensus, ganz anders als noch in den 1980er Jahren, kaum Boykottaufrufe vernehmen lassen, deutet zugleich auf einen neuen Umgang mit Daten in weiten Teilen der Bevölkerung.

Bei all diesen Entwicklungen wird die Diskussion um Daten oft unter den Vorzeichen ethischer Bedenken gegen die Datennutzung geführt. So wichtig und wertvoll der Schutz der informationellen Selbstbestimmung und des geistigen Eigentums ist, so bemerkenswert erscheint doch, dass die ethischen Fragen der Nichtnutzung von Daten in den Diskussionen kaum eine Rolle spielen. Dieses White Paper nimmt dazu eine explizit andere Sichtweise ein.

Datenstrategien werden hier als Datennutzungsstrategien verstanden. Ziel jeder Datenstrategie öffentlicher Stellen muss es demnach sein, die Nutzung von Daten zu erleichtern, um so Innovationen zu ermöglichen. Entsprechend fokussieren die Ausführungen gerade nicht auf mögliche Restriktionen und Risiken, sondern spannen einen Bogen von der Datennutzung bei der menschlichen Wahrnehmung über die Bedeutung von Daten als Produktionsfaktoren und Innovationstreiber bis hin zum normativen Gebot, die eigenen Daten möglichst Vielen zugänglich zu machen, um so neue Lösungen zu ermöglichen.

Die Argumentation entfaltet dabei eine stark wirtschaftsliberale Sichtweise, die Daten, Datennutzung und in der Folge auch Datenmärkte als natürlich ansieht. Durch die Entfesselung solcher Datenmärkte, so das Postulat, treibt der Wettbewerb die Innovationsaktivitäten der Datennutzenden. Verbleiben die Daten demgegenüber weiter in ihren Silos, werden Innovationstätigkeiten

und damit Produktivitätssteigerungen verhindert. Der Raum zwischen den Silos eröffnet neue Möglichkeiten, die dank der Besonderheiten von Daten, insbesondere ihrer Nicht-Rivalität im Konsum, von Vielen genutzt werden können.

Aus der wirtschaftsliberalen Perspektive entfaltet sich ein neuer Blick auf die Ausrichtung von Datenstrategien. Als Datennutzungsstrategien werfen sie die Frage auf, wie die neu entstehenden Datenmärkte zum Vorteil möglichst vieler ausgestaltet werden sollten, ohne die berechtigten Schutzinteressen zu verletzen. Dies betrifft umso mehr die soziale Konstruktion freier Datenmärkte, für die in diesem Papier plädiert wird. Die Entwicklung von Datennutzungsstrategien wird so zu einem Aufruf, zu neuen Datenerfahrungen aufzubrechen – und so mit der Entwicklung von Datenstrategien einen politischen Unterschied zu markieren, der einen gesellschaftlichen und wirtschaftlichen Unterschied nach sich zieht. Dass bei einem so breiten und anspruchsvollen Thema nicht alle Aspekte in gleicher Intensität beleuchtet werden können, bedarf keiner Erwähnung: Damit die Überlegungen praktische Wirkung entfalten können, erfordert beispielsweise Datenkompetenzen auf allen Ebenen, eine tragfähige Datenethik, das entsprechende Datenmanagement und organisationale Flexibilität.

Für den Debattenimpuls in Form dieses White Papers möchten wir uns bei dem Autor bedanken. Dem besonderen Dank des Autors an Prof. Dr. Viktor Mayer-Schönberger vom Oxford Internet Institute, der die hier formulierten Gedanken maßgeblich geprägt und die Entstehung des White Papers begleitet hat, schließen wir uns ebenfalls gern an. Datenpolitisch interessierten Lesenden empfehlen wir zudem Ausgabe 18 des ÖFIT-Podcasts mit Prof. Mayer-Schönberger.

Wir wünschen eine anregende und aufrüttelnde Lektüre!

Ihr Kompetenzzentrum Öffentliche IT

ERST DURCH NUTZUNG

SCHAFFEN DATEN WERT.

## INHALTSVERZEICHNIS

	<b>Vorwort</b>	<b>3</b>
<b>1.</b>	<b>Thesen</b>	<b>5</b>
<b>2.</b>	<b>Auf dem Weg zur strategischen Datennutzung</b>	<b>7</b>
<b>3.</b>	<b>Vorgelagerte Annahmen im Kontext von Datenstrategien und -wirtschaft</b>	<b>10</b>
<b>4.</b>	<b>Thesendiskussion und Handlungsempfehlungen</b>	<b>19</b>
4.1	Menschen und Gesellschaften sind datengetrieben	19
4.2	Erst durch Nutzung schaffen Daten Wert	21
4.3	Daten nutzen sich nicht ab, lassen sich perfekt klonen und sind nicht-rivalisierend nutzbar	24
4.4	Innovationen benötigen Zugang zu mehr Daten	29
4.5	Daten benötigen eine gemeinsame Sprache	34
4.6	Datengetrieben zu sein, verändert Organisationen	40
<b>5.</b>	<b>Ausblick</b>	<b>46</b>



# 1. THESEN

## **Menschen und Gesellschaften sind datengetrieben.**

Das Verarbeiten von Daten ist alltäglicher Teil und sogar eine wesentliche Grundlage von Gesellschaften. Daten ermöglichen Arbeitsteilung in und die Koordination von Gesellschaften. Wer beispielsweise Handel betreibt, erfasst und bewertet notwendigerweise Daten, etwa um den Wert gehandelter Waren einzuschätzen und abzustimmen. Neu sind in Zeiten der Datenökonomie und datengetriebener Entscheidungen nicht Daten an sich, sondern die zunehmende *Konzentration von Datenbeständen*, deren sich daraus ergebende *oftmalige Unzugänglichkeit* sowie die *ungleiche Verteilung von Fähigkeiten und Ressourcen zur Datenverwertung*. Diese Hindernisfaktoren für die individuell und gesellschaftlich sinnvolle Nutzung von Daten abzuschwächen, sollte primäres Anliegen staatlicher Datenstrategien sein.

## **Erst durch Nutzung schaffen Daten Wert.**

Mehr als die Hälfte der globalen Datenbestände wird *nie* genutzt. Das reine Erheben und Horten von Daten an sich erzeugt aber keinen Wert. Wert entsteht nur, wenn Daten *genutzt* werden, sei es, um neue medizinische Therapien zu entwickeln, Geschäftsprozesse zu optimieren oder Ressourcenprobleme zu lösen. Datenstrategien sollten daher nicht das Sammeln von Daten, sondern primär die Nutzung von bereits bestehenden Datenbeständen in den Fokus nehmen.

## **Daten nutzen sich nicht ab, lassen sich perfekt klonen und sind nicht-rivalisierend nutzbar.**

Für die Förderung von Innovationen und Produktivität sind Daten eine ideale Ressource. Denn anders als andere Ressourcen lassen sie sich tatsächlich nachhaltig verteilen, sei es per Gießkanne als Open Data für alle oder als geteilte Daten für ausgewählte Nutzende. Daten können unendlich oft genutzt und perfekt kopiert werden, ohne dass sie sich abnutzen oder ihre Qualität anderweitig leidet. Anders als materielle Ressourcen lassen sich Daten damit nicht-rivalisierend nutzen. Rivalität kann allenfalls zwischen konkurrierenden Geschäftsmodellen, Innovationen und Algorithmen aufkommen. Damit entsteht Wettbewerb. Der ist zwar für die beteiligten Akteure unbequem, für ein nachhaltiges Innovations-Ökosystem aber umso förderlicher.

## **Innovationen benötigen Zugang zu mehr Daten.**

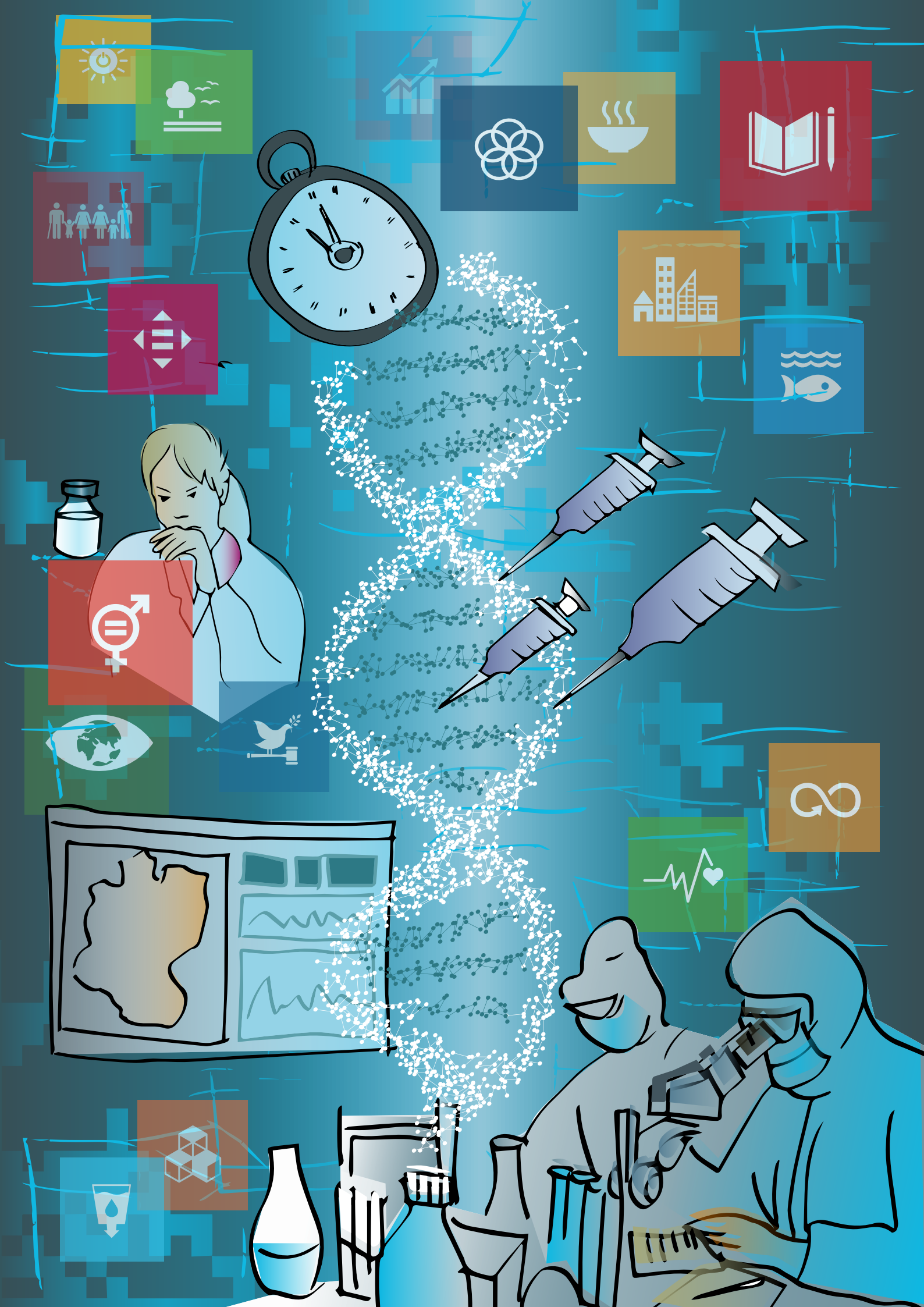
Ökonomisch betrachtet sind die Treiber von Innovationen heute Arbeit, Kapital *und* Daten. Gerade die begrenzte Verfügbarkeit von Daten ist aber für zahlreiche Innovatoren, ob als Unternehmen oder bürgerschaftliche Initiativen, ein zentrales Hindernis. Zwar profitieren einige wenige datenreiche Unternehmen von fast unendlich tiefen Datentöpfen, um scheinbar in rasendem Tempo immer neue Innovationen hervorzubringen. Die Produktivitätszuwächse führender Industrienationen bleiben aber trotz dieser Ausreißer-Unternehmen seit etwa 2005 hinter historischen Wachstumsraten zurück. Datenzugangs- und Teilungspflichtigen können Innovatoren nicht nur Zugang zur Innovationsressource Daten verschaffen, sondern auch die Halter großer Datenbestände dazu anspornen, diese noch besser auszuschöpfen.

## **Daten benötigen eine gemeinsame Sprache.**

Damit Daten besser nutzbar werden, bauen verschiedene Akteure immer öfter Plattformen für das Veröffentlichen und Teilen von Daten auf. Dabei stehen nicht selten sektorspezifische Lösungen im Vordergrund, die jedoch aufgrund mangelnder Interoperabilität eine Datennutzung über die bestehenden Sektorengrenzen kaum erleichtern. Datenstrategien sollten im Auge behalten, dass bahnbrechende Innovationen nicht selten von sektorfremden Akteuren kommen. Diesen Akteuren sollten Datenstrategien daher den Zugang erleichtern, vor allem durch domänenübergreifende und interoperable Standards, Schnittstellen, Taxonomien und Ontologien.

## **Datengetrieben zu sein, verändert Organisationen.**

Datengetriebene Organisationen unterscheiden sich erheblich von traditionelleren Organisationen, die den Zugang zu Daten aus verschiedenen Gründen funktional und hierarchisch konzentrierten. Die konkreten Organisationsformen zu finden, die eine hocheffektive und -effiziente Verwertung von Daten ermöglichen, stellt Organisationen vor erhebliche Herausforderungen. Schnelligkeit, Offenheit und Agilität erscheinen immer mehr als Instrumente, um diesen Herausforderungen zu begegnen. Datenstrategien sollten daher explorative Förderinstrumente für die datengetriebene Transformation von Organisationen bereitstellen.



## 2. AUF DEM WEG ZUR STRATEGISCHEN DATENNUTZUNG

Ein wachsendes Interesse an der Rolle von Daten für eine fortschrittliche Wirtschaft, Politik und Verwaltung begleitet die öffentliche Debatte in Deutschland und Europa seit fast einer Dekade. Konzepte wie Big Data, Industrie 4.0 und Smart Cities versprechen durch Daten verbesserte Anwendungen, Produkte und Verwaltungsservices. Gemessen an den mannigfaltig dokumentierten Ansprüchen erscheint dabei keine Herausforderung zu komplex: 2015 legten die Vereinten Nationen in ihrem Global Sustainability Development Report dar, wie Daten bei der Erreichung der Sustainable Development Goals helfen können – von der Identifizierung von Diskriminierung per Natural Language Processing bis zum Aufdecken illegaler Abholzungen per Satellitendaten.<sup>1</sup> Dass die UN mit ihrer implizit vertretenen These zur Bedeutung von Daten für das Wohl und Wehe der globalen Gesellschaft kaum übertrieben haben, verdeutlicht seit Ende 2019 die Coronapandemie.

Zwei Jahre nach Ausbruch der Pandemie ist die datengetriebene Produktion wissenschaftlicher Evidenz und deren Anwendung für die Weiterentwicklung des medizinischen Instrumentenkastens die global einzig erfolgreiche Strategie zur Virusbekämpfung. Nirgends wird dabei die Bedeutung hürdenfreier Datenzugänge, internationaler Kollaboration und Wettbewerb deutlicher als bei der Impfstoffentwicklung. Ein kurzer Rückblick: Nachdem Patient 0 am 8. Dezember 2019 in Wuhan registriert worden war, veranlassten die Ärzte am 25. Dezember eine Genomsequenzierung des noch unbekanntes Virus. Gerade einmal zweieinhalb Wochen später, am 10. Januar 2020, hatte ein chinesisch-australisches Konsortium die Genomsequenz entschlüsselt. Die chinesischen Behörden reichten die entsprechenden Daten am 11. Januar 2020 zur Publikation an die WHO<sup>2</sup> weiter. Öffentlich zugänglich sind diese seitdem unter anderem auf der Open-Access-Plattform virological.org<sup>3</sup> sowie auf der offenen GenBank-Datenbank<sup>4</sup>, die von der US-amerikanischen National Library of Medicine betrieben

wird. Auf dieser Grundlage entwickelte ein US-amerikanisches Biotech-Unternehmen innerhalb von 72 Stunden die Sequenz für den ersten Impfstoff, dessen Produktion für die ersten Versuchsreihen am 7. Februar begann. Weitere Hersteller folgten. In kurzer Taktfolge präsentierten drei Kandidaten im November 2020 die Ergebnisse der Stufe 3 der klinischen Prüfungen und beantragten weltweit Impfstoffzulassungen. Dass diese Entwicklung nicht die oftmals üblichen 10 Jahre, sondern nur knapp 10 Monate dauerte, ist unzweifelhaft Resultat außergewöhnlicher Dringlichkeiten. Der Erfolg liegt aber auch daran, dass Daten zu Genomsequenzen schnell und vor allem ungehindert verfügbar waren<sup>5</sup>, parallel von konkurrierenden Nutzenden verwertet werden konnten und damit ein gleichsam kollaboratives und kompetitives Innovationssystem entstand.<sup>6</sup>

Die Coronakrise bewirkte damit aus datenökonomischer Perspektive die Beschleunigung einer Lernkurve, auf die Politik- und Wirtschaftsstrategen bereits zuvor eingebogen waren. Anfänglich beschäftigten sich datenstrategische Initiativen lange mit der *Bereitstellung* von Daten als Schlüssel zum Erfolg: Zahlreiche Open-Government-Data-Portale auf verschiedenen Verwaltungsebenen sowie nationale und transnationale Referenzportale wie das deutsche Govdata und das Europäische Datenportal traten primär mit dem Ziel an, Daten vermehrt und einfacher

<sup>5</sup> Neben Genomsequenzdatenbanken wie der der National Library of Medicine und der des European Molecular Biology Laboratory (EMBL; <https://www.ebi.ac.uk/>) wurden dazu unter anderem auch spezifische Portale wie das Covid-19-Portal der Europäischen Kommission (<https://www.covid19dataportal.org/>) geschaffen, um Genomsequenzen und Versuchsdaten schnell zu veröffentlichen.

<sup>6</sup> So sehr dieses Beispiel den im Idealfall extrem positiven Nutzen von Daten und kompetitiver Datennutzung illustriert, hat die Erfolgsgeschichte aber auch eine Schattenseite: In einem Report für die WHO kritisierte das Independent Panel for Pandemic Preparedness and Response Anfang 2021, dass Behörden weltweit schneller Eindämmungsmaßnahmen basierend auf vorliegenden Frühindikatoren und ersten Infektionsdaten hätten ergreifen können. Im Vergleich zu anderen Krisen erscheint der zeitliche Abstand zwischen Datenerhebung, Datenanalyse und Verwertung in Form politischer Entscheidungen sehr schnell. Gemessen an der Dynamik der Corona-Pandemie trübt sich dieser Eindruck aber bei kritischer Betrachtung. Zahlreiche öffentliche Debatten, sei es um zu spät beschlossene Kontaktbeschränkungen, zahme Klimaschutzmaßnahmen oder ausbleibende Rentenreformen im demografischen Wandel verdeutlichen dabei: Wie sehr Daten im konkreten Fall Wirksamkeit entfalten können, hängt erheblich von der Entscheidungsfreude relevanter Akteure im Angesicht kritischer Herausforderungen ab. Das objektive Potenzial von Daten als Evidenzmittel, Innovationstreiber und Entscheidungsstütze wird dadurch aber nicht geschmälert. The Independent Panel for Pandemic Preparedness and Response: »Second report on progress«. Januar 2021. [https://theindependentpanel.org/wp-content/uploads/2021/01/Independent-Panel\\_Second-Report-on-Progress\\_Final-15-Jan-2021.pdf](https://theindependentpanel.org/wp-content/uploads/2021/01/Independent-Panel_Second-Report-on-Progress_Final-15-Jan-2021.pdf).

<sup>1</sup> Vereinte Nationen (2015): »Global Sustainable Development Report«. 2015 Edition. Advance Unedited Version. <https://sustainabledevelopment.un.org/content/documents/1758GSDR%202015%20Advance%20Unedited%20Version.pdf>.

<sup>2</sup> WHO: World Health Organization, Weltgesundheitsorganisation.

<sup>3</sup> <https://virological.org/t/novel-2019-coronavirus-genome/319>.

<sup>4</sup> [https://www.ncbi.nlm.nih.gov/nuccore/NC\\_045512.2](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2) und <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

DIE ERWARTUNG IST, DASS EINE  
INNOVATIVE DATENNUTZUNG MENSCHEN,  
GEMEINWESEN UND WIRTSCHAFT GÄNZLICH  
NEUE MÖGLICHKEITEN ERÖFFNEN WIRD,  
FÜR ZAHLREICHE HERAUSFORDERUNGEN  
BISHER UNENTDECKTE LÖSUNGEN ZU ERSCHLIESSEN.

verfügbar zu machen. Graduell wurde in den Folgejahren der angebotsgetriebene Fokus erweitert und durch nachfrageorientierte Instrumente ergänzt. Erkennbar ist das in der Konzeption und Umsetzung von Datenräumen von der kommunalen bis zur europäischen Ebene. Besonders im Kontext der Europäischen Datenstrategie sind sogenannte Europäische Datenräume zentrale Infrastrukturen, um primär Unternehmen das Teilen von Daten unter rechtlich und technisch kontrollierten, meist sektorspezifischen Rahmenbedingungen zu ermöglichen. Darüber hinaus sollen Europäische Datenräume aber auch technische Lösungen bereitstellen, die eine anwendungsorientierte Nutzung und Verwertung von geteilten Datenbeständen unterstützen.<sup>7</sup> In ähnlicher Form soll die European Open Science Cloud (EOSC) im Wissenschaftssektor durch Speicher- und Rechnerkapazitäten dazu beitragen, die Nutzbarkeit von Forschungsdaten EU-weit drastisch zu erhöhen.<sup>8</sup> Auch der Aufbau einer Nationalen Forschungsdateninfrastruktur (NFDI) in Deutschland verfolgt vergleichbare Zielsetzungen.<sup>9</sup>

Datenstrategien sollen diese Entwicklungen zusammenführen und koordinieren, unabhängig davon, ob es sich dabei um staatliche Strategien oder um solche einzelner privater Organisationen handelt. Dabei definieren diese Strategien in der Regel Rahmenbedingungen für den Umgang mit und die Nutzung von Daten über deren gesamten Lebenszyklus, von der Bereitstellung über die Verwertung bis hin zur etwaigen erneuten Bereitstellung. Dass unter anderem die EU<sup>10</sup>, Deutschland<sup>11</sup> und

Großbritannien<sup>12</sup> ihre Datenstrategien in dichter Folge während der Coronapandemie veröffentlicht haben, ist zwar eher durch Zufall denn durch Kausalität begründet. Sie folgen dabei dem Vorbild der Vereinigten Staaten<sup>13</sup>, die bereits 2017 auf Bundesebene eine Datenstrategie erarbeiteten. Die Pandemie dürfte aber dennoch die strategische Relevanz eines nachhaltig innovativen Umgangs mit Daten als (inter-)nationale Priorität in den Köpfen zahlreicher Entscheider:innen eingebrannt haben. Der Umgang mit den neu gewonnenen Analysefähigkeiten, mit den gelernten Lektionen in evidenzbasierter Politikgestaltung und letztlich auch mit den zur Pandemieeindämmung mobilisierten Datenressourcen ist zweifelsohne eine Herausforderung. Datenstrategien sind Instrumente, um sich dieser Herausforderung als Fortschrittschance zu widmen.

Dabei stellt sich eine grundlegende Frage: Auf welchen Grundlagen sollten Datenstrategien aufbauen, damit diese richtig gedacht und gemacht werden? Hinter dieser Frage stehen eine Erwartung und eine These. Die Erwartung ist, dass eine innovative Datennutzung Menschen, Gemeinwesen und Wirtschaft gänzlich neue Möglichkeiten eröffnen wird, für zahlreiche Herausforderungen bisher unentdeckte Lösungen zu erschließen. Die These ist, dass, um diese Potenziale nutzbar zu machen, Entscheider:innen von einigen zentralen Grundannahmen ausgehen sollten. Diese dienen als Grundlage für tatsächlich fortschrittliche Datenstrategien, welche Potenziale richtig erkennen, priorisieren und umsetzen. Dieses White Paper diskutiert dazu sechs Thesen und umreißt entsprechende Handlungsempfehlungen. Das Papier richtet sich vorwiegend an Entscheidungsträger:innen und Analyst:innen aus Politik, Verwaltung und Unternehmen, die eine Datenstrategie entwickeln möchten.

<sup>7</sup> Europäische Kommission (2022): »Europäische Datenstrategie«. [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy\\_de](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_de).

<sup>8</sup> <https://eosc-portal.eu/>.

<sup>9</sup> <https://www.nfdi.de/>.

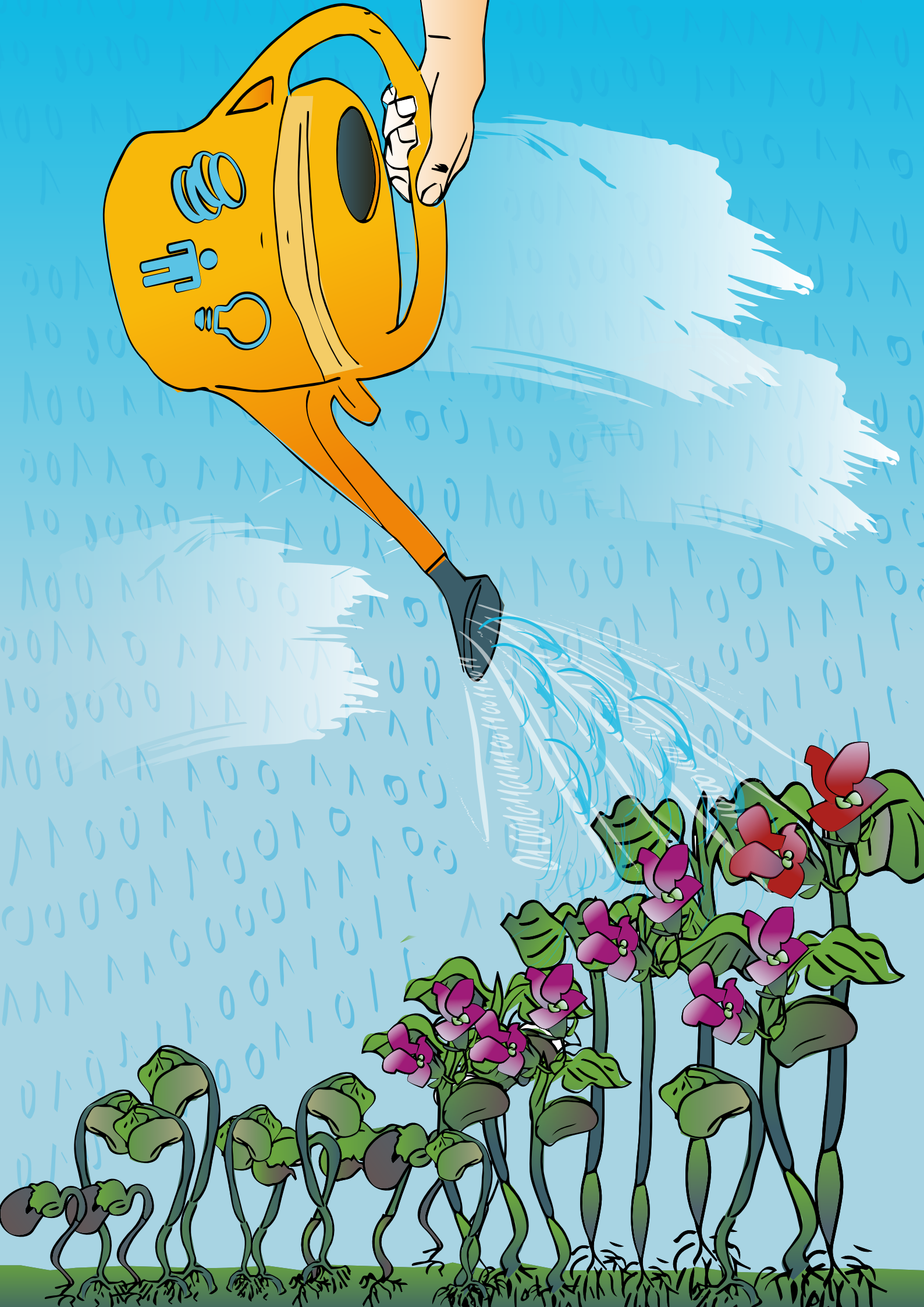
<sup>10</sup> Europäische Kommission (2022): »Europäische Datenstrategie«. [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy\\_de](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_de).

<sup>11</sup> Die Bundesregierung (2021): »Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum«. <https://www.bundesregierung.de/breg-de/suche/datenstrategie-der-bundesregierung-1845632>.

<sup>12</sup> UK Department for Digital, Culture, Media & Sport (2019): »National Data Strategy«. <https://www.gov.uk/guidance/national-data-strategy>.

<sup>13</sup> U.S. Office of Management and Budget: »Federal Data Strategy – Data, accountability, and transparency: creating a data strategy and infrastructure for the future«. <https://strategy.data.gov/>.





# 3. VORGELAGERTE ANNAHMEN IM KONTEXT VON DATENSTRATEGIEN UND -WIRTSCHAFT

Wie Daten strategisch genutzt werden sollen, erscheint nicht wenigen Entscheidungsträger:innen als eine komplizierte Frage. Bei allen notwendigen Abwägungen und Kontextualisierungen müssen strategische Antworten auf diese Frage aber nicht unbedingt hochkomplex sein. Denn längst nicht alle Ansätze, auf denen fortschrittliche Datenstrategien aufbauen sollten, sind notwendigerweise wirklich neu und bisher unbekannt. Erschwert werden Diskussionen vielmehr durch ungeeignete Analogien und falsche Fährten, die Diskussionen über »die Daten« in ähnlicher Form erschweren, wie frühere Debatten über »das Internet« und »die sozialen Medien«.

Um die im weiteren Verlauf dieses Papiers vertieften Thesen besser einzuordnen, ist es daher notwendig, vorab einige vorgelagerte Annahmen zu erläutern. Diese Annahmen bilden die Grundlage für die folgende Diskussion der Thesen und grenzen zugleich den Gegenstand dieses Papiers ein.

## Warum Datenstrategien politisch sind

Erstens geht dieses Papier davon aus, dass zahlreiche Fragen, die sich im Kontext von Datenstrategien stellen, letztlich nur durch politische Priorisierungen und Abwägungen zu beantworten sind. Im Sinne des US-amerikanischen Juristen Lawrence Lessig sind technische Rahmenbedingungen und technische Data Governance architekturelle Variablen, die in einem Politikfeld wie der Datenpolitik regulierende Wirkung haben.<sup>14</sup> Die Möglichkeiten und Unmöglichkeiten von beispielsweise Datenmodellen, Speichertechnologien und Datenverarbeitungssystemen beeinflussen in diesem Sinne natürlich, ob und wie komplex politische Handlungsoptionen umsetzbar sind. Sie determinieren entsprechende Entscheidungen aber nicht. Wie sehr beispielsweise eine Datenstrategie den Zugang zu und die Nutzung von bestimmten Daten vereinfachen kann, ist primär das Ergebnis einer (wirtschafts-)politischen Abwägung darüber,

wie offen Datenzugänge sein sollen<sup>15</sup>, wie viele Marktteilnehmer es als Datennutzende geben soll und in welchem Wettbewerb diese miteinander stehen sollen. Im datenwirtschaftlichen Kontext werden diese Fragestellungen immer wieder diskutiert, derzeit vor allem im Kontext der in der Europäischen Datenstrategie eingeführten Datenintermediäre. Bei den genannten Fragen handelt es sich erkennbar um originär politische Fragestellungen. Diese können nicht in der Hoffnung darauf an Entwickler:innen delegiert – oder outgesourct – werden, dass sich politische Differenzen durch bestimmte Technik- oder Architekturentscheidungen auf technischer Ebene wegimplementieren lassen. Derart magische Kräfte besitzen weder Technologie noch vermeintliche Datenmagier im Gewand von Daten- und Technologieexpert:innen aller Art. Für Entscheidung:innen ist es stattdessen wichtiger, zu verstehen, wann Technologie im Bereich der Datenpolitik *nicht* primär dazu beitragen kann, Probleme zu lösen. Dabei geht es darum, besser als bisher zu erkennen, in welchen Fällen beispielsweise wettbewerbs- oder innovationspolitische Entscheidungen geeigneter zur Problemlösung sind als technologiepolitische.<sup>16</sup> Dies dürfte deutlich öfter der Fall sein, als es viele rein technisch oder technikrechtlich geführte Debatten und Argumente um Datenstrategien, Datenpolitik und Datenteilung bzw. -öffnung vermuten lassen. Daher diskutieren die hier präsentierten Thesen die Grundlagen von Datenstrategien stärker aus einem (wirtschafts-)politischen Blickwinkel, als das technisch daher kommende Thema »Daten« es zunächst vermuten lassen mag.

<sup>14</sup> Lawrence Lessig (2006): »Code: Version 2.0«. [https://upload.wikimedia.org/wikipedia/commons/f/fd/Code\\_v2.pdf](https://upload.wikimedia.org/wikipedia/commons/f/fd/Code_v2.pdf).

<sup>15</sup> Der Grad der Offenheit bemisst sich dabei an der technischen Zugänglichkeit von Daten (d. h. der Art der Bereitstellung – beispielsweise über Programmierschnittstellen – und deren Niedrigschwelligkeit) sowie deren (Weiter-)Nutzungsmöglichkeiten, die durch Lizenzbedingungen mehr oder weniger stark reglementiert werden. Zum besseren Verständnis der unterschiedlichen Offenheitsgrade von Daten wird auf das »Data Spectrum« des Open Data Institute verwiesen: <https://theodi.org/about-the-odi/the-data-spectrum/>.

<sup>16</sup> Michael Veale (2019): »A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence«. [https://discovery.ucl.ac.uk/id/eprint/10084302/7/Veale\\_HLEG\\_preprint\\_revised.pdf](https://discovery.ucl.ac.uk/id/eprint/10084302/7/Veale_HLEG_preprint_revised.pdf).

### Welche Daten strategische Aufmerksamkeit verdienen

Zweitens stellt sich die Frage, um welche »Daten« es im Kontext von Datenstrategien nach dem Verständnis dieses Papiers gehen soll. Wenn die Nutzung von Daten durch Datenstrategien, wie nicht selten propagiert, wirklich erheblich gesteigert werden soll, führt in letzter Konsequenz kaum ein Weg daran vorbei, auch den Nutzendenkreis substanziell zu erweitern. Damit steht eine teilweise Öffnung oder ein verstärktes Teilen von Datenbeständen im Raum, die bisher geschlossen waren und nur wenigen Nutzenden innerhalb einer Organisation zur Verfügung standen. Geeignete Daten müssten also zugänglich gemacht werden – entweder durch anreizgetriebene Freiwilligkeit oder aber durch Datenteilungspflichten. Derartige Vorhaben wecken natürlich Begehrlichkeiten, Sensibilitäten und Befürchtungen, vor allem im Hinblick auf die Wahrung von Datenschutz, geistigem Eigentum und Geschäftsinteressen. Entscheidend sind aber aus Sicht dieses Papiers politische Abwägungen zwischen Innovations- und Wettbewerbsförderung, weiteren politischen Zielsetzungen (beispielsweise in der Umwelt-, Wohnungs- und Sozialpolitik), Individualrechten sowie Geschäftsinteressen und -geheimnissen. Dabei wird in diesem Papier die Auffassung vertreten, dass bisherige Debatten zur freiwilligen oder mandatierten Teilung oder Öffnung von Daten – sowohl solche von Verwaltungen als auch solche von Unternehmen – zu zaghafte verlaufen.

Zunächst sei vor diesem Hintergrund gesagt: Wenn im Weiteren von »Daten« gesprochen wird, sind damit in der Regel nicht-personenbezogene und nicht-geschäftsgeheimnisrelevante Daten gemeint. Ziele einer von Datenstrategien zu fördernden nachhaltigen und innovativen Datennutzung können nicht Verletzungen der Privatheit von Einzelpersonen oder die Erosion von berechtigten Geschäftsgeheimnissen sein. Grundsätzlich ist aber auch die Frage danach, bei welchen spezifischen Bedingungen, Bereichen und Anwendungsfällen eine Nutzung personenbezogener Daten erlaubt sein sollte, eine

genuin politische Frage. So erlaubt die DSGVO<sup>17</sup> in Artikel 23 den Mitgliedstaaten, unter bestimmten Bedingungen<sup>18</sup> abweichende Datenschutzregelungen zu treffen und eröffnet in Kapitel 9 (Artikel 85 – 91) Möglichkeiten zur Datenverarbeitung in spezifischen Situationen. Für im öffentlichen Interesse liegende Archivzwecke, zu wissenschaftlichen oder historischen Forschungszwecken und zu statistischen Zwecken beinhaltet dies sogar die Verarbeitung sensibler Daten. Der Ball wird in der DSGVO also bewusst und mit weitreichenden Handlungsmöglichkeiten versehen an die nationalen Gesetzgeber zurückgespielt. Eine erweiterte Nutzung, Teilung und gegebenenfalls sogar Öffnung auch personenbezogener Daten ist damit eine politische Möglichkeit.

Ähnlich sollten auch die zunehmende Öffnung und das Teilen von Daten behandelt werden, die Geschäftsinteressen von Unternehmen betreffen. Nicht selten wird dagegen eine Mischung aus urheber-, datenbank- oder geschäftsgeheimnisrechtlichen Argumenten vorgebracht. Aus Sicht dieses Papiers sollte es im Rahmen von Datenstrategien nicht darum gehen, geistiges Eigentum, beispielsweise an Daten oder Datenbanken, sofern die-

<sup>17</sup> DSGVO, Datenschutzgrundverordnung: »Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)«. <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32016R0679>.

<sup>18</sup> Zentral ist dabei laut Art. 23 DSGVO vor allem die Bedingung, dass abweichende Regelungen grundlegende Rechte und Freiheiten respektieren und es sich um in einer demokratischen Gesellschaft erforderliche und verhältnismäßige Maßnahmen handelt.

BEI DER VERARBEITUNG SENSIBLER  
DATEN SPIELT DIE DSGVO BEWUSST  
DEN BALL AN DIE NATIONALEN  
GESETZGEBER ZURÜCK – UND DAS  
MIT WEITREICHENDEN  
HANDLUNGSMÖGLICHKEITEN.

se tatsächlich eine ausreichende Schöpfungshöhe aufweisen<sup>19</sup>, zu erodieren oder z. B. eine generelle Aussetzung von Patenten für datengetriebene Innovationen voranzutreiben. Ebenso

<sup>19</sup> Nach dem aktuellen Stand der europäischen Gesetzgebung und Rechtsprechung kann jedoch nur für solche Daten ein urheberrechtlicher Anspruch reklamiert werden, in denen eine persönliche Schöpfung des Schaffenden oder die geistige Höhe im Werk erkennbar ist. Für reine Mess- und Beobachtungs(roh)daten, Metadaten oder sonstige Fakten wird dies verneint. Konkret betroffen sind davon regelmäßig geografische Daten, Statistikdaten, Verwaltungsregister, Umwelt-, Wirtschafts- und Sozialdaten, aber auch Produktbeschreibungen und Komponentenlisten. Fraglich kann aber – vor allem für in Datenbanken aggregierte Datenbestände – sein, ob Daten im Sinne der europäischen Datenbankrichtlinie geschützt sind. Diese schützt Daten auch ohne urheberrechtlichen Schöpfungsanspruch, schon wenn allein die Datenbank vom Erstellenden substantielle Investitionen für deren Erstellung, Verifizierung und Bereitstellung erfordert. Besonders für aggregierte, verarbeitete oder anders aufbereitete Daten kann dies relevant sein. Allerdings hat der Europäische Gerichtshof in einer Reihe von Urteilen den Wirkungsbereich der Datenbankrichtlinie insofern eingegrenzt, dass deren Schutzrechte nicht auf Datenbanken anwendbar sind, die lediglich als Nebenprodukt der Haupttätigkeiten einer Organisation entstehen. Davon ausgehend lässt sich die Auffassung vertreten, dass die Datenbankrichtlinie keine Datenbanken schützt, die Verwaltungen oder Unternehmen zu Administrationszwecken unterhalten oder um die Funktionalität ihrer Produkte oder Dienstleistungen zu gewährleisten. Davon betroffen wären möglicherweise zahlreiche Registerdaten von Verwaltungen oder automatisch generierte Daten, beispielsweise von IoT-Geräten.

Forschungsdaten.info (2021): »Urheberrecht«.

<https://www.forschungsdaten.info/themen/rechte-und-pflichten/urheberrecht/>.

Kanzlei Plutte (2022): »Rechtlicher Schutz von Datenbanken: Alles Wichtige im Überblick«. <https://www.ra-plutte.de/rechtlicher-schutz-von-datenbanken-alles-wichtige-im-ueberblick/#datenbank-persoennliche-geistige-schoepfung>.

Hans Graux (2021): »The peculiar case of data sharing: are intellectual property rights relevant to begin with?«. [https://eudatasharing.eu/news/peculiar-case-data-sharing-are-intellectual-property-rights-relevant-begin#footnote4\\_j10oit6](https://eudatasharing.eu/news/peculiar-case-data-sharing-are-intellectual-property-rights-relevant-begin#footnote4_j10oit6).

Court of Justice of the European Union: »Judgment of the Court (Grand Chamber) of 9 November 2004. Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE (OPAP)«. Case C-444/02.

European Court Reports 2004 I-10549. ECLI:EU:C:2004:697. CELEX Number 62002CJ0444. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62002CJ0444>.

Court of Justice of the European Union: »Judgment of the Court (Grand Chamber) of 9 November 2004. The British Horseracing Board Ltd and Others v William Hill Organization Ltd.«. Case C-203/02. European Court Reports 2004 I-10415. ECLI:EU:C:2004:695. CELEX Number 62002CJ0203. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62002CJ0203>.

Court of Justice of the European Union: »Opinion of Advocate General Stix-Hackl delivered on 8 June 2004. Fixtures Marketing Ltd v Svenska Spel AB.«. Case C-338/02. European Court Reports 2004 I-10497. ECLI:EU:C:2004:338. CELEX Number 62002CC0338. <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:62002CC0338>.

Datenbankrichtlinie: »Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11. März 1996 über den rechtlichen Schutz von Datenbanken«. <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A31996L0009>.

sollten Unternehmensdaten geschlossen bleiben dürfen, die zweifelsfrei Geschäftsgeheimnisse betreffen und von den Unternehmen auch tatsächlich als solche behandelt werden. Dieses Recht sollte auch nicht dadurch angreifbar werden, dass einige Unternehmen aus freien strategischen Erwägungen teils sensible Daten mit ausgewählten Geschäftspartnern teilen. Dazu gehören beispielsweise Kundendateien, detaillierte Finanz- und Buchungsdaten sowie Logistik- und Beschaffungsdaten.<sup>20</sup>

Anders verhält es sich jedoch in Fällen, in denen marktbeherrschende Akteure Daten verschlossen halten, um den Markteintritt von Konkurrenten zu verhindern. Neben dem Ziel der Wettbewerbsförderung sind zudem weitere übergeordnete politische Interessen denkbar, die eine Öffnung bzw. Teilung von Unternehmensdaten trotz entgegenstehender Unternehmensinteressen vertretbar machen könnten. Aus Sicht dieses Papiers könnte dies z. B. die folgenden Fälle betreffen:

- Daten zu Produktsuchanfragen und historischen Preistrends bei Onlinehändlern und Marktplätzen, um die Marktteilhabe und den Wettbewerb gegenüber plattformartigen, marktbeherrschenden Akteuren zu fördern.
- Daten zu Fahrplänen, Ist-Zeiten und Routenauslastungen im öffentlichen Personennahverkehr und von Mobilitätsunternehmen, insbesondere im Falle durch Konzessionen begünstigter regionaler Monopolisten. Ziel kann hier beispielsweise sein, multimodale Mobilitätsangebote und Routenplanungen zu verbessern, um letztlich die Attraktivität des ÖPNV zu erhöhen. Spätestens zum Zeitpunkt der Neuvergabe von Konzessionen kann durch eine bessere Informationsversor-

<sup>20</sup> Interessant ist in diesem Zusammenhang, dass gerade Logistik- und Beschaffungsdaten im Industriekontext bei der Verwendung lieferantengesteuerter Bestandssysteme vielfach geteilt werden (s. dazu auch Abschnitt 4.6). Dabei handelt es sich allerdings um freie strategische Erwägungen der entsprechenden Unternehmen, woraus sich keine Mandatierung zur Datenöffnung oder -teilung ergeben sollte.



gung potenzieller Mitbewerber mehr Wettbewerb geschaffen werden.

- Daten zu Buchungsanfragen und (ggf. aggregierten) Vermietungsdaten von Kurzzeitvermietungsplattformen<sup>21</sup>, beispielsweise um lokale Tourismusangebote zu verbessern, um andere Kurzzeitvermieter wie etwa Hotels im Wettbewerb zu stärken oder um neue Instrumente zur Steuerung der Wohnungspolitik zu entwickeln.

Zusammengefasst führen diese Betrachtungen dazu, dass in diesem Papier davon ausgegangen wird, dass es grundsätzlich möglich wäre, deutlich mehr Daten zugänglich zu machen, als es heute der Fall ist. Auch wenn rechtliche oder technische Debatten dies teils anders suggerieren, ist die Frage, ob dies gewollt ist, eine politische.

### **Weshalb Wettbewerb wichtig ist**

Drittens setzen die Thesen folgende Annahme voraus: Wettbewerb ist der beste und wichtigste Schlüssel, um die innovativen Potenziale einer intensivierten Datennutzung zu erschließen. Ein in sämtlichen Dimensionen möglichst hürdenloser Zugang zu Daten ist dabei essenziell, um einen freieren Wettbewerb sicherzustellen.

In diesem Zusammenhang zeigte der Innovationsforscher Clayton M. Christensen bereits 1997 das zentrale Dilemma von Innovatoren auf.<sup>22</sup> Die Entwicklungsdynamiken so unterschiedlicher Industrien wie der Festplattenindustrie der 1980er und 1990er Jahre oder der Schaufelbaggerindustrie der 1950er bis

1960er Jahre teilten nach Christensens Erkenntnis ein Muster, das auch für die moderne Datenökonomie ein Lehrstück bietet: Kurz gesagt genügt anscheinend eine noch so gute Ressourcenausstattung nicht, um vor allem große oder marktbeherrschende Unternehmen dazu anzuregen, diese wirkungsvoll zur Entwicklung innovativer Produkte einzusetzen. Förderlich ist stattdessen schlichtweg gehöriger Wettbewerbsdruck.

Christensen beobachtete in seiner Studie, wie kleinere Wettbewerber regelmäßig marktbeherrschende Unternehmen mit innovativen, aber einfacheren Produkten herausforderten, damit disruptive Marktdynamiken beförderten und schließlich die Platzhirsche geradezu überfallartig ablösten. Das gelang den Außenseitern, obwohl diese zunächst nur einen Nischenmarkt bedienten und obgleich ihre Produkte nicht selten weniger performant waren als die der Platzhirsche. Den Raum für die Disruption von unten schafften die marktführenden Unternehmen selbst: Unter Fortführung ihres im Wesentlichen bekannten Produktportfolios und der bekannten Produktionsmethoden entwickelten sie zu lange immer hochwertigere, teurere, performantere Produkte vorwiegend für ihre profitabelsten und anspruchsvollsten Kunden. Damit übertrafen sie regelmäßig die Bedürfnisse von Kunden aus niedrigeren oder mittleren Marktsegmenten – und das war ihr Verhängnis. Denn sobald die zwar weniger leistungsfähigen, aber auch deutlich günstigeren Produkte der (noch) kleineren Mitbewerber gut genug waren, die Bedürfnisse dieser Kunden zu befriedigen, lief die Kundschaft in Scharen zur Konkurrenz über.

Das ist nicht nur eine Vorstellung, vor der sich die globale Verbrenner-Autoindustrie beim Umstieg auf den Elektromotor fürchtet. Aus datenökonomischer und innovationspolitischer Sicht gibt es noch eine weitere, tragische Lehre: Obwohl die Marktführer über mehr Geld verfügten, dichtere Vertriebsnetzwerke gestrickt hatten, mindestens ebenbürtige Entwicklungsabteilungen betrieben und direkten Zugriff auf deutlich größere

<sup>21</sup> Kurzzeitvermietungsplattformen sind Plattformen, über die meist private Nutzende Wohnraum für kurze Zeit, in der Regel als Ferienwohnung, anbieten und buchen können. Bekannte Beispiele sind AirBnB oder Booking.com; letztgenanntes vermittelt neben Hotels auch private Ferienwohnungen und Zimmer.

<sup>22</sup> Clayton M. Christensen (1997): »The innovator's dilemma. When new technologies cause great firms to fail«. Harvard Business School Press, (Boston).

WETTBEWERB DURCH FREIE

DATENNUTZUNG KANN ETABLIERTE

MARKTTEILNEHMER UND MARKTFÜHRER

ZU MEHR INNOVATIONEN ANTREIBEN.

Datenbestände<sup>23</sup> hatten, nutzten sie diese Ressourcen nicht, um radikale, auch produktkostensenkende Innovationssprünge zu unternehmen. Maßgeblich war dabei laut Christensen ironischerweise vor allem ihre Größe: Diese trieb Unternehmen dazu, ihr Geschäft primär in Massenmärkten sowie mit Produkten und Preisstrukturen zu suchen, die Umsätze und Gewinne in Größenordnungen versprachen, die für ihre eigenen, großen Unternehmen relevant waren. Weniger profitable Nischenmärkte wurden allenfalls halbherzig bedient und damit den kleineren Mitbewerbern überlassen – die damit ungestört Zeit hatten, die Leistungsfähigkeit ihrer Produkte zu steigern. Als diese Angebote dann massenmarktfähig waren, war es für die Noch-Marktführer fast immer längst zu spät zum Umsatteln.

Christensens Analysen erregten naturgemäß die Aufmerksamkeit von CEOs, Entrepreneuren und Unternehmensberater:innen weltweit. Zahlreiche Innovationslabore, Konzernausgründungen und Start-up-Kooperationen zeugen von den Bemühungen der letzten beide Jahrzehnte, die Innovationsfähigkeit von (Groß-)Unternehmen zu steigern und ihre Fähigkeiten als Disruptoren zu entwickeln. Experten in der Erfindung radikal neuer Produkte und der Eroberung und Entwicklung von Nischenmärkten sind viele Großunternehmen zumeist dennoch nicht geworden. Stattdessen versucht in letzter Zeit ironischerweise gerade die Speerspitze der vermeintlich hoch-innovativen großen Tech-Unternehmen, potenziell disruptive Konkurrenten durch eine aggressive Akquisepolitik auszuschalten. Aus Sicht von Kritiker:innen schaffen diese milliardenschweren Übernahmen nicht nur Raum für einen lukrativen Exit von Gründer:innen,

sondern auch eine »Kill-Zone« für Start-ups und deren Innovationen.<sup>24</sup> Um das eigene Produktportfolio zu erhalten, nutzen Tech-Großkonzerne nach dieser Lesart zunehmend die für Unternehmen am einfachsten zu mobilisierende Ressource: Geld.

Eher fraglich erscheint aus Sicht dieses Papiers daher der Versuch, die unternehmerische Datennutzung durch staatliche Maßnahmen anzuregen, die im Wesentlichen auf die Formel *mehr Geld für bessere und exklusivere Daten* hinausläuft. Christensens Analysen sprechen kaum dafür, dass dies vor allem Marktführern oder Monopolisten die tatsächlich erforderlichen Anreize bieten kann, eine disruptiv-innovative Datennutzung effektiv und mit Nachdruck zu verfolgen. Die Hoffnung, dass sich durch Exklusivrechte wie Dateneigentum, patentähnliche Datennutzungsprivilegien oder verschiedenartig faktisch gegen sektor- oder netzwerkfremde Wettbewerber geschützte Datenräume vornehmlich etablierte Akteure mit hohen Marktanteilen zu mehr grundlegender Innovation bewegen lassen, wäre demnach wohl vergeblich. Denn trotz anders lautender Beteuerungen liegt diese, ohne externen Wettbewerbsdruck, gar nicht in ihrem Interesse.

Aus innovationspolitischer und datenökonomischer Sicht sind fehlgelagerte Anreize zur Datennutzung das eigentliche Problem – und nicht zu wenig oder zu schlechte Daten, ineffektiv verteilte Informations- und Datenflüsse oder unklare Rechte an Daten. Zu versuchen, die innovationsdynamischen Defizite und fehlgelagerten (unternehmens-)internen Anreize von Marktakteuren durch die Gewährung zusätzlicher Privilegien sowie eine staatlich subventionierte oder sanktionierte Optimierung

<sup>23</sup> Beispielsweise in Form von Kundendateien oder Entwicklungs- und Produktionsdaten.

<sup>24</sup> Raghuram Rajan, Sai Krishna Kamepalli, Luigi Zingales (2020): »Killzone«. [https://www8.gsb.columbia.edu/faculty-research/sites/faculty-research/files/finance/Finance%20Seminar/kill\\_zone\\_Sep\\_13.pdf](https://www8.gsb.columbia.edu/faculty-research/sites/faculty-research/files/finance/Finance%20Seminar/kill_zone_Sep_13.pdf).  
Tiago S. Prado (2021): »Kill Zones? Effects of Big Tech Start-up Acquisitions on Innovation«. 23rd Biennial Conference of the International Telecommunications Society (ITS). <https://www.econstor.eu/bitstream/10419/238049/1/Prado-Kill-Zones.pdf>.

IM MITTELPUNKT EINER  
DATENSTRATEGISCHEN ABWÄGUNG  
SOLLTE DIE FRAGE STEHEN, OB DER  
GESAMTGESELLSCHAFTLICHE NUTZEN  
DER ÖFFNUNG DENJENIGEN EINES  
WEITEREN VERSCHLUSSES ÜBERWIEGT.

von Datenressourcen auszugleichen, kann aber in einer freien Marktwirtschaft nicht die dauerhafte Aufgabe von Politik sein.

Wer stattdessen Ja zu mehr datengetriebenen Innovationen sagt, muss auch Ja zum freien Wettstreit der Ideen sagen. Nur ein freier Wettbewerb in der Datennutzung sichert die Breite dieses Wettstreites der Ideen und bewegt mitunter auch Marktteilnehmer in komfortablen Wettbewerbspositionen dazu, konsequent Verbesserungen ihrer Produkte oder Neuschöpfungen voranzutreiben. Noch wichtiger: Er treibt alle Marktteilnehmer dazu an, aus purem Interesse am eigenen wirtschaftlichen Überleben nicht davor zurückzuschrecken, einfachere, preislich kompetitivere Geschäftsmodelle auch in vermeintlichen Nischenmärkten zu suchen.

### Wie Zielkonflikte zu verstehen sind

Diese Annahme führt unweigerlich zu einer letzten vierten vorangestellten Frage: Wie sollten Datenstrategien mit Zielkonflikten hinsichtlich der Nutzung oder Verfügbarmachung von Daten umgehen? Als scheinbar besonders knifflige Varianten dieser Fragestellung werden gerne reale oder fiktive Szenarien skizziert, in denen eine Nutzung von Daten im Widerspruch zu Privatsphäreinteressen oder der Realisierbarkeit von Geschäftsmodellen steht. Um Diskussionen in diesem Kontext abzukürzen, sei gleich zu Beginn gesagt: Realistischer Anspruch von Datenstrategien kann nur sein, grundlegende Prinzipien und den politischen Rahmen für einen »Entscheidungs- und Handlungsbias« zu definieren.<sup>25</sup> Wie bereits weiter oben erwähnt,

kann es nicht Ziel von Datenstrategien sei, legitime Privatsphärenansprüche zu beschädigen. Was »legitim« ist, kann jedoch, wie bereits erläutert, durch nationale Gesetzgeber über in der DSGVO explizit vorgesehene abweichende Datenschutzregelungen und weitere technische oder rechtliche Vorgaben zur Anonymisierung eingegrenzt werden.<sup>26</sup> Sind derartige Voraussetzungen gegeben, sollte eine von Datenstrategien getriebene Nutzung oder sogar Öffnung von Daten zumindest deutlich seltener zu Zielkonflikten mit dann noch als legitim betrachteten Privatsphäreinteressen führen, als dies heutige Debatten vermuten lassen.

Etwas komplizierter stellt sich die Lage im Falle von Geschäftsinteresse dar: Insbesondere wenn Unternehmen dazu verpflichtet würden, bestimmte Daten zu teilen oder gänzlich zu öffnen, sind Zielkonflikte durchaus vorstellbar. Bestehende Geschäftsmodelle können dadurch im Extremfall so weit unter Druck geraten, dass Umsätze bis zum Totalausfall sinken. Dieses Papier nimmt allerdings an, dass Datenstrategien gesellschaftlich und volkswirtschaftlich Nutzen und Mehrwerte durch Innovation heben sollen. Eine Abwägung von Zielkonflikten im Kontext von bisher geschlossenen Daten sollte daher gerade nicht danach fragen, ob mögliche Innovationen in bestehende Geschäftsmodelle eingreifen. Stattdessen sollte im Mittelpunkt einer datenstrategischen Abwägung die Frage stehen, ob der zu erwartende gesamtgesellschaftliche und volkswirtschaftliche Nutzen, der aus einer Datennutzung durch breitere Nutzendengruppen zu erwarten ist, den gesamtgesellschaftlichen und volkswirtschaftlichen Nutzen eines weiteren Verschlusses eben jener Daten überwiegt. Solange Schritte zur Datenteilung

<sup>25</sup> Mit anderen Worten: Dass Datenstrategien so formuliert werden könnten, dass sich sämtliche, im Kontext einer größeren Datennutzung und -verfügbarkeit gemeinhin formulierten Interessenkonflikte programmatisch auflösen, wäre eine überzogene Erwartung. Ebenso ist es überzogen zu erwarten, dass Datenstrategien alle derartigen Konflikte berücksichtigen oder auch nur antizipieren könnten. Gut formuliert setzen Datenstrategien aber klare Prinzipien, um zu entscheiden, welche Interessen Vorrang genießen sollten und wie – unter Berücksichtigung der technischen, rechtlichen und ökonomischen Gestaltungsmöglichkeiten im Einzelfall – mit den dann noch verbleibenden tatsächlichen Zielkonflikten umgegangen werden sollte.

<sup>26</sup> Dass der Gesetzgeber gerade in diesen Fragen in Deutschland mit der DSGVO gegebene Handlungsmöglichkeiten bisher nicht ausreichend genutzt hat, stellten beispielsweise mehrere Parlamentarier:innen von Bündnis 90 / Die Grünen in einem gemeinsamen Strategiepapier im August 2021 fest. Der in diesem Entscheidungsvakuum entstandene Regelungs- und Rechtsprechungsdschungel hat zu erheblicher Rechtsunsicherheit geführt, die ihren Ausdruck auch in extensiven Datenschutzdebatten findet: [https://annachristmann.de/wp-content/uploads/2021/08/Strategiepapier\\_Daten-fuer-den-Wandel-nutzen\\_layout.pdf](https://annachristmann.de/wp-content/uploads/2021/08/Strategiepapier_Daten-fuer-den-Wandel-nutzen_layout.pdf).

DER ZUSATZNUTZEN TECHNOLOGIEGETRIEBENER,

DISRUPTIV WIRKENDER GESCHÄFTSMODELLE

LÄSST SICH NICHT UNBEDINGT MIT

KLASSISCHEN BIP-METRIKEN ABBILDEN.

und -öffnung nicht grundlegend in Geschäftsgeheimnisse eingreifen, geht dieses Papier daher davon aus, dass der Gesamtnutzen eines breiteren Datenzugangs den Gesamtnutzen eines weiteren Datenverschlusses nahezu immer überwiegt.

Vor allem im Hinblick auf die großen Datenbestände von Monopolisten und Marktführern in hochkonzentrierten Märkten stehen die Aussicht auf erheblich erweiterte Innovationsmöglichkeiten bei Datenöffnungen und die Bewahrung von faktischen Datenmonopolen sowie darauf aufbauenden Geschäftsmodellen einander gegenüber. Wie der Zusatznutzen neuer Geschäftsmodelle gemessen werden kann, ist allerdings gerade im digitalwirtschaftlichen Kontext nicht immer einfach zu bestimmen. Die britische Ökonomin Diane Coyle zeigt in ihren Arbeiten<sup>27</sup>, dass sich die Zusatznutzen (insbesondere im Sinne von Konsumentenrenten, Englisch »consumer surplus«) technologiegetriebener, disruptiv wirkender Geschäftsmodelle nicht unbedingt mit Hilfe klassischer BIP<sup>28</sup>-Metriken abbilden lassen. Vereinfacht dargestellt: Wenn Kund:innen ihre Nachfrage durch neue Produkte oder Services günstiger befriedigen können, und damit einem Unternehmen oder sogar ganzen Wirtschaftssektoren weniger Umsätze bescheren, führt dies unter Umständen zu stagnierendem oder gar rückläufigem Wachstum. Vor allem wenn Produkte oder Services kostenlos sind, also keinen monetären Preis haben, erfolgt mittels BIP-basierter Metriken überhaupt keine Messung ökonomischer Aktivität – wengleich

deren Mehrwerte enorm sein können.<sup>29</sup> Aus Sicht dieses Papiers entkräftet dieses Mess- und Indikatorikproblem allerdings nicht die Hypothese, dass eine breitere Datennutzung und Datenöffnung überwiegende Mehrwerte schaffen kann – auch und gerade, wenn davon bestehende Geschäftsmodelle nicht unberührt bleiben. Für politische Entscheidungsträger:innen bleiben diese Abwägungen aber unzweifelhaft komplex, vor allem weil eindimensionale, rein ökonomische Nutzenbewertungen zumindest a priori irreführend sein können. Daher soll ein abschließendes Beispiel dreierlei verdeutlichen: Erstens, wie sehr die oben angesprochenen Zielkonflikte oftmals eine Frage politischer Priorisierung sind. Zweitens, wie groß die »Innovationsrendite« einer verpflichtenden Datenöffnung sein kann. Und drittens, dass Regierungen in dieser Frage grundlegend und entschieden regulieren können, wenn sie nur wollen.

Im März 2000 verkündeten der damalige US-Präsident Bill Clinton und der britische Premierminister Tony Blair in einer gemeinsamen Erklärung, dass Genomsequenzierungsdaten öffentlich und frei verfügbar sein sollten. Sie untergruben damit die Geschäftsmodelle hoch bewerteter Biotechunternehmen, die auf eine Patentierung ebenjener Sequenzierungen setzten und beendeten damit, quasi im Alleingang, den Boom der Biotech-Branche um die Jahrtausendwende.<sup>30</sup> Gleichzeitig finanzierten die beiden Regierungen als Teil eines internationalen Konsortiums das bereits seit 1993 laufende Human Genome Project. Dieses verfolgte das Ziel, das menschliche Genom erstmalig vollständig zu sequenzieren und in einer Datenbank

<sup>27</sup> Diane Coyle (2016): »Digitally disrupted GDP«.

<https://voxeu.org/article/digitally-disrupted-gdp>.

Diane Coyle (2015): »Modernising Economic Statistics: Why It Matters«. National Institute Economic Review No. 234.

<https://journals.sagepub.com/doi/pdf/10.1177/002795011523400108>.

Diane Coyle (2015): »Why GDP Statistics Are Failing Us«.

<https://www.uschamberfoundation.org/article/why-gdp-statistics-are-failing-us>.

<sup>28</sup> BIP: Bruttoinlandsprodukt.

<sup>29</sup> Einfache Beispiele sind hier die Nutzung kostenloser E-Mail- und Kartendienste im Gegensatz zu klassischen Briefen bzw. Produkten von Kartenverlagen. Beide Produkte werden tagtäglich millionenfach genutzt, kreieren dadurch mutmaßlich erhebliche Mehrwerte. Da für diese Dienstleistungen aber niemand unmittelbar bezahlt, erfasst eine BIP-basierte Wertschöpfungsmessung diese nicht.

<sup>30</sup> Alex Berenson, Nicholas Wade (2000): »Clinton-Blair Statement on Genome Leads to Big Sell-Off«. <https://archive.nytimes.com/www.nytimes.com/learning/students/pop/articles/031500sci-human-genome.html>.

»Accidental death of a bubble«. Nature Biotechnology Vol. 18, S. 469 (2000). [https://www.nature.com/articles/nbt0500\\_469a.pdf](https://www.nature.com/articles/nbt0500_469a.pdf).

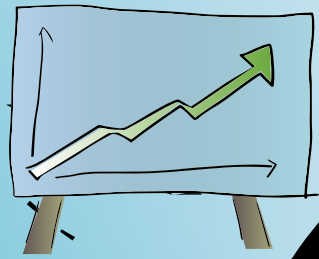
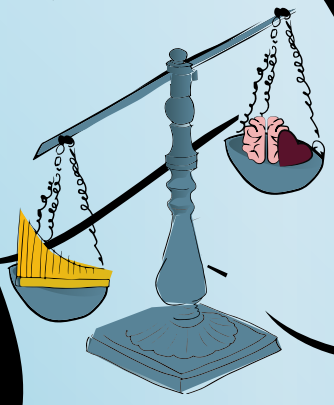
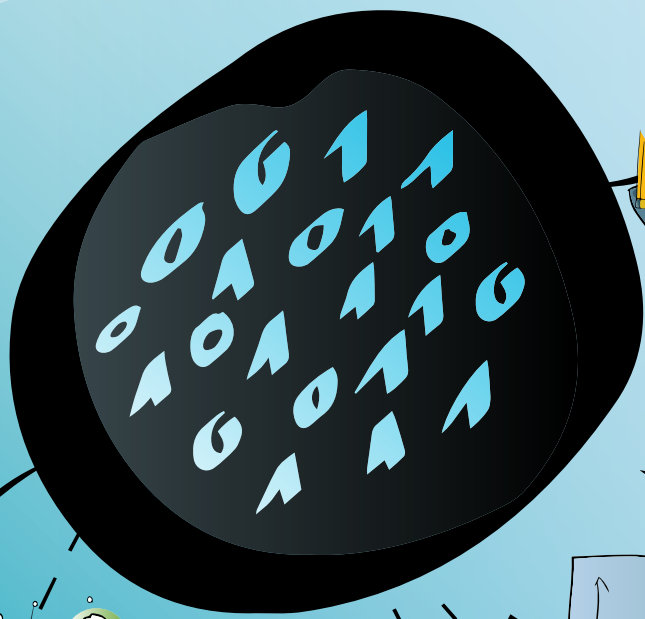
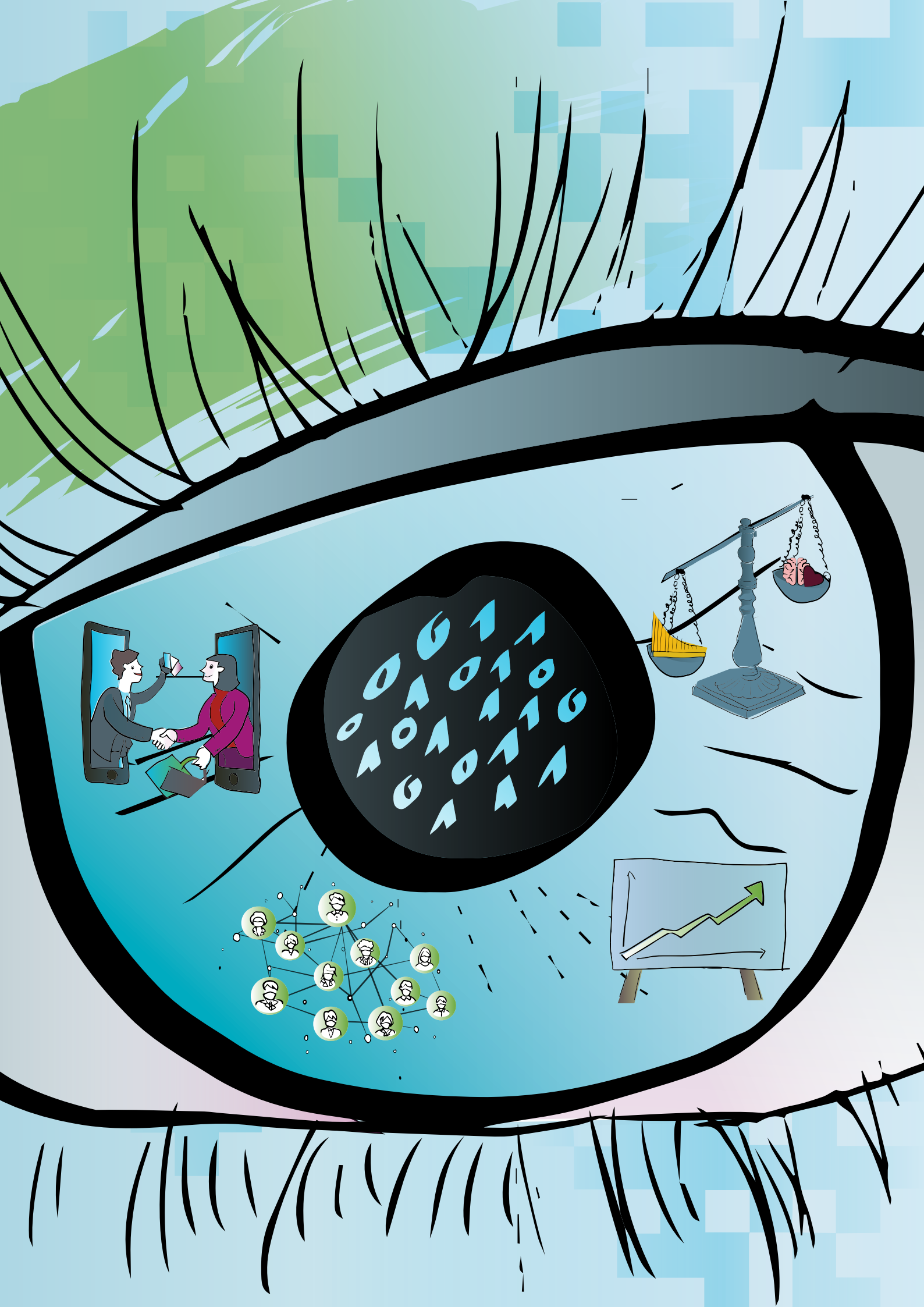
DIE ÖFFENTLICH UND FREI  
ZUGÄNLICHEN  
GENOMSEQUENZIERUNGSDATEN  
HATTEN IN DEN USA  
EINEN BETRÄCHTLICHEN  
VOLKSWIRTSCHAFTLICHEN IMPACT.

frei zur Verfügung zu stellen. Rückblickend brachte diese Richtungsentscheidung nicht nur einen Quantensprung in der Biologie als datengetriebene, digitalisierte Wissenschaft hervor, sondern beförderte auch die Entwicklung von gänzlich neuen Therapien gegen Mukoviszidose oder Chorea Huntington.<sup>31</sup> Auch die Entwicklung mRNA-basierter Coronaimpfstoffe wäre ohne eine umfassende Genomsequenzierung und ohne Genomeditierungsverfahren nicht möglich gewesen. Obendrein warf das Human Genome Project auch eine beachtenswerte volkswirtschaftliche Rendite ab. Laut einer als Battelle Report bekannt gewordenen Studie investierte die US-Regierung als mit Abstand größter Financier von 1988 bis 2010 3,8 Mrd. US-Dollar in das internationale Konsortium des Human Genome Projects. Durch weitere Forschungsprojekte, Spin-offs oder andere Industrietätigkeiten trugen diese Investitionen bis 2013 allein in den USA zu einem wirtschaftlichen Output von 965 Milliarde US-Dollar bei, generierten Einkommen in Höhe von 293 Milliarde US-Dollar und finanzierten 4,3 Millionen. Vollzeit-Beschäftigungsjahre.<sup>32</sup> Ob diese Effekte auch erreicht worden wären, wenn Genomsequenzierungsdaten heute geistiges Eigentum einiger weniger Unternehmen wären, darf bezweifelt werden.

---

<sup>31</sup> Richard A. Gibbs (2020): »The Human Genome Project changed everything«. Nature Reviews Genetics Vol. 21, S. 575 – 576. <https://www.nature.com/articles/s41576-020-0275-3.pdf>.

<sup>32</sup> U.S. Department of Energy, Human Genome Project (2019): »Report Details Economic Benefits of the HGP«. [https://web.ornl.gov/sci/techresources/Human\\_Genome/project/economics.shtml](https://web.ornl.gov/sci/techresources/Human_Genome/project/economics.shtml).



# 4. THESENDISKUSSION UND HANDLUNGSEMPFEHLUNGEN

Wenn Datenstrategien mit dem Ziel definiert werden, die Datennutzung tatsächlich zu fördern, müssen sie auf soliden Grundlagen aufbauen. Ausgehend von den vorgelagerten Annahmen werden in diesem Abschnitt sechs Thesen diskutiert, die das Fundament für Datenstrategien bilden sollten. Daraus werden jeweils übergeordnete Handlungsempfehlungen für die Gestaltung von Datenstrategien abgeleitet.

## 4.1 MENSCHEN UND GESELLSCHAFTEN SIND DATENGETRIEBEN

Wer mit Suchmaschinen das Internet nach Kombinationen der Wörter »Daten«, »Mensch« sowie »künstliche Intelligenz« oder »Data Analytics« durchsucht, findet denkwürdige Resultate: Einträge wie »Wie menschlich ist KI?«, »So geht digital menschlich!« oder, auf Englisch, »Why data needs humans as much as humans need data«. Der Weg über die Suchmaschinen ist dabei vor allem interessant, weil diese uns, aus rein technisch-methodischen Gründen, den Spiegel vorhalten. In den Resultaten offenbart sich der bekannte Topos des Menschen versus die Maschine, an dem man sich gerade in Deutschland nicht erst seit Fritz Langs »Metropolis«, sondern mindestens seit E.T.A. Hoffmanns »Der Sandmann« kritisch abarbeitet. Weil Daten als Medium der Automatisierung und Futter für Algorithmen gesehen werden, schlägt ihnen auch heute nicht selten eine gewisse Grundskepsis entgegen. Die reicht von der Auffassung, Daten seien ein neues Phänomen und daher (noch) gesellschaftsfremd, bis zu offener Ablehnung, weil Daten in Konkurrenz zu menschlicher Intuition und Autonomie ständen. Allein, dieser Blickwinkel täuscht – und führt bei der Formulierung von Datenstrategien auf die falsche Fährte.

Daten sind nicht menschen- und noch weniger gesellschaftsfremd. Ganz im Gegenteil sind Menschen von Natur aus Datenverarbeitungswesen: Ihren optischen Zugang zur Welt erschließen sie über 10 bis 100 Millionen Bits pro Sekunde, die über die Retina erfasst und über den Sehnerv an das Gehirn

gesendet werden.<sup>33</sup> Das überdurchschnittlich große und leistungsfähige Gehirn ermöglicht beispielsweise eine hoch performante Mustererkennung, die es Menschen erlaubt, sich auch in unbekanntem Umgebungen schnell und präzise zu orientieren. Sie meistern damit im Vorbeigehen eine Aufgabe, mit der auch mit modernster Sensorik ausgestattete Roboter immer noch hadern.<sup>34</sup>

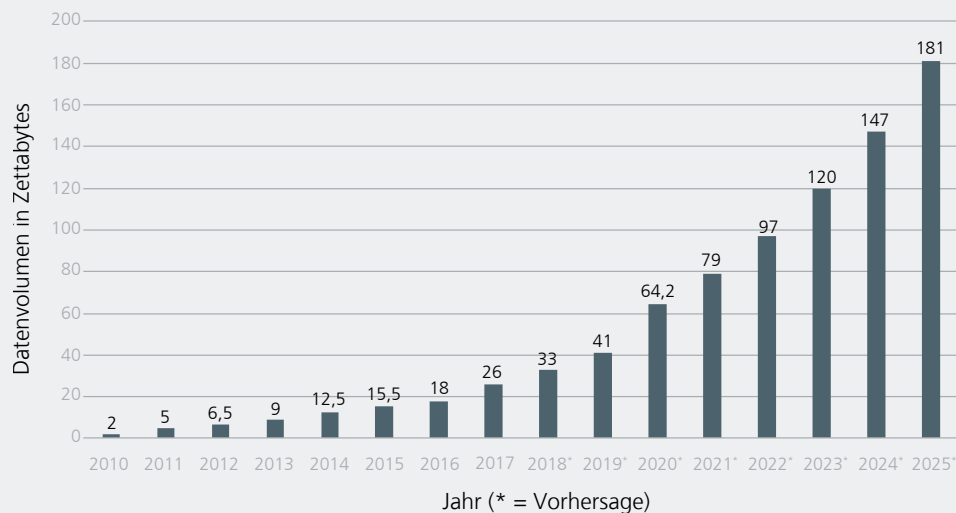
Überragend sind diese Fähigkeiten jedoch erst in Kombination. Der Historiker Yuval Harari argumentiert, dass die kognitiven Fähigkeiten des Menschen erlaubten, ab einem Zeitpunkt vor ca. 70.000 Jahren gemeinsame Glaubens-, Vertrauens- und Wissenssysteme aufzubauen. Die in der Folge entstandenen Gesellschaften ermöglichen es Menschen seit Jahrtausenden, flexibel und in großen, sogar stetig steigenden Zahlen miteinander zu kooperieren. Kritisch ist dabei, dass Kooperation möglich ist, ohne dass sich die Beteiligten unmittelbar kennen oder gar begegnen müssen. Für Steinzeitmenschen wäre es vermutlich unvorstellbar gewesen, einen Wocheneinkauf zu tätigen – und das auch noch, ohne dass man dazu auch nur einen Erzeuger kennen muss. Für die Koordination dieser Kooperationen in Gesellschaften spielt die Verarbeitung von Daten eine zentrale Rolle, weil sie fiktive gemeinsame Bezugspunkte für globale Arbeitsteilungen, Rechtssysteme und andere Ideenräume schafft. Aus unserem Alltag ist die Datenerfassung und -nutzung daher nicht wegzudenken: Daten erfassen und verarbeiten wir beispielsweise, wenn wir Handel treiben, bereits beim Tauschhandel, vor allem aber, wenn wir mit Goldmünzen oder Papiergeld bezahlen oder bargeldlos unsere Geschäfte abwickeln. Mit jeder Transaktion erheben und verarbeiten wir Daten, unter

<sup>33</sup> Kristin Koch et al. (2006): »How much the eye tells the brain«. Current Biology Vol. 16 Issue 14, S. 1428 – 1434. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1564115/>.

<sup>34</sup> Im Englischen wird diese Fähigkeit mit dem Akronym SLAM abgekürzt, das für »Simultaneous Localisation and Mapping« steht. SLAM bleibt ein zentrales Problem der Robotik, das in der Entwicklung autonomer Systeme eine zentrale Herausforderung darstellt. SLAM-Probleme sind regelmäßig prominente Elemente der DARPA Grand Challenge, wie z.B. im Jahr 2006 ([https://en.wikipedia.org/wiki/DARPA\\_Grand\\_Challenge](https://en.wikipedia.org/wiki/DARPA_Grand_Challenge), <http://robots.stanford.edu/papers/thrun.stanley05.pdf>). Auch Privatunternehmen wie der Chiphersteller Intel investieren in Wettbewerbe zum Lösen von SLAM-Problemen: <https://lifelong-robotic-vision.github.io/>.



Abbildung 1: Globales Datenwachstum 2010 – 2025 (Schätzungen der International Data Corporation)<sup>35</sup>



anderem darüber, für wie wertvoll wir Objekte oder Dienstleistungen einschätzen – und darüber, ob unsere Mittel es uns erlauben, diese zu kaufen. Nur wenn die Daten zur Werteschätzung des Verkäufers und des Käufers ähnlich sind, kann letztlich eine Transaktion erfolgen.

Während dieser Umgang mit Daten schon seit Jahrtausenden eine zentrale implizite Stütze von Gesellschaften ist, haben moderne Technologien zu Umbrüchen in drei Bereichen geführt: Erstens verfügen wir über immer einfachere technische Mittel, um Daten von ihren ursprünglichen Erhebungsformen und -kontexten zu trennen. Vor allem mit der massiven Verbreitung moderner Tabellenkalkulations- und Datenbanksoftware ab den 1980er Jahren wurde diese Fähigkeit massentauglich. Waren Daten früher beispielsweise implizit an eine Transaktion oder explizit auf Papier gebunden, lassen sie sich heute in nahezu unendlich flexible Datenbanken eingeben, von wo aus sie in immer neuen Kontexten und Analysen genutzt werden können. Mit papiergefassten Tabellen oder Lochkarten war das zwar nicht unmöglich, aber doch sehr viel komplizierter. In einer Umkehrung von Marshal McLuhans berühmtem Sprichwort ist also das Medium *nicht* länger die Nachricht.

Zweitens erlauben immer feinere Sensorik und im Sinne von Moore's Law immer günstigere Speichertechnologien die Sammlung von immer mehr Daten. War früher die Erfassung von Daten eine aufwändige händische Angelegenheit, sammeln z. B. moderne Düsentriebwerke heute pro Flug zwischen einem halben und mehreren Terabyte (TB) an Daten, um Fehlfunktionen frühzeitig zu erkennen, Verschleiß zu berechnen und den Teileaustausch optimal zu terminieren.<sup>36</sup> Weltweit ist das jähr-

lich gesammelte und digital gespeicherte Datenvolumen infolge dieser Entwicklung alleine seit 2010 rapide von ca. 2 Zettabyte auf geschätzte 64 Zettabyte im Jahr 2020 gewachsen. Hält das exponentielle Wachstum an, wird dieses Volumen bis 2025 auf 181 Zettabyte, d. h. 181 Milliarden Terabyte, projiziert.<sup>37</sup>

Drittens besitzen wir Fähigkeiten, die es uns erlauben, diese gigantischen Datenmengen auch zu verwerten, um neue Erkenntnisse, Ideen, Produkte und Services zu kreieren. Ob mit Hilfe klassischer Datenanalysesoftware, künstlicher Intelligenz oder mit dem Wissen speziell trainierter Datenanalytist:innen: Wir können Daten mehrfach analysieren, zusammenführen und in vollständig neuen, für den originären Datenerhebungszweck nicht vorgesehenen Zusammenhängen auswerten. Gerade diese, vom ursprünglichen Erhebungszweck entkoppelte und anwendungsfallübergreifende Verwertung ist dabei das zentrale Charakteristikum und der methodisch-qualitative Quantensprung datengetriebener Innovationsprozesse.

Quelle der wahrnehmbaren Verunsicherung ist in diesem Kontext nicht, dass der relativ plötzliche Sprung ins Datenzeitalter Menschen oder Gesellschaften mit einem gänzlich neuen Phänomen konfrontiert, sondern dass der Datenreichtum und die Chancen zu dessen Verwertung nicht annähernd gleich verteilt sind: Das sich aufblähende globale Datenvolumen befindet sich zu erheblichen Teilen in den Händen weniger (Superstar-)Firmen – und die meisten Daten kommen dabei nicht aus IoT<sup>38</sup>-Sensoren, sondern entstehen im Internet in sozialen Netzwerken und Onlinemärkten. Besonders problematisch ist aktuell, dass die Chancen zur Datennachnutzung für viele *potenzielle* Nutzende, gemessen am Gesamtvolumen der global existierenden Daten, gegen Null tendieren. Dies liegt schlichtweg daran,

<sup>35</sup> David Reinsel, John Rydning, John F. Gantz (2021): »Worldwide Global Data Sphere Forecast, 2021-2025: The World Keeps Creating More Data – Now, What Do We Do With It All?«. <https://www.idc.com/getdoc.jsp?containerId=US46410421>.

<sup>36</sup> RTInsights Team (2016): »How Rolls-Royce Maintains Jet Engines With the IoT«. <https://www.rtinsights.com/rolls-royce-jet-engine-maintenance-iot/>.

<sup>37</sup> Ulrike Hack (2021): »What's the real story behind the explosive growth of data?«. <https://www.red-gate.com/blog/database-development/whats-the-real-story-behind-the-explosive-growth-of-data>.

<sup>38</sup> IoT: Internet of Things, Internet der Dinge.



DATENSTRATEGIEN SOLLTEN  
THEMATISIEREN, WIE DIE  
WIRKUNG VON DATEN ALS HEBEL  
FÜR GESELLSCHAFTLICHEN,  
WIRTSCHAFTLICHEN UND POLITISCHEN  
NUTZEN MAXIMIERT WERDEN KANN.

dass die meisten Daten unter Verschluss gehalten werden, nicht nur von Superstarfirmen des Silicon Valley. Diesem Status quo der Datenwirtschaft lässt sich aber dank einiger besonderer Eigenschaften von Daten gezielt entgegenwirken. Davon handeln die folgenden Thesen.

### Handlungsempfehlungen

Datenstrategien sollten thematisieren, wie die Wirkung von Daten als *Hebel* für gesellschaftlichen, wirtschaftlichen und politischen Nutzen maximiert werden kann. Vor allem wenn sie von staatlicher Seite definiert sind, sollten Datenstrategien daher grundsätzlich von der Annahme ausgehen, dass Daten nicht nur gesellschaftsnah, sondern an vielen Stellen eine bisher oft unsichtbare Grundlage für gesellschaftliches Zusammenleben sind. Teilhabe, ob in ökonomischer, politischer oder sozialer Form, wird aber in der datengetriebenen Gesellschaft und Wirtschaft durch neue Asymmetrien in Frage gestellt. Im Mittelpunkt von Datenstrategien sollte daher das Interesse stehen, diese Asymmetrien in Bezug auf Datenzugänge auszugleichen und neue Chancen zur Datennutzung für mehr Datennutzende zu schaffen.

## 4.2 ERST DURCH NUTZUNG SCHAFFEN DATEN WERT

Daten lassen sich zwar erheben und massenhaft speichern, aber als Immaterialgut sind sie nicht greifbar. Daher führen die physischen Behelfsmetaphern des »Datenschatzes« und der Daten als Öl des 21. Jahrhunderts auf eine falsche Fährte: Sie suggerieren, dass Daten einen immanenten Wert besäßen, also *selbst* wertvoll seien. Allein der Besitz oder die Kontrolle von Daten erscheint damit schon als verwertbarer Vorteil oder gar unmittelbarer Gewinn. Bleibt man in dieser Gedankenwelt, ist es konsequent, den »Datenschatz« am besten streng gesichert zu lagern und keinesfalls allzu freigiebig zugänglich zu machen, denn die Verwertbarkeit der Daten hängt ja auch

vermeintlich mit deren Exklusivität zusammen. Weil Daten von sich aus wertvoll sind, so lässt sich der Gedanke fortführen, ist es in jedem Fall nicht schädlich, mehr Daten zu sammeln und zu horten, vor allem wenn dies mit Sensoren, (halb-)automatischen Geschäftsprozessen und durch das extensive Abgreifen von Kommunikationsdaten immer einfacher geht. Ob die so angehäuften Daten jemals genutzt werden oder als sogenannte »Dark Data« im Speicher versauern, erscheint zweitrangig. Dass diese Logik in der Realität verfangt, zeigt eine paradoxe Entwicklung: Wie oben bereits beschrieben werden zwar einerseits immer mehr Daten gesammelt, andererseits wird aber ein Großteil dieser Daten nicht ein einziges Mal genutzt. Der Anteil sogenannter Dark Data am globalen Datenvolumen beträgt je nach Schätzung zwischen 55 und 90 Prozent.<sup>39</sup>

Ein solcher Umgang mit Daten schafft aber keinen Wert. Ganz im Gegenteil ist er eine garantiert schlechte Alternative mit stets negativen Opportunitätskosten, weil er Chancen nicht nutzt, das Mehrwertpotenzial von Daten verfallen lässt und darüber hinaus auch noch Ressourcen unmittelbar bindet.

Erstens kann jedes unzureichend oder gar nicht genutzte Datum dazu beitragen, dass wissenschaftliche Erkenntnisse nicht gewonnen, neue Geschäftsmodelle nicht erdacht und politische oder soziale Fragestellungen nicht beantwortet werden. Sei der Mehrwert einer Nutzung auch noch so gering, so stimmt in jedem Fall der Grundsatz, dass *ungenutzte* Daten per

<sup>39</sup> Deutsche Unternehmen saßen im Jahr 2015 laut einem Survey von Veritas Technologies im Vergleich zu Unternehmen in Europa, dem Nahen Osten und Afrika auf der größten Dark-Data-Halde: [http://images.info.veritas.com/Web/Veritas/%7B364a7ca5-e05c-4fce-971b-88e18c62eafb%7D\\_45145\\_EMEA\\_Veritas\\_Strike\\_Report\\_Gulf.pdf](http://images.info.veritas.com/Web/Veritas/%7B364a7ca5-e05c-4fce-971b-88e18c62eafb%7D_45145_EMEA_Veritas_Strike_Report_Gulf.pdf).

Dazu führen verschiedene andere IT- und Data-Analytics-Unternehmen in unregelmäßigen Abständen Schätzungen durch, z. B. Forrester (<https://go.forrester.com/blogs/hadoop-is-datas-darling-for-a-reason/>), Gartner (<https://www.gartner.com/en/documents/3869701/market-guide-for-file-analysis-software0>), Seagate (<https://www.seagate.com/files/www-content/our-story/rethink-data/files/rethink-data-report-2020-de-de.pdf>), Splunk ([https://www.splunk.com/en\\_us/form/the-state-of-dark-data.html](https://www.splunk.com/en_us/form/the-state-of-dark-data.html)) und IBM (<https://siliconangle.com/2015/10/30/ibm-is-at-the-forefront-of-insight-economy-ibminsight/>).

DATEN SELBST SIND NICHT WERTVOLL.

DIE NICHTNUTZUNG ANGEHÄUFTER

DATEN FÜHRT JEDOCH ZU NEGATIVEN

OPPORTUNITÄTSKOSTEN.

se nicht nützlich – und damit auch nicht wertvoll – sein können. Jede ausbleibende Nutzung bedeutet folglich den Verzicht auf Chancen zur Wertschöpfung.

Zweitens verlieren viele Daten, wie etwa Kommunikations-, Mobilitäts- und Produktionsdaten, mit der Zeit an Aussagekraft. Menschliche Verhaltensmuster beispielsweise verändern sich mit der Zeit, sodass Daten, die alte Muster abbilden, für aktuelle Analysen bestenfalls eingeschränkt brauchbar sind. Ebenso können maschinelle Sensordaten nutzlos werden, weil Maschinen schlichtweg ausgetauscht oder Produktionsprozesse verändert wurden.<sup>40</sup> Diese stetig abnehmende Nützlichkeit führt gleichzeitig zur schleichenden Abwertung sämtlicher, noch so marginaler Investitionen in Prozesse und Technologien zur Erhebung der betroffenen Daten. Sollten Datenbestände bilanzierbar werden, müssten Buchhalter für Dark Data wohl besonders große Abwertungsfaktoren ansetzen. Mit der Zeit treibt die Nichtnutzung angehäufter Datenbestände somit die Opportunitätskostenbilanz immer weiter ins Negative.

Drittens bindet das Anhäufen ungenutzter Daten aber auch unmittelbar Ressourcen, denn die Speicherung von Daten mag zwar immer günstiger werden, kostenfrei ist sie deshalb aber noch lange nicht. Veritas Technologies schätzte 2015, dass die Speicherung und das Management von durchschnittlich ca. 1.000 TB ungenutzten Dark Data und redundanten Daten ein mittelgroßes Unternehmen jährlich ca. 650.000 US-Dollar kosten.<sup>41</sup> Das Horten von Daten ohne eine Nutzung, die im Ge-

genzug Wert erzeugt, verzehrt somit unmittelbar finanzielle Ressourcen.<sup>42</sup> Dass Unternehmen oder auch Verwaltungen diese Kosten regelmäßig indirekt auf Kund:innen, Aktionär:innen oder Steuerzahler:innen umlegen können, rettet vielleicht die buchhalterische Bilanz. Dass damit die Kosten für einen verschwenderisch ineffizienten Umgang mit Daten vergemeinschaftet werden, macht die Sache aus gesellschaftlicher oder volkswirtschaftlicher Perspektive aber nicht besser.

Daten *selbst* sind also nicht wertvoll. Die *Nichtnutzung* angehäufter Daten führt eher zu negativen Opportunitätskosten und ist eine schlechte Alternative. Sinnvoller erscheint daher die Betrachtung, dass *nur die Nutzung* von Daten Wert schafft. Wie konkret diese Logik schon vor über 20 Jahren selbst in traditionellen Industrien griff, zeigt das folgende Beispiel: Ende der 1990er Jahre suchte der CEO der GoldCorp Inc. und ehemalige Merrill-Lynch-Banker Rob McEwen mit Hochdruck nach Wegen, seine 1989 gekaufte Red-Lake-Mine aus den roten Zahlen zu führen. Um die Jahrtausendwende förderte die Mine gerade einmal 50.000 Unzen Gold pro Jahr mit Kosten von 360 US-Dollar pro Unze. In einem hochkompetitiven Marktumfeld mit historisch niedrigen Goldpreisen von unter 300 US-Dollar je Unze war die Mine hochdefizitär.<sup>43</sup> Bodengutachten und die Tatsache, dass benachbarte Minen deutlich effizienter produzierten, deuteten zwar darauf hin, dass auch die Red-Lake-Mine auf größere Goldvorkommen hoffen konnte, wo genau wussten die Geolog:innen des Unternehmens aber nicht.

<sup>40</sup> Ausnahmen bestätigen auch hier die Regel: Für Langzeitanalysen sind gerade historische Daten nützlich; ebenso verlieren geologische Daten, beispielsweise für den Bergbau, nur langsam ihre Aussagekraft. Die meisten IoT- und Prozessdaten dürften demgegenüber jedoch eine vergleichsweise kurze Halbwertszeit besitzen.

<sup>41</sup> Veritas (2016): »Veritas Global Databerg Report Finds 85% of Stored Data Is Either Dark or Redundant, Obsolete, or Trivial (ROT)«. <https://www.veritas.com/news-releases/2016-03-15-veritas-global-databerg-report-finds-85-percent-of-stored-data>.

<sup>42</sup> Ebenfalls hinzuzurechnen wären hier grundsätzlich auch die Kosten für die Datenerhebung. Je nach Grad der Automatisierung können diese sehr hoch oder vergleichsweise niedrig sein. Kostenfrei ist die Datenerhebung aber in keinem Fall.

<sup>43</sup> »Open Innovation: Goldcorp Challenge«. <https://www.ideaconnection.com/open-innovation-success/Open-Innovation-Goldcorp-Challenge-00031.html>.

DATENSTRATEGIEN SOLLTEN  
SICH AUF INSTRUMENTE UND  
RAHMENBEDINGUNGEN KONZENTRIEREN,  
DIE EINE TATSÄCHLICHE NUTZUNG  
VON DATEN ERMÖGLICHEN.

Eine Linux-Konferenz am Massachusetts Institute of Technology inspirierte Rob McEwen schließlich, »die Welt zu fragen, wo das Gold in seiner Mine sei«. <sup>44</sup> Im März 2000 rief er »The Challenge« aus: Für ein Preisgeld von 575.000 US-Dollar wurde die beste, d. h. zutreffendste und gewinnbringendste Modellierung der Goldlagerstätten gesucht. Als Arbeitswerkzeug stellte GoldCorp Inc. die gesamten geologischen Daten der ca. 22.000 ha großen Mine öffentlich über das Internet zur Verfügung. Die Daten reichten zurück bis in das Jahr 1948 und umfassten aus heutiger Sicht überschaubare 400 MByte. Nicht jeder war überzeugt von dem Unterfangen. Skeptiker:innen hielten McEwen vor, dass er proprietäre Daten verscherbele, geistiges Eigentum preisgebe und obendrein auch noch öffentlich zugebe, unfähig zu sein, die Goldbestände selbst zu finden. Dem globalen virtuellen Goldtausch, der mit »The Challenge« losbrach, tat dies aber keinen Abbruch. Innerhalb weniger Wochen werteten mehr als 1.000 Wettbewerber:innen, oft ohne Bergbau- oder Geologiehintergrund, die Daten aus. Ohne die Mine jemals selbst besucht zu haben, errangen zwei australische Unternehmen den Sieg mit einem damals neuartigen 3-D-Modell der Mine. <sup>45</sup>

Insgesamt wurden durch »The Challenge« 110 Lagerstätten identifiziert, wovon die Hälfte bisher unbekannt war. 80 Prozent dieser Lagerstätten enthielten signifikante Goldreserven im Wert von rund 6 Milliarden US-Dollar. Die Explorationszeit zur Erschließung der Lagerstätten konnte durch die zutreffende Modellierung um ca. 2 – 3 Jahre reduziert und die Produktionskosten konnten auf unter 60 US-Dollar je Unze gedrückt werden. <sup>46</sup> Für den Aufstieg der GoldCorp zu einem der größ-

ten Goldproduzenten weltweit war diese Kehrtwende von entscheidender Bedeutung.

»The Challenge« ist ein frühes Beispiel für die Umsetzung einer Open-Innovation-Strategie, die externe Kollaborateure zur Erreichung wichtiger Ziele einbindet und sie dazu an unternehmens- oder organisationseigenen Ressourcen, in diesem Fall Daten, teilhaben lässt. Nach dem Erfolg von GoldCorp schrieben auch andere Goldunternehmen Wettbewerbe aus, um Goldbestände basierend auf offenen Daten zu finden. <sup>47</sup> Ganz im originären Sinne wären diese Schätze ohne die breite *Nutzung* von Daten nicht gehoben worden. »The Challenge« war aber auch ein wichtiger Meilenstein für die Entwicklung einer globalen Industrie von 3-D-Modellierungsunternehmen für geologische Anwendungen. Laut Aussage der beteiligten Unternehmen eröffnete es diesen neue Geschäftsfelder und, aufgrund ihrer neu erlangten Bekanntheit, den Zugang zu wichtigen finanziellen Ressourcen. Von der medialen Aufmerksamkeit dürften letztlich auch andere Unternehmen dieser damals noch jungen Industrie profitiert haben. <sup>48</sup>

Zahlreiche Hackathons sind gewissermaßen die Nachfahren dieser Anfänge offener datengetriebener Innovationsprozesse. Während solche Formate aber oftmals nur punktuell wirken und eingesetzt werden, sollten fortschrittliche Datenstrategien die *Datennutzung* priorisieren. Wie das mithilfe einer breiteren Datennutzung und durch eine bessere Zugänglichkeit von Daten *dauerhaft* erreicht werden kann, erläutern die folgenden Thesen.

<sup>44</sup> Myra P. Saefong (2016): »How a gold miner turned a \$1 million investment into \$3 billion«. <https://www.marketwatch.com/story/meet-the-mining-chief-who-wants-to-get-his-company-into-the-sp-500-2016-09-28>.

<sup>45</sup> William (Bill) Power (2000): »The GoldCorp Challenge«. [https://www.csaglobal.com/wp-content/uploads/2019/11/MEGWA-2018\\_The-GoldCorp-Challenge-2000\\_Bill-Power\\_March\\_2018.pdf](https://www.csaglobal.com/wp-content/uploads/2019/11/MEGWA-2018_The-GoldCorp-Challenge-2000_Bill-Power_March_2018.pdf).

<sup>46</sup> Wikiwand (2014): »Goldcorp«. <https://www.wikiwand.com/de/Goldcorp>.

<sup>47</sup> Myra P. Saefong (2016): »How a gold miner turned a \$1 million investment into \$3 billion«. <https://www.marketwatch.com/story/meet-the-mining-chief-who-wants-to-get-his-company-into-the-sp-500-2016-09-28>.

<sup>48</sup> Linda Tischler (2002): »He Struck Gold on the Net (Really)«. <https://www.fastcompany.com/44917/he-struck-gold-net-really>.

KONKURRENZ ENTSTEHT NICHT  
BEI DER NUTZUNG VON DATEN,  
SONDERN ZWISCHEN  
GESCHÄFTSMODELLEN,  
DATENANALYSEN UND ALGORITHMEN.

### Handlungsempfehlungen

Funktionales Ziel von Datenstrategien sollte es sein, die nutzenorientierte Verwendung von Daten zu maximieren. Datenstrategien sollten sich daher auf die Schaffung von Instrumenten und die Herstellung von Rahmenbedingungen konzentrieren, die eine tatsächlich einfache und hürdenfreie Nutzung von Daten ermöglichen. Berührt wird damit die Frage, wie Datenstrategien etwaige Förder- und Anreizmaßnahmen für Datenerhebung, -speicherung und -management sowie -nutzung untereinander priorisieren sollten. Die hier formulierte Empfehlung ist klar: Maßnahmen für ein besseres Datenmanagement und auch die fortlaufende Datenerhebung müssen konsequent so ausgestaltet werden, dass sie dem Ziel einer erhöhten Datennutzung funktional zuarbeiten. Dem Ziel der Datennutzung sollten Datenerhebung und -speicherung folglich strategisch untergeordnet werden. Maßnahmen, die direkt oder indirekt pauschale Anreize zur reinen Speicherung oder Erhebung von Daten ohne klaren Anwendungsfall bzw. ohne Nutzungs- oder Verwertungsszenario setzen, sollten zumindest für staatliche Datenstrategien keine Priorität besitzen.<sup>49</sup>

<sup>49</sup> Problematisch erscheinen in diesem Kontext vor allem Maßnahmen, die keine ausreichende Verknüpfung zu einer tatsächlichen Datennutzung herstellen. Grundsätzlich können und sollten auch Datenstrategien das Sammeln und eine adäquate Speicherung von Daten fördern, beispielsweise wenn es darum geht, Anreize für bisher unzureichend digitalisierte und *datafizierte* Unternehmen (z. B. im Handwerk) zu setzen. Dabei sollte aber immer eine *nutzungsgebundene* Förderkomponente enthalten sein. Kritisch zu bewerten sind aus Sicht dieses Papiers im Kontext von Datenstrategien daher Vorschläge wie die Einführung eigentumsähnlicher Rechtspositionen (»Dateneigentum«), exklusive Datenverwertungsrechte oder auch pauschale Abschreibungsmöglichkeiten für Digitalisierungsinvestitionen mit dem Ziel der reinen Datenerhebung oder -speicherung.

## 4.3 DATEN NUTZEN SICH NICHT AB, LASSEN SICH PERFEKT KLONNEN UND SIND NICHT-RIVALISIEREND NUTZBAR

Die Daten-Challenge für die Red-Lake-Mine war nur aufgrund einer charakteristischen Eigenschaft von Daten umsetzbar: Wenn Daten genutzt werden, nutzen sie sich nicht ab, lassen sich perfekt klonen, d. h. kopieren, und nicht-rivalisierend nutzen. Anders also als ein physischer Schatz können sie gleichzeitig von unterschiedlichen Nutzenden verwertet werden. Die Datennutzung des einen steht einer Datennutzung eines anderen nicht im Weg. Nur so konnten tausende Wettbewerber den GoldCorp-Datenbestand gleichzeitig auswerten. In Konkurrenz standen hingegen die Ideen der Wettbewerber. Aus Sicht fortschrittlicher Datenstrategien ist dies aber höchst wünschenswert, denn wie der US-amerikanische Chemie-Nobelpreisträger Linus Pauling erkannte: Wer gute Ideen haben möchte, braucht viele Ideen.

Dass der möglichst ungehinderte Wettbewerb der besten Ideen der effektivste Innovationsmotor ist, scheint im Fall von »The Challenge« klar. Die gleiche Logik greift aber auch auf dem offenen Markt, wenn Unternehmen und staatliche Organisationen Daten veröffentlichen, damit diese von anderen für neue Geschäfts- und Servicemodelle genutzt werden. Anders als in so mancher politischen Debatte suggeriert, entsteht Konkurrenz in diesem Fall nicht bei der Nutzung von Daten, sondern allenfalls zwischen Geschäftsmodellen, Datenanalysen und Algorithmen. Im Gegenzug winkt, auch aus Sicht einzelner datenveröffentlichender Marktakteure, die Chance auf bessere Ideen, die teure zuvor brach liegende Investitionen verwerten, und darauf, ein Ökosystem an komplementären Geschäftsideen aufzubauen. Im Idealfall kann das Ökosystem von Datennutzenden Innovationen zudem auch noch schneller hervorbringen – und bevor aufwändig erhobene Daten an Aussagekraft verlieren. Linus

DIE LONDONER  
VERKEHRSBEHÖRDE TFL SPART  
DURCH IHRE OPEN-DATA-STRATEGIE  
MEHRERE MILLIONEN PFUND JÄHRLICH EIN.

Paulings Sprichwort ließe sich im Kontext der Datenökonomie daher noch erweitern: Wer gute Ideen haben möchte, braucht viele Ideen – und die möglichst schnell.

Wie man die Eigenschaften der nicht rivalisierenden und abnutzungsfreien Datennutzung einsetzt, um die breite zeitnahe Datennutzung zu fördern, den datenbasierten Innovationswettbewerb zu stimulieren und gleichzeitig *eigene* strategische Ziele zu erreichen, verdeutlicht das Beispiel der Londoner ÖPNV-Betreiber, der Verkehrsbehörde Transport for London (TfL): Im Zuge der Finanzkrise sah sich TfL von 2009 an mit strikten Sparmaßnahmen konfrontiert. Gleichzeitig sollten Ziele zur Ausweitung des Nahverkehrsangebotes, zur Steigerung der Passagierzahlen und zu erhöhter Transparenz umgesetzt werden.<sup>50</sup> Ein wenig profitabler Kostenfaktor waren für TfL Marketingkampagnen zur ständigen Information von Kund:innen, der Unterhalt SMS-basierter Systeme zur Versendung von Nachrichten zu Verspätungen und Streckensperrungen sowie die Entwicklung von mobilen Apps. Statt die dazu erforderlichen Daten weiter selbst über verschiedene Kommunikationskanäle an ÖPNV-Passagiere zu vertreiben, setzte TfL auf die Möglichkeit, Daten kostenfrei bereitzustellen und freien Wettbewerb zwischen solchen Unternehmen zu ermöglichen, die komplementäre Dienstleistungen zum TfL-Kerngeschäft, der Personenbeförderung, anbieten. Zügig baute TfL ab 2010 ein Open-Data-Angebot von inzwischen über 50 Datensätzen auf, die zunächst als statische Dateien, Echtzeit-Data-Feeds und über APIs zur Verfügung gestellt wurden.<sup>51</sup> Bereits 2011 hatten sich für die Letztgenannten mehr als 1.000 Entwickelnde registriert, die bis 2013 hunderte neue Apps entwickelten. Endnutzende, die sich direkt bei TfL informieren möchten, stand seitdem nur noch die TfL-Webseite zur Verfügung; erst seit August 2020 bietet TfL seinen Passagieren

wieder direkt eine funktional stark reduzierte App an, die in der Pandemie helfen soll, Passagierströme gleichmäßiger über den Tag zu verteilen.<sup>52</sup> TfL sparte mit dieser Strategie jährliche Ausgaben von rund 3 Millionen Pfund für Kundenbenachrichtigungen und Marketingkampagnen. Entwicklungskosten von bis zu 150.000 Pfund und jährliche Supportkosten von durchschnittlich 75.000 Pfund je nicht selbst entwickelter App entfielen ebenfalls. Stattdessen fokussierte man sich bei TfL konsequent auf eine hochwertige, möglichst offene und nutzungsfreundliche Datenbereitstellung.

2015 präsentierte TfL dann in einem weiteren Entwicklungsschritt die »Unified API«. Diese eröffnet den Zugriff auf Millionen in Echtzeit aktualisierte und harmonisierte Datenpunkte für sämtliche von TfL betriebene Verkehrsmodi<sup>53</sup> sowie ergänzende Daten aus Unfallstatistiken und zur aktuellen Luftqualität. Alle semantisch ähnlichen oder zusammengehörigen Daten verschiedener Verkehrsträger und Systeme werden seitdem über die Unified API harmonisiert angeboten, d. h. in konsistenten Formaten und Strukturen sowie mit normalisierten Identifikatoren unter anderem für Haltestellen, Umsteigepunkte und Bahnsteige.<sup>54</sup> Nutzende können dadurch TfL-Daten sofort nutzen und kombinieren, ohne erst zeitaufwändige Datenbereinigungen oder formatbedingte Harmonisierungen durchführen zu müssen. Die Daten der Unified API adoptieren zudem den nationalen NaPTAN-Standard, der in Vereinigtem Königreich dazu genutzt wird, mehr als 450.000 ÖPNV-Zugangspunkte

<sup>50</sup> Becky Hogge (2016): »Transport for London get set, go!«. <https://odimpact.org/files/case-studies-transport-for-london.pdf>.

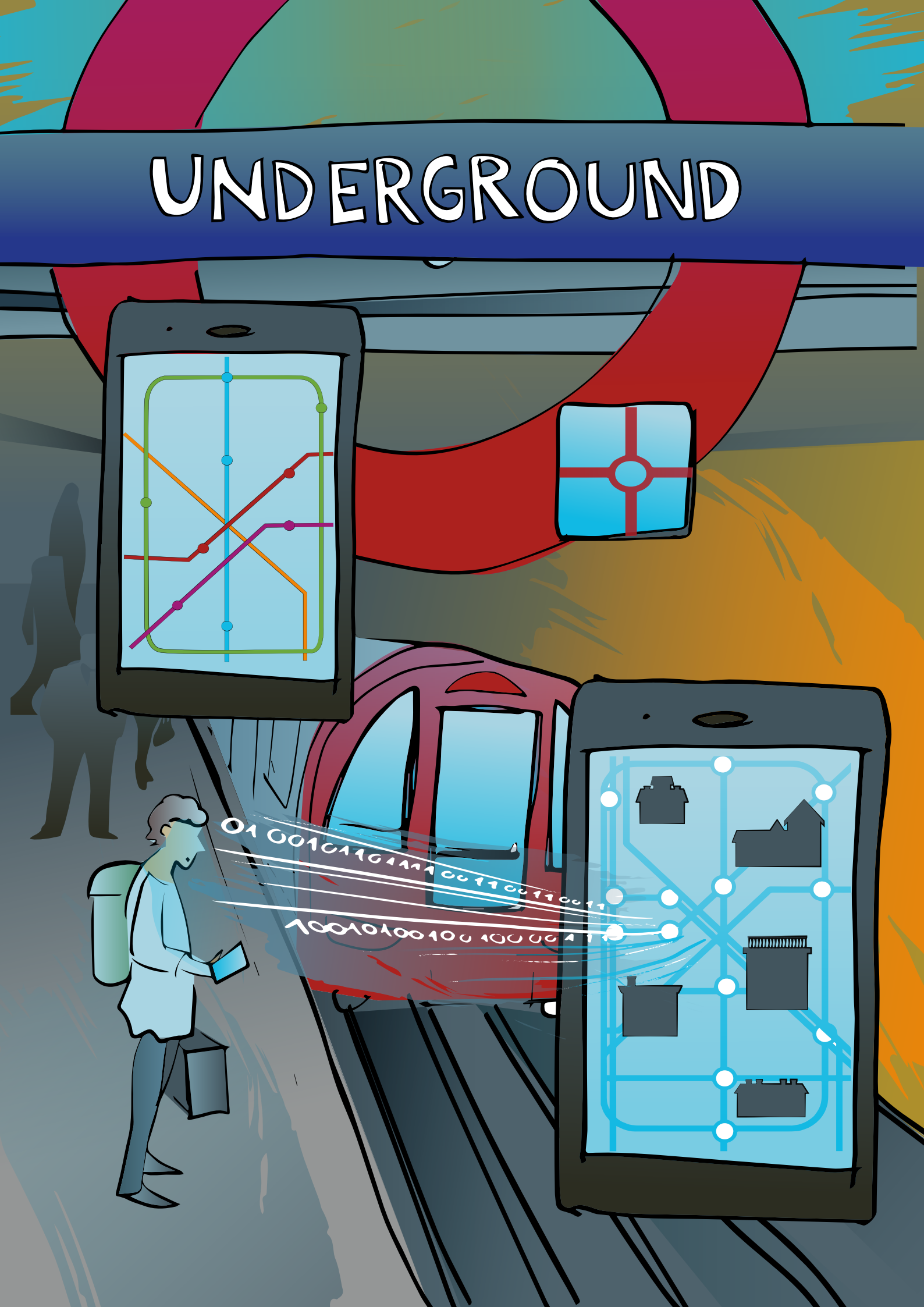
<sup>51</sup> Transport for London: »Our open data«. <https://tfl.gov.uk/info-for/open-data-users/our-open-data>.

<sup>52</sup> Transport for London (2020): »New TfL app to help Londoners plan ahead and travelsafely«. <https://tfl.gov.uk/info-for/media/press-releases/2020/august/new-tfl-app-to-help-londoners-plan-ahead-and-travel-safely>.

<sup>53</sup> Darunter fallen U-Bahnen, Straßenbahnen, Busse, S-Bahnen (Overground), Regionalbahnen, die Docklands Light Railway, die Seilbahn Emirates Air Line sowie die als Boris Bikes bekannten städtischen Leihfahrräder. Diese werden überwiegend von Franchisenehmer:innen betrieben. Zum Datenangebot gehören auch Echtzeitdaten zu Straßensperrungen und aktuellen Verzögerungen auf Straßen durch Staus und Unfälle.

<sup>54</sup> Transport for London: »Unified API«. <https://tfl.gov.uk/info-for/open-data-users/unified-api>.

# UNDERGROUND





UM DIE DATEN VON TFL IST  
EIN INNOVATIVES ÖKOSYSTEM  
SEHR UNTERSCHIEDLICHER  
ANWENDUNGEN ENTSTANDEN.

eindeutig und dauerhaft zu identifizieren. Die Daten der Unified API sind daher auch ohne Probleme mit den Daten der NaPTAN-Datenbank<sup>55</sup> des britischen Verkehrsministeriums sowie mit anderen NaPTAN-konformen Datensätzen kombinierbar. Nicht zuletzt aufgrund dieser auf Nutzbarkeit optimierten Datenbereitstellung hatten sich laut einem Report von Deloitte bis 2017 mehr als 13.000 Entwickelnde als Nutzende der API registriert. Mehr als 600 Apps verwenden Daten der API, davon rund 20 mit mehr als einer Million Downloads.<sup>56</sup>

Um die Daten von Tfl ist damit ein innovatives Ökosystem sehr unterschiedlicher Anwendungen entstanden, wovon nicht alle direkt miteinander konkurrieren: Unter anderem nutzen Makler:innen die Daten, um präzisere interne Wertkalkulationen für Immobilien zu erstellen, andere Daten werden kollaborativ von Wissenschaftler:innen analysiert, um die Pandemiebekämpfung zu unterstützen.<sup>57</sup> Vor allem für die Routenplanung ist hingegen ein hochkompetitiver Markt entstanden, in dem etwa Anwendungen von Citymapper, Google und Mapway um Kund:innen buhlen. Diese Anbieter stehen unter hohem Innovationsdruck, laufend neue Funktionen zu entwickeln und damit den Nutzen ihrer für die Anwendenden in der Regel kostenlosen Apps ständig zu steigern. Damit machen sie aber auch den Londoner ÖPNV insgesamt attraktiver, denn zusammen mit Daten von Uber und anderen Nicht-Tfl-Diensten ermöglichen diese Apps unter anderem eine vereinfachte multimodale Routenplanung, welche die Nutzung des gesamten Angebotsspektrums des öffentlichen Nahverkehrs für Nutzende einfacher macht. Allein die daraus im öffentlichen Nahverkehr

resultierenden Mehrfahrten schätzten Tfl und Deloitte 2017 auf einen Wert von 20 Millionen Pfund. 2017 wurde der durch zusätzliche Einnahmen und Einsparungen entstehende ökonomische Mehrwert von Tfls Open-Data-Initiative zudem auf jährlich 130 Millionen Pfund taxiert. Dazu zählt auch eine jährliche Bruttowertschöpfung von 12–15 Millionen Pfund durch neue Unternehmen und Arbeitsplätze in London. Auf der Gegenseite wendet Tfl nach eigenen Angaben pro Jahr eine Million Pfund für die Veröffentlichung der Daten auf. Dieser Aufwand kann allerdings kaum als reine Publikationskosten deklariert werden, denn die offenen APIs werden inzwischen auch intern als ein primärer Zugang zu den Daten genutzt und haben damit – ganz nebenbei – auch das interne Datenteilen erheblich vereinfacht.<sup>58</sup> Anders als in Finnland<sup>59</sup> sind konzessionierte Transportunternehmen in London zwar nicht dazu verpflichtet, Verkehrs- und Transportdaten, beispielsweise Fahrplandaten, bereitzustellen. Dennoch unterhält Tfl inzwischen mit einigen Tfl-Daten nutzenden Technologieunternehmen, zu denen z. B. Waze, Google, Bosch und Citymapper gehören, Partnerschaften. Durch diese teilen die Partnerunternehmen wiederum eigene Daten, wodurch die Verkehrsbetriebe wichtige Daten in bisher nur unzureichend abgedeckten Bereichen erhalten. Dazu gehört etwa das Echtzeit-Monitoring von Verkehrsflüssen, womit Tfl Nutzendennachfragen präziser abbilden und beispielsweise für optimierte Fahrplanangebote nutzen kann.<sup>60</sup>

Das Prinzip der nicht-rivalisierenden Datennutzung und des innovationsfördernden Wettbewerbs zwischen datengetriebenen Geschäftsmodellen hat Tfl damit entlang der eigenen

<sup>55</sup> In der NaPTAN-Datenbank wird jeder ÖPNV-Haltestelle registriert, an dem Passagiere im Vereinigten Königreich Zugang zu öffentlichen bzw. öffentlich angebotenen Verkehrsmitteln erhalten können. Dazu gehören unter anderem alle Bus- und Bahnhaltstellen, Bus- und Fernbusendhaltestellen, Flug- und Fährhäfen sowie Taxistände. <https://beta-naptan.dft.gov.uk/>.

<sup>56</sup> Deloitte (2017): »Assessing the value of Tfl's open data and digital partnerships«. <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf>.

<sup>57</sup> Transport for London: »Coronavirus publications«. <https://tfl.gov.uk/corporate/publications-and-reports/coronavirus-publications#on-this-page-1>.

<sup>58</sup> Becky Hogge (2016): »Transport for London get set, go!«. <https://odimply.org/files/case-studies-transport-for-london.pdf>.

<sup>59</sup> Ministry of Transport and Communications Finland (2017): »Act on Transport Services«. [https://www.finlex.fi/fi/laki/kaannokset/2017/en20170320\\_20180731.pdf](https://www.finlex.fi/fi/laki/kaannokset/2017/en20170320_20180731.pdf).

<sup>60</sup> Deloitte (2017): »Assessing the value of Tfl's open data and digital partnerships«. <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf>.  
Becky Hogge (2016): »Transport for London get set, go!«. <https://odimply.org/files/case-studies-transport-for-london.pdf>.

WEIL DATEN TECHNISCH ABNUTZUNGSFREI,  
PERFEKT REPLIZIERBAR UND  
NICHT-RIVALISIEREND  
VERWENDBAR SIND, BIETEN  
SIE OPTIMALE VORAUSSETZUNGEN  
FÜR EINE VIELFÄLTIGE UND INTENSIVE NUTZUNG.

strategischen Interessen optimiert. Dass TfL sich dabei weitgehend auf die Rolle des Datenbereitstellers konzentriert und den Wert seiner Investition dadurch steigert, dass optimale Bedingungen zur breiten Nutzung von Daten geschaffen wurden, ist gewissermaßen vorbildhaft. Für die lokale Wirtschaft wurden Daten somit zu einer Subventionsressource, die tatsächlich mit der Gießkanne verteilt werden kann.

Im Gegensatz zum Innovationswettbewerb der Red-Lake-Mine folgt TfL einem datengetriebenen offenen Inside-out-Innovationsprozess. Die Verkehrsbetriebe stellen ihre Ressourcen nicht nur bereit, um mit der Hilfe anderer vorab definierte TfL-eigene Probleme zu lösen, sondern insbesondere, um Dritten zu ermöglichen, eigene Produktideen zu entwickeln.<sup>61</sup> So ist ein Ökosystem entstanden, in dem der Datenveröffentlicher TfL von komplementären Anwendungen indirekt, durch den Aufbau neuer kommerzieller Beziehungen, oder durch direkte Mehreinnahmen im Kerngeschäft profitiert. Neben der restriktionsfreien Lizenzierung und Bereitstellung der Daten ist deren durch die Unified API auf Daten- und Technologieebene garantierte Interoperabilität essenziell, um multimodale und sektorfremde Anwendungen zu ermöglichen.

Dass diese Logik nicht nur im ÖPNV, sondern auch in traditionell deutlich kompetitiveren und proprietärer denkenden Industrien helfen kann, zunächst fehlgeschlagene Investitionen doch noch zu verwerten, stellten die Innovationsforscher

Henry Chesbrough und Eric L. Chen 2015 in einer Studie der Pharmaindustrie fest.<sup>62</sup> Kostentreiber in der pharmazeutischen Produktentwicklung sind vor allem aufgegebene »Fehlentwicklungen«, die nach der Aufgabe im schlechtesten Fall brachliegen. Chesbrough und Chen zeigten mittels Interviews auf, wie Arzneimittelunternehmen wie Roche oder AstraZeneca Daten zu eingestellten Entwicklungen mit Dritten teilen und auf dieser Basis versuchen, bisher unentdeckte Anwendungsfelder zu finden, von deren Kommerzialisierung sie teilweise profitieren können.

Klar ist also: Daten schaffen nur durch Nutzung Wert. Und weil Daten von vielen gleichzeitig genutzt werden können ohne sich abzunutzen, steigt mit einer breiteren Datennutzung auch das gesamtgesellschaftliche Mehrwertpotenzial. Ein Nullsummenspiel ist die Verwertung von Daten durch konkurrierende Nutzungsideen daher folglich nicht. Im Gegenteil: Während das Horten ungenutzter Datenbestände eine garantiert schlechte Alternative mit stets negativen Opportunitätskosten ist, kann zwar eine intensive Datennutzung theoretisch auch zu keinem Mehrwert führen. Allerdings steigert jede weitere Nutzung eines Datums dessen Potenzial, tatsächlich einen Mehrwert zu schaffen, sodass, bei ausreichend breiter Nutzung, die Chancen auf eine positive Opportunitätskostenbilanz gesteigert werden können. Zugespitzt ließe sich sagen: Jede einzelne weitere Nutzung ist eine Chance, die es verdient, verfolgt zu werden. Warum bei der Erreichung dieses Anspruchs der Zugang zu Daten eine zentrale Rolle spielt, davon handelt die nächste These.

## Handlungsempfehlungen

Datenstrategien sollten eine möglichst *breite* Nutzung von Daten als Ziel verfolgen. Weil Daten technisch abnutzungsfrei,

<sup>61</sup> Unterschieden werden nach Chesbrough (s. z.B. die folgende Fußnote) zwei grundsätzliche Modelle des Wissens- und Informationsflusses bei offenen Innovationsprozessen: Im Falle des »outside-in« Innovationsprozesses öffnet eine Organisation ihre Innovationsprozesse für Dritte, um mit deren Ideen interne Neuentwicklungen anzustoßen. Für dieses Modell steht der GoldCorp-Wettbewerb. Im Falle des »inside-out« Innovationsprozesses stellen Organisationen, so wie TfL, Dritten ihre Ressourcen zur Verfügung, damit diese eigene, gegebenenfalls komplementäre Services und Produkte entwickeln können.

Frank T Piller, Joel West (2014): »Firms, users, and innovation: An interactive model of coupled open innovation«. [https://www.researchgate.net/profile/Frank-Piller/publication/297763555\\_Firms\\_users\\_and\\_innovation\\_An\\_interactive\\_model\\_of\\_coupled\\_open\\_innovation/links/5725e2ab08ae586b21def06a/Firms-users-and-innovation-An-interactive-model-of-coupled-open-innovation.pdf](https://www.researchgate.net/profile/Frank-Piller/publication/297763555_Firms_users_and_innovation_An_interactive_model_of_coupled_open_innovation/links/5725e2ab08ae586b21def06a/Firms-users-and-innovation-An-interactive-model-of-coupled-open-innovation.pdf).

<sup>62</sup> Henry Chesbrough, Eric L. Chen (2015): »Using Inside-Out Open Innovation to Recover Abandoned Pharmaceutical Compounds«. *Journal of Innovation Management* Vol. 3 No. 2, S. 21 – 32. [https://journals.ojs3.fe.up.pt/index.php/jim/article/download/2183-0606\\_003.002\\_0005/182/1310](https://journals.ojs3.fe.up.pt/index.php/jim/article/download/2183-0606_003.002_0005/182/1310).



PRODUKTIVITÄTSZUWÄCHSE SIND  
DIE LANGFRISTIG ZENTRALE  
MESSLATTE FÜR DIE  
VOLKSWIRTSCHAFTLICHE  
WIRKSAMKEIT VON INNOVATIONEN.

perfekt replizierbar und nicht-rivalisierend verwendbar sind, bieten sie optimale Voraussetzungen für eine vielfältige und intensive Nutzung. Die Innovationspotenziale von Daten können so optimal gehebelt und maximiert werden. Ein erhöhter Wettbewerb zwischen datenbasierten Ideen, Geschäftsmodellen und Algorithmen ist dabei kein Hinderungsgrund, sondern ein förderlicher Begleiteffekt. Datenstrategien, ob unternehmerisch oder staatlich definiert, sollten folglich aufmerksam evaluieren, wo eine möglichst breite Datennutzung – auch durch Dritte – innovationsstrategische Vorteile bietet, um unternehmerische oder gemeinnützige Zwecke zu verfolgen. Datenstrategien, die von staatlicher Seite auch mit Blick auf Wirtschaft und Unternehmen formuliert werden, sollten in diesem Sinne für alle Dateninteressenten möglichst einheitliche, offene Datenzugangsregeln als Ausgangsbasis für den Datennutzungswettbewerb priorisieren.

## 4.4 INNOVATIONEN BENÖTIGEN ZUGANG ZU MEHR DATEN

Wer den Keynotes aus dem Silicon Valley folgt, wähnt sich schnell in einem Zeitalter radikalen Fortschritts, in dem eine Innovation die nächste jagt. Joseph Schumpeter, könnte man meinen, wäre entzückt. Schließlich legte Schumpeter bereits in der ersten Hälfte des 20. Jahrhunderts mit seiner Idee der schöpferischen Zerstörung die Grundlage für die unter anderem von Clayton M. Christensen entwickelte Theorie disruptiver Innovationen. Für Schumpeter waren die stetige kreative Zerstörung, das Scheitern und die Neuentstehung nicht nur von Produkten, sondern ganzer Industrien ein Wesensmerkmal kapitalistischer Marktwirtschaften. Denn die ständige Ablösung von führenden Unternehmen durch neue Herausforderer bietet Raum für wirtschaftliche Erneuerung und die Verbreitung von Innovationen. Als Motor für Produktivität und damit langfristige Quelle von Wohlstand erlangten die von Schumpeter zunächst primär mikroökonomisch betrachteten Innovationsdynamiken auch volkswirtschaftliche Bedeutung.

Abgesehen von neuen Produkten und Patenten gibt es aber einen grundlegenden Indikator dafür, wie wirksam Innovationen Gesellschaften und einzelne Organisationen vorantreiben. Innovationen, vor allem als Technologien und Produktionstechniken, werden von Unternehmen und anderen Organisationen in der Regel mit der Erwartung eingesetzt, bei gleichbleibendem Ressourceneinsatz mehr Ertrag zu erzielen. Der Wirtschaftsnobelpreisträger Robert Solow zeigte in seiner Forschung die volkswirtschaftliche Bedeutung dieses Zusammenhangs auf. Wirtschaftliches Wachstum, so Solows Erkenntnis, wird langfristig nicht von steigenden Arbeits- oder Kapitaleinsätzen getragen, sondern von der Fähigkeit von Volkswirtschaften, technologische Innovationen hervorzubringen und zu adoptieren.<sup>63</sup> Produktivitätszuwächse, vor allem in der Form von Arbeits- und Kapitalproduktivitätszuwächsen, sind damit also die langfristig zentrale Messlatte für die Wirksamkeit von Innovationen.

Betrachtet man die Daten zum Produktivitätswachstum entwickelter Industrienationen der letzten 50 Jahre, stellt sich die Lage allerdings weniger rosig dar als vermutet. Es zeigt sich nämlich, dass sich dieses Wachstum bereits seit 2005, also noch vor Beginn der Finanzkrise, unterdurchschnittlich entwickelt.<sup>64</sup> Lag das Arbeitsproduktivitätswachstum beispielsweise in Deutschland, dem Vereinigten Königreich, Japan und den USA zuvor überwiegend über 2 Prozent, hat sich dieses seit 2005 relativ stabil um 1 Prozent eingependelt.<sup>65</sup> Wenn wir also seit ungefähr diesem Zeitpunkt in einem Zeitalter radikal

<sup>63</sup> Nobel Prize Outreach AB (1987): »Press Release«.

<https://www.nobelprize.org/prizes/economic-sciences/1987/press-release/>.

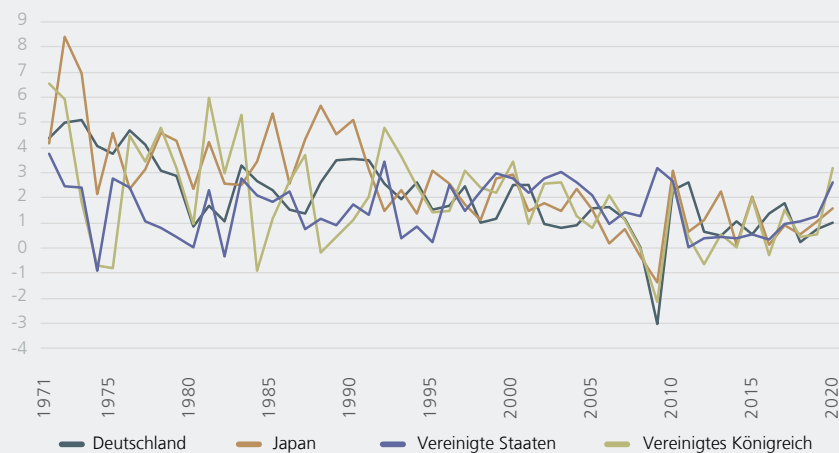
Robert M. Solow (1956): »A Contribution to the Theory of Economic Growth«. The Quarterly Journal of Economics Vol. 70 No. 1, S. 65 – 94.

<https://academic.oup.com/qje/article/70/1/65/1903777>.

<sup>64</sup> Jane Olmstead-Rumsey (2020): »Market Concentration and the Productivity Slowdown«. [https://cpb-us-e1.wpmucdn.com/sites.northwestern.edu/dist/c/5134/files/2020/12/jmp\\_draft.pdf](https://cpb-us-e1.wpmucdn.com/sites.northwestern.edu/dist/c/5134/files/2020/12/jmp_draft.pdf).

<sup>65</sup> Eine deutliche Verlangsamung des Produktivitätswachstums ab 2005 dokumentieren auch Statistiken zur Gesamtfaktorproduktivität. Diese Metrik berücksichtigt neben der Arbeits- auch die Kapitalproduktivität sowie andere laterale Produktivitätsfaktoren.

Abbildung 2: Jährliches Wachstum der Arbeitsproduktivität in Deutschland, UK, Japan und den USA, 1971 – 2020<sup>66</sup>



beschleunigter Innovationen leben, dann scheinen diese auf die langfristig entscheidende, weil mit Wohlstandszuwächsen stark korrelierende Metrik nicht durchzuschlagen.

Was hat das alles mit Schumpeter und mehr noch mit Daten zu tun? Denn natürlich schaffen es einige Unternehmen, namentlich sogenannte digitale Superstarfirmen, deutlich überdurchschnittliche Produktivitätszuwächse zu erzielen.<sup>67</sup> Die eher ernüchternde volkswirtschaftliche Performanz gleichen die Ausreißer aber nicht aus. Schumpeter dürfte deshalb allerdings nicht nur enttäuscht, sondern eher erschrocken sein. 1942 formulierte der Nationalökonom in seinem Spätwerk »Kapitalismus, Sozialismus und Demokratie«<sup>68</sup> eine düstere Prognose: Dass nämlich zunehmende Marktkonzentration, Oligopole und letztlich sogar Monopole den Markteintritt für kleinere Wettbewerber immer schwieriger machen würden. Damit würde der Innovationsmotor erschlaffen und letztlich sogar die Demokratie durch Sozialismus ersetzt. Erhebliche Mitschuld an dieser zu erwartenden Misere hätten Informationsasymmetrien, vor allem in Form unklarer und unzureichender Zugänge zu Wagniskapital für Entrepreneurere.

Dass es nach dem Ende des Zweiten Weltkriegs nicht genau so kam, hat auch mit der Entwicklung der Wagniskapitalindustrie seit Ende der 1940er Jahre zu tun. So unterscheiden sich viele Superstarfirmen heute von unterlegenen Mitsreitern zumindest nicht ursprünglich durch einen fundamental abweichenden Zugang zu Kapital. Auch der Zugang zu Humanressourcen, in Form von Datenanalytist:innen und anderen Spezialist:innen, erscheint nicht als wesentliches Hindernis. Substanziell differenzieren sie sich eher durch ihre datengetriebenen Geschäftsmodelle

und vor allem dadurch, dass sie für diese schneller als andere massenweise Daten sammeln, kuratieren und mit mehr oder weniger guten Algorithmen und Prozessen verwerten.<sup>69</sup> Geht man also von der Hypothese aus, dass für Innovationen heute Kapital, Arbeit *und* Daten erforderlich sind, so sind letztgenannte die tatsächlich knappe Ressource – zumindest für diejenigen, die keinen Zugriff auf große Datensilos haben. Wer im Schumpeter'schen Sinne ein breiteres Innovationssystem mit Aussicht auf letztlich größeres volkswirtschaftliches Produktivitätswachstum möchte, muss diesen Zustand ändern.

Daten möglichst breit und einfach zugänglich zu machen, ist dabei aus mindestens zwei innovations- und produktivitätspolitischen Gründen geboten: Disruptive Innovationen, die auf grundlegend neuen Produkt- oder Servicedesigns aufbauen und damit bestehende Produktionsparadigmen sprengen, kommen nur in Ausnahmefällen von etablierten Marktführern. Wie der bereits zitierte Clayton M. Christensen in seinen Arbeiten zum Innovationsdilemma zeigte, vereiteln strategische Pfadabhängigkeiten und organisationsinterne Anreizstrukturen in etablierten Unternehmen mit hohem Marktanteil oftmals eine disruptive Selbstkannibalisierung.<sup>70</sup> Fortschreitende Marktkonzentration bis hin zu Monopolen trägt zudem marktdynamisch zu nachlassendem Innovationswillen einzelner Unternehmen bei, beispielsweise indem sie sich mangels Wettbewerbs strategisch auf inkrementelle, im Vokabular von Christensen »erhaltende«, und vermeintlich sichere

<sup>66</sup> OECD Data (2021): »Labour productivity and utilisation«. <https://data.oecd.org/lprdty/labour-productivity-and-utilisation.htm>.

<sup>67</sup> David Autor et al. (2019): »The Fall of the Labor Share and the Rise of Superstar Firms«. <https://economics.mit.edu/files/12979>.

<sup>68</sup> Joseph A. Schumpeter (2005): »Kapitalismus, Sozialismus und Demokratie«. 8. Auflage. UTB (Stuttgart).

<sup>69</sup> Maarten De Ridder (2021): »Market Power and Innovation in the Intangible Economy«. <https://www.lse.ac.uk/CFM/assets/pdf/CFM-Discussion-Papers-2019/CFMDP2019-07-Paper-updated.pdf>.

Maximilian Schäfer, Geza Sapi (2020): »Learning from Data and Network Effects: The Example of Internet Search«. [https://www.diw.de/de/diw\\_01.c.798446.de/publikationen/diskussionspapiere/2020\\_1894/learning\\_from\\_data\\_and\\_network\\_effects\\_the\\_example\\_of\\_internet\\_search.html](https://www.diw.de/de/diw_01.c.798446.de/publikationen/diskussionspapiere/2020_1894/learning_from_data_and_network_effects_the_example_of_internet_search.html).

<sup>70</sup> Clayton M. Christensen (1997): »The innovator's dilemma. When new technologies cause great firms to fail«. Harvard Business School Press (Boston).

OPEN-DATA-POLICIES BIETEN  
EIN KONZEPT, EIN MINDESTMAß  
AN WIRTSCHAFTLICHER UND  
GESELLSCHAFTLICHER TEILHABE  
UND AN CHANCEN ZUR  
INNOVATION SICHERZUSTELLEN.

FuE-Investitionen konzentrieren.<sup>71</sup> Das verspricht einerseits stetiges, relativ gut vorhersehbares, wenngleich niedrigeres Gewinnwachstum durch die möglichst lange Erhaltung eigener etablierter Geschäftsmodelle und Produktionsparadigmen. Andererseits schmälert dieses Verhalten aber auch die Chancen auf wirklich umwälzende Neuentwicklungen sowie deutlichere Gewinn- und Produktivitätssprünge. Wenn Datenstrategien folglich Innovationen nicht nur inkrementell fördern wollen, sondern auch neuen Marktakteuren eine Chance geben *und* den Schumpeter'schen Innovationseifer auch der Etablierten anfeuern sollen, tun sie gut daran, einen möglichst breiten und hürdenlosen Datenzugang zu priorisieren.

Neue Entwicklungen nicht durch einseitige Datenknappheit abzuwürgen, ist aber auch volkswirtschaftlich aus der Perspektive des Produktivitätswachstums relevant. Anhand der Elektrifizierung amerikanischer Fabriken zu Beginn des 20. Jahrhunderts zeigte der Wirtschaftshistoriker Paul David von der Stanford University, dass neue Basistechnologien<sup>72</sup> ihr volles Potenzial zur wirtschaftlichen Produktivitätssteigerung erst im Zusammenspiel mit komplementären Innovationen ausschöpfen.<sup>73</sup> Kurz gesprochen eröffneten beispielsweise erst tayloristische und viel später »Lean Manufacturing«-Produktionsprinzipien das volle Produktivitätspotenzial elektrifizierter Fabriken. Dass oftmals sehr vielfältige komplementäre Innovationen im Falle der datengetriebenen Wirtschaft allein oder primär von etablierten Unternehmen erdacht und adoptiert werden, ist kaum zu erwarten. Denn diese arbeiten meist fokussiert in ihren

Sektoren und mit ihren Produkten – seien ihre Datentaschen noch so groß. Folglich gilt es auch, die produktivitätssteigernde Entwicklung von komplementären Innovationen auf viele Schultern zu verteilen. Dazu muss der Zugang zu Daten so weit wie möglich vereinfacht und geöffnet werden.

Politik und öffentliche Verwaltung unternehmen in diesem Zusammenhang mit der Umsetzung von Open-Data-Policies bereits konsequente, wenn auch in Deutschland oftmals ausbaufähige Schritte. Vor allem besitzen sie damit ein Konzept, wie im Zeitalter der Datenökonomie im Sinne einer digitalen Daseinsvorsorge ein Mindestmaß an wirtschaftlicher und gesellschaftlicher Teilhabe und an Chancen zur Innovation sichergestellt werden kann. Mit immer weiter gehenden Verpflichtungen zur Publikation von Forschungsdaten treibt auch die Wissenschaft, teilweise getrieben von Forschungsförderern, die Zugänglichkeit ihrer Daten voran. Fraglich ist allerdings, wie weit diese Bemühungen tragen und wie attraktiv das Datenangebot von Politik, Verwaltung und Wissenschaft aus wirtschaftlicher Sicht ist. Die bisherigen Erfolge scheinen – trotz der essenziellen Rolle offener Daten in der Pandemiebekämpfung – eher auf einen Nischen- denn einen Massenmarkt hinzudeuten. Für einen durchschlagenden wirtschaftlichen Erfolg müssten daher wohl auch Datenbestände zugänglich gemacht werden, die sich aktuell in privaten Händen, vor allem von datenreichen Marktführern, befinden. Die Instrumente dazu können entweder auf Freiwilligkeit oder Pflichten setzen.

Welchen Vorteil das freiwillige Teilen von Daten haben kann, zeigten in den vorigen Abschnitten die Beispiele der GoldCorp Inc. und der Londoner Verkehrsbetriebe. Dem Freiwilligkeitsprinzip folgt auch der Aufbau von Datenräumen in unterschiedlichen Wirtschaftssektoren von der transnationalen bis zur kommunalen Ebene. Diese sollen das Teilen von Daten so attraktiv machen, dass Unternehmen und andere Organisationen freiwillig eintreten wollen. Bei besten Intentionen besteht dabei aber auch die Gefahr, dass in Datenräumen letztlich nur bestehende sektorale Produktionsnetzwerke abgebildet werden. Diese

<sup>71</sup> Jane Olmstead-Rumsey (2020): »Market Concentration and the Productivity Slowdown«. [https://cpb-us-e1.wpmucdn.com/sites.northwestern.edu/dist/c/5134/files/2020/12/jmp\\_draft.pdf](https://cpb-us-e1.wpmucdn.com/sites.northwestern.edu/dist/c/5134/files/2020/12/jmp_draft.pdf).

<sup>72</sup> Basistechnologien, im Englischen General Purpose Technologies, sind universell einsetzbare Technologien, die die Produktionsgrundlagen von Gesellschaften grundlegend verändern. Dazu gezählt werden regelmäßig beispielsweise Dynamos zur elektrischen Stromerzeugung, Informationstechnologien und Big Data.

<sup>73</sup> Paul A. David (1989): »The Modern Productivity Paradox in a Not-Too Distant Mirror«. <https://warwick.ac.uk/fac/soc/economics/research/workingpapers/1989-1994/twerp339.pdf>.



DATENTEILUNGSPFLICHTEN SOLLTEN

ABGESTUFT GESTALTET SEIN,

UM GRÖßERE UND KLEINERE

MARKTTEILNEHMER IN EINEM

ANGEMESSENEN UMFANG

GEGENSEITIG ZU VERPFLICHTEN.

würden Innovationen weiterhin nur inkrementell hervorbringen und letztlich innovationsdynamisch ineffiziente Strukturen eher verfestigen, als diese zu revolutionieren.

Vor allem in hochkonzentrierten Märkten sollten daher zur Förderung eines hürdenfreien Datenzugangs als Ultima Ratio auch Datenteilungspflichten in Betracht kommen. Wie Viktor Mayer-Schönberger und Thomas Ramge in »Machtmaschinen« erläutern, ist die Mandatierung von Datenteilungspflichten keine radikal neue Idee<sup>74</sup>: Um die Marktkonzentration zu reduzieren, verpflichtete die amerikanische Kartellbehörde im Jahr 2011 Google bei Übernahme des Flugbuchungsdienstleisters ITA dazu, den Zugang zu den erworbenen Daten auch für direkte Mitbewerber zu öffnen.<sup>75</sup> Wie essenziell der Datenzugang für die Wertmaximierung der Akquise war, lässt sich daran ablesen, dass Google den Datenzugang unmittelbar nach Ablauf der auferlegten Fünfjahresfrist wieder schloss.

Wie Mayer-Schönberger und Ramge vorschlagen, sollten Datenteilungspflichten abgestuft gestaltet sein, um größere und kleinere Marktteilnehmer in einem angemessenen Umfang gegenseitig zu verpflichten. Größere Marktteilnehmer könnten durchaus mehr (Roh-)Daten teilen müssen als mittelgroße und kleinere Unternehmen, solange dies in beiden Fällen nicht direkt Geschäftsgeheimnisse berührt. Kleinunternehmen könnten bis zu einer bestimmten Größe sogar vollständig von Datenteilungspflichten ausgenommen werden. Für Onlinemarktplätze wie Amazon könnten beispielsweise Daten zu Produktsuchen teilungspflichtig werden. Im Sinne einer abgestuften Gegenseitigkeit könnten aber auch kleinere und mittlere Unternehmen

dazu verpflichtet werden, ab einer gewissen Mindestgröße oder einer bestimmten Menge bezogener Daten eigene Daten mit eben jenen größeren Unternehmen zu teilen, von denen sie selbst Daten nutzen. Konsequenterweise eingesetzt, verwandelt eine solche Datenteilungspflicht Daten von einer hortbaren Ressource in eine Wettbewerbsinfrastruktur, auf deren Basis breitere Innovationen möglich werden.<sup>76</sup> Mit dem GWB-

Digitalisierungsgesetz<sup>77</sup> wurde Anfang 2021 erstmals eine begrenzte Möglichkeit einer Datenteilungspflicht eingeführt. § 20 Abs. 1a GWB erkennt nun die Möglichkeit einer datenbasierten Abhängigkeit zwischen Unternehmen an. Eine Datenteilungspflicht ergibt sich zwar nur, wenn die betreffenden Daten von einem Unternehmen mit Marktmacht kontrolliert werden, nicht auf anderem Weg zu erwerben und schließlich für die Datennutzenden auch unbedingt erforderlich sind.<sup>78</sup> Jedoch können diese Änderungen ein erster Schritt hin zu einer datenbasierten Wettbewerbsinfrastruktur sein.

Umfragen kamen in der Vergangenheit zu dem Ergebnis, dass deutsche Unternehmen mehrheitlich ablehnend gegenüber dem Datenteilen eingestellt sind. In einer 2020 durchgeführten Umfrage unter ca. 500 Unternehmen äußerten rund 88 Prozent der befragten Unternehmen, keine Bereitschaft zum Teilen von Daten zu haben, drei Viertel sahen das Teilen von Daten als

<sup>74</sup> Thomas Ramge, Viktor Mayer-Schönberger (2020): »Machtmaschinen. Warum Datenmonopole unsere Zukunft gefährden und wie wir sie brechen«. Murmann Verlag (Hamburg).

<sup>75</sup> Greg Sivinski, Alex Okuliar, Lars Kjolbye (2017): »Is big data a big deal? A competition law approach to big data«. European Competition Journal Vol. 13 Issue 2-3, S. 199 – 227. <https://www.tandfonline.com/doi/pdf/10.1080/17441056.2017.1362866?needAccess=true>.

<sup>76</sup> Charlotte Ducuing (2019): »Mandating Data Sharing to Establish Data as an Infrastructural Resource«. Network Industries Quarterly Vol. 21 No. 3, S. 21 – 25. <https://www.network-industries.org/wp-content/uploads/2019/09/Mandating-Data-Sharing-to-Establish-Data.pdf>.

<sup>77</sup> »Gesetz zur Änderung des Gesetzes gegen Wettbewerbsbeschränkungen für ein fokussiertes, proaktives und digitales Wettbewerbsrecht 4.0 und anderer Bestimmungen (GWB-Digitalisierungsgesetz) vom 18. Januar 2021«. Bundesgesetzblatt Jahrgang 2021 Teil I Nr. 1. [https://www.bgbl.de/xaver/bgbl/start.xav#\\_bgbl\\_\\_%2F%2F\\*%5B%40attr\\_id%3D%27bgbl121s0002.pdf%27%5D\\_\\_1611043118542](https://www.bgbl.de/xaver/bgbl/start.xav#_bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl121s0002.pdf%27%5D__1611043118542).

<sup>78</sup> Klaus-Heiner Röhl, Lennart Bolwin, Paula Hüttl (2021): »Datenwirtschaft in Deutschland. Wo stehen die Unternehmen in der Datennutzung und was sind ihre größten Hemmnisse?«. <https://www.iwkoeln.de/studien/klaus-heiner-roehl-lennart-bolwin-wo-stehen-die-unternehmen-in-der-datennutzung-und-was-sind-ihre-groessten-hemmnisse.html>.



ESSENZIELL IST, DASS ZUGANGSREGELN

SO EINHEITLICH WIE MÖGLICH

UND MIT SO WENIG SONDERREGELN

WIE NÖTIG KONZEPTIONIERT WERDEN.

nicht erwünscht oder möglich an.<sup>79</sup> Vor diesem Hintergrund sei abschließend auf ein oft vorgetragenes Argument eingegangen: Der Einwand, dass Unternehmen gar nicht mehr in die Erhebung von Daten investieren würden, wenn sie erwarteten, gesammelte Daten teilen zu müssen. Dieses erscheint aus mehreren Gründen wenig überzeugend – und gerade eine Datenteilungspflicht könnte hier zu einem Sinneswandel beitragen. Sofern die betroffenen Daten bisher von den erhebenden Unternehmen nicht genutzt werden, stellt sich die Frage, warum die »Investition« überhaupt getätigt wird. Sollten die entsprechenden Daten nicht einmal einen prospektiven Nutzen für das Unternehmen besitzen und es sich im Angesicht einer Datenteilungspflicht dazu entschließen, diese nicht mehr zu erheben, wäre dies unternehmerisch nur konsequent. Sollten die Daten hingegen entweder bereits genutzt werden oder einen substantiellen, aber noch nicht validierten Nutzen besitzen, dürfte der Druck potenzieller Konkurrenten dazu beitragen, dass Chancen zur Innovation und weiteren Produktiterationen schnell und konsequent verfolgt werden – anstatt in FuE-Abteilungen auf unbestimmte Zeit auf Perfektion zu warten. So gedacht hat ein möglichst offener Datenzugang, auch per Datenteilungspflicht, das Potenzial zum wettbewerbs- und innovationspolitischen Katalysator. Warum neben den formellen Zugangsregeln aber auch die Interoperabilität eine entscheidende Rolle für den tatsächlichen Zugang zu Daten spielt, diskutiert die nächste These.

### Handlungsempfehlungen

Ein zentrales Anliegen von Datenstrategien sollte die Reduzierung von Asymmetrien beim Datenzugang sein. Ziel einer intensivierte Datennutzung sind aus gesellschaftlicher und

volkswirtschaftlicher Sicht nicht einfach Innovationen, sondern es ist langfristig Wohlstand sicherndes Produktivitätswachstum. Dazu ist ein breiterer Wettbewerb zwischen datengetriebenen Ideen und Geschäftsmodellen unumgänglich. Dieser dient dazu, dass Chancen für inkrementelle *und* disruptive Entwicklungen verfolgt und auch Möglichkeiten für komplementäre Innovationen nach dem Vielaugenprinzip entdeckt werden. Zwar unterstützen öffentliche Verwaltung und Politik bereits seit Jahren diese Agenda durch die Veröffentlichung von Open Data, zum wirklichen Durchbruch müssten aber auch Daten privater Unternehmen zugänglicher werden – durch Öffnen oder Teilen. Instrumentell können Datenstrategien dazu auf Freiwilligkeit, wie im Fall von Datenräumen, oder auf Datenteilungspflichten setzen. Letztere könnten als universell bindendes Instrument formuliert werden, sodass unterschiedliche Klassen von Datenbereitstellenden und -nutzenden entsprechend ihren Fähigkeiten verpflichtet werden. Essenziell ist dabei, dass Zugangsregeln so einheitlich wie möglich und mit so wenig Sonderregeln wie nötig konzeptioniert werden. Im besten Fall vereinen und harmonisieren weitreichende einheitliche Regeln bisherige sektorale Zugangsregime oder -praktiken. Sonderregeln können und sollten nur in drei Fällen eingesetzt werden: Erstens, um in spezifischen Fällen besondere Härten betroffener Organisationen oder Sektoren abzufedern, beispielsweise wenn besonders aufwändig zu erhebende Daten betroffen sind. Zweitens, um zu vermeiden, dass liberalisierte Datenzugänge bestehende Ungleichgewichte in der Datennutzung zwischen großen und kleinen Marktakteuren nur verfestigen. Hier können z. B. spezifisch formulierte Zugangsregeln Vorteile für kleine Organisationen oder individuelle Nutzende sicherstellen. Drittens können Sonderregeln dabei helfen, einen adäquaten Umgang mit Daten sicherzustellen, insbesondere um sensible Interessen verschiedener eventuell schutzwürdiger Gruppen auszugleichen. Solche (sektoralen) Zugangsregelungen müssen allerdings kritisch auf ihre tatsächliche Notwendigkeit sowie Kompatibilität und Interoperabilität zu anderen Zugangsregeln überprüft werden. Ziel

<sup>79</sup> Can Azkan (2022): »Anreizsysteme und Ökonomie des Data Sharings«. <https://ieds-projekt.de/wp-content/uploads/2022/03/EDS-Whitepaper.pdf>. Klaus-Heiner Röhl, Lennart Bolwin, Paula Hüttl (2021): »Datenwirtschaft in Deutschland. Wo stehen die Unternehmen in der Datennutzung und was sind ihre größten Hemmnisse?«. <https://www.iwkoeln.de/studien/klaus-heiner-roehl-lennart-bolwin-wo-stehen-die-unternehmen-in-der-datennutzung-und-was-sind-ihre-groessten-hemmnisse.html>.

dessen ist es, im Sinne einer kohärenten strategischen Herangehensweise sicherzustellen, dass der tatsächliche Datenzugang nicht von sektoralen Zufällen abhängt.

## 4.5 DATEN BENÖTIGEN EINE GEMEINSAME SPRACHE

Um den Zugang zu Daten zu verbessern, hat besonders der öffentliche Sektor in der vergangenen Dekade eine milliarden-schwere Investitionswelle losgetreten. Im Bereich der Open Government Data wurden Dateninfrastrukturen in Form zahlreicher Open-(Meta-)Data-Portale geschaffen: Auf europäischer Ebene fällt darunter das Europäische Datenportal, das 2020 mit dem Datenportal der EU-Kommission zusammengelegt wurde.<sup>80</sup> Deutsche öffentliche Verwaltungen haben unter anderem das bundesweite Metadatenportal Govdata.de und Landesportale wie Open.NRW<sup>81</sup> aufgebaut; hinzu kommen kommunale Portale<sup>82</sup> sowie Lösungen einzelner Behörden wie die des Statistischen Bundesamtes<sup>83</sup> oder des Deutschen Wetterdienstes. Für Außenstehende noch unübersichtlicher sind oft Infrastrukturen für Forschungsdaten. Neben spezifischen Angeboten, wie dem bereits erwähnten Covid-19-Datenportal der Europäischen Kommission, gibt es sowohl domänenübergreifende Infrastrukturen, wie die paneuropäische OpenAIRE-Plattform<sup>84</sup>, als auch disziplinspezifische Portale, wie die Ensembl-Genom-Datenbank des European Bioinformatics Institute<sup>85</sup>. Mit der Initiative Nationale Forschungsdateninfrastruktur verfolgt auch Deutschland den Trend zum Aufbau

umfassender domänenagnostischer Dateninfrastrukturen in der Wissenschaft.<sup>86</sup>

Vorwiegend getrieben durch die Wirtschaft schreitet zudem die Entwicklung von Infrastrukturen zum Teilen von Daten immer zügiger voran. Dabei werden Daten nicht für jedermann geöffnet, sondern nur ausgewählten Nutzenden zur Verfügung gestellt, die sich beispielsweise auf einer Plattform oder bei der teilenden Organisation direkt registrieren müssen. Vor allem das Konzept sogenannter Industrial Data Spaces und das europäische Dateninfrastrukturprojekt GAIA-X propagieren den Aufbau sektorübergreifender Datenräume. Diese sollen das Teilen von Daten in ganzen Wirtschaftssektoren, also nicht nur zwischen einzelnen, vorab bekannten Akteuren, vorantreiben. Auch für den öffentlichen Sektor haben Lösungen zum Teilen von Daten zunehmende Bedeutung, z. B. im Kontext von Smart-City-Initiativen oder für die Bewältigung der Coronakrise. Gleichermassen relevant sind dabei sowohl der Datenaustausch mit Unternehmen als auch zwischen öffentlichen Einrichtungen.<sup>87</sup> Mit FIWARE<sup>88</sup> oder dem vom Fraunhofer-Institut FOKUS entwickelten piveau<sup>89</sup> werden für diese Anwendungsszenarien agnostische Datenplattformen aufgebaut, die sowohl für das Öffnen als auch für das Teilen von Daten genutzt werden können.

Die aktuellen Dateninfrastrukturinitiativen erscheinen in der Summe somit ähnlich ambitioniert wie der Turmbau zu Babel. Das ist einerseits gut, denn dem Ziel einer intensiveren, effektiveren Datennutzung arbeitet jedes einzelne dieser Projekte zu. Andererseits gibt es dabei aber auch besondere

<sup>80</sup> <https://data.europa.eu/euodp/en/home>.

<sup>81</sup> <https://open.nrw/open-data>.

<sup>82</sup> <https://www.offenedaten.frankfurt.de/>.

<sup>83</sup> <https://www-genesis.destatis.de/genesis/online>.

<sup>84</sup> <https://explore.openaire.eu/>.

<sup>85</sup> <http://www.ensembl.org/index.html>.

<sup>86</sup> <https://www.nfdi.de/>.

<sup>87</sup> Carla Dietmair (2021): »Privatwirtschaftliche Daten für öffentliche Aufgaben – was ist möglich?«. <https://www.boell.de/de/2021/05/18/privatwirtschaftliche-daten-fuer-oeffentliche-aufgaben-was-ist-moeglich>. Bluebell Drummond, Lorna Christie (2022): »Sharing public sector data«. <https://post.parliament.uk/research-briefings/post-pn-0664/>.

<sup>88</sup> <https://www.fiware.org/>.

<sup>89</sup> <https://www.piveau.de/>.

PRAKTISCHE FALLSTRICKE SIND

DOKUMENTATIONEN IN UNVERSTÄNDLICHEN

FACHJARGONS, NICHT-KOMPATIBLE

VOKABULARE, TAXONOMIEN UND

ONTOLOGIEN SOWIE DIE STRUKTURIERUNG

VON DATEN IN DEN AUSGANGSFORMATEN.

Herausforderungen: Denn ein breiterer Datenzugang wird nicht automatisch dadurch erfolgreich und vollständig sichergestellt, dass vorhandene (Roh-)Daten über neue Portale lizenzrechtlich offen bereitgestellt oder mit niedrighöchlichen Nutzungshürden geteilt werden. Vielmehr müssen Daten auch in Formen und auf Wegen bereitgestellt werden, die es erlauben, dass potenzielle Nutzende mit unterschiedlichen sektoralen Hintergründen für ihren Anwendungsfall relevante Daten zuverlässig identifizieren, verstehen und einsetzen können.

In der Wissenschaft ist dies eine besonders oft diskutierte Forderung, die sich aus zwei auch in anderen Sektoren relevanten Anwendungsfällen ergibt: Für Wissenschaftler:innen ist erstens die Wiederverwendung von Daten Tagesgeschäft, um Studienergebnisse zu replizieren und zu validieren. Allgemein geht es dabei darum, eine Anwendung oder Analyse mit Daten zu replizieren, um sie so auf Richtigkeit zu überprüfen. Der Begriff der wissenschaftlichen Replikationskrise fasst dabei die Beobachtung zusammen, dass Daten, Berechnungsmethoden und Forschungsdesigns jedoch oft so schlecht dokumentiert werden, dass in vielen akademischen Disziplinen weniger als die Hälfte aller Studien replizierbar sind.<sup>90</sup> Bemerkenswert ist dabei, dass Daten zwar technisch und rechtlich zugänglich, gleichzeitig aber so dürftig dokumentiert sind, dass selbst Nutzende aus der gleichen Domäne diese nicht zuverlässig nutzen können.

Der zweite Anwendungsfall dreht sich um die Wieder- bzw. Zweitverwendung von Daten in neuen Studien, insbesondere über wissenschaftliche Disziplinen, Domänen und Sektoren

hinaus. Die Sekundärnutzung von Daten, oftmals durch »Fachfremde«, kann ein wichtiger Erkenntnismotor sein, beispielsweise um Untersuchungen zu triangulieren oder mittels bereichsfremder Daten Phänomene in anderen Domänen aufzudecken.<sup>91</sup> Derartige Ambitionen können jedoch schnell daran scheitern, dass Daten für den neuen »erhebungszweckfremden« Anwendungsfall nicht als relevant erkannt werden, nur für Expert:innen der ursprünglich erhebenden Domäne verständlich dokumentiert oder so strukturiert sind, dass sie nur sehr schwer für »fachfremde« Nutzungszwecke eingesetzt werden können. Stetig wiederkehrende praktische Fallstricke sind neben schlichtweg intransparenten Infrastrukturen unter anderem Dokumentationen in unverständlichen Fachjargons<sup>92</sup>, die Verwendung nicht-kompatibler Vokabulare, Taxonomien und Ontologien<sup>93</sup> sowie die Strukturierung von Daten in den Ausgangsformaten von Backend- bzw. Sensorsystemen<sup>94</sup>.

<sup>91</sup> Der Begriff der Triangulierung beschreibt ein wissenschaftliches Vorgehen, bei dem unterschiedliche Methoden angewandt oder verschiedene Datenquellen ausgewertet werden, um ein vollständigeres Verständnis eines untersuchten Phänomens zu erreichen.

<sup>92</sup> Selbst wenn Nutzende schon wissen, welche Daten sie bei welchem Bereitstellenden suchen, kann das Auffinden der gewünschten Daten aufgrund der Undurchsichtigkeit von Bereitstellungsinfrastrukturen eine komplexe Aufgabe sein. Erschwert wird dies beispielsweise dadurch, dass Datensätze mit fachjargon- oder sogar bereitstellendenspezifischen Titeln versehen werden und nur eine titelbasierte Suche angeboten wird. Regelmäßig werden zudem verfügbare Daten nicht über einen webseitenbasierten, durchsuchbaren Katalog dokumentiert, sondern in tabellenlastigen PDF-Dokumenten aufgelistet. Nutzende müssen Daten dann z.B. einzeln über unübersichtliche und veraltete FTP-Dateiserver oder sogar per E-Mail beziehen.

<sup>93</sup> In vielen Fällen ist es grundsätzlich möglich, Vokabulare, Taxonomien und Ontologien in sogenannten »Cross-Mapping-Verfahren« auf andere Vokabulare, Taxonomien oder Ontologien abzubilden. Allerdings haben diese Verfahren zugleich meist logische Grenzen, beispielsweise, weil gewünschte Informationen mit den übertragenen Daten nicht in der gewünschten Granularität des Zielformates abbildbar sind.

<sup>94</sup> Besonders problematisch ist es, wenn derartig bereitgestellte Datensätze von Nutzenden nicht nur ein Verständnis der Daten selbst sowie des ursprünglichen Erhebungszwecks erfordern, sondern auch noch Kenntnisse technischer Spezifika der Erhebungssysteme (wie z.B. IoT-Sensoren oder Backend-Kommunikationssysteme).

<sup>90</sup> Richard A. Klein et al. (2017): »Many Labs 2: Investigating Variation in Replicability Across Sample and Setting«. <https://osf.io/ux3eh/>.

Colin F. Camerer (2018): »Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015«. *Nature Human Behaviour* Vol. 2 Issue 9, Abstract / S. 637 – 644.

<https://www.nature.com/articles/s41562-018-0399-z>.

Open Science Collaboration (2015): »Estimating the reproducibility of psychological science«. *Science* Vol. 349 Issue 6251, Abstract / S. 4716-1 – 4716-8. <https://www.science.org/doi/10.1126/science.aac4716>.



Die mit diesen Anwendungsszenarien verbundenen Anforderungen hat die Wissenschaftscommunity prägnant mit dem Akronym *FAIR* umschrieben: Demnach müssen Daten zunächst auffindbar (*findable*) und zugreifbar (*accessible*) sein. Für die Auffindbarkeit von Daten sind Dateninfrastrukturen, beispielsweise in Form von Datenportalen, ein zentrales Instrument. Sie ermöglichen die Bereitstellung von Daten in strukturierter, auffindbarer Form, z. B. unterstützt durch anwendungsspezifische Suchmechanismen und die Speicherung von Daten in vernetzten, verteilten Repositorien. Ebenso sind Dateninfrastrukturen ein generell geeignetes Mittel, um die Zugreifbarkeit von Daten für Berechtigte sicherzustellen, idealerweise über einheitliche Zugangsprozesse und transparent dokumentierte Schnittstellen.<sup>95</sup>

Darüber hinaus müssen Daten nach den sogenannten FAIR-Prinzipien aber auch interoperabel (*interoperable*) und reproduzierbar (*reproducible*) sein. Aus Sicht von Datennutzung fördernden Datenstrategien sind diese Eigenschaften besonders wichtig, denn sie sorgen dafür, dass Daten auch tatsächlich durch unterschiedliche Nutzende und Nutzendengruppen nutzbar sind. Sie stellen, um im Bild des Turmbaus zu Babel zu bleiben, die Verständlichkeit von Daten durch unterschiedliche Nutzendengruppen, ähnlich einer gemeinsamen Sprache, sicher. In der Praxis wird eine hohe Reproduzierbarkeit auf Datenseite dadurch erreicht, dass Daten nicht nur vollständig sind, sondern auch ausreichend detailliert, nutzungsfreundlich und für Neulinge verständlich beschrieben werden. Dies wird einerseits erreicht durch eine logische, konsistente und transparente Struktur der betreffenden Datensätze oder -quellen selbst.<sup>96</sup> Darüber hinaus müssen die betreffenden Datensätze aber auch durch gut verständliche, die wesentlichen Charakteristika

beschreibende Metadaten ergänzt werden. Für das Ziel einer breiten Datennutzung sind qualitativ hochwertige Metadaten besonders wichtig, findet doch der erste Kontakt zu bisher unbekanntem Daten durch neue Nutzende in der Regel über Metadaten statt. Ob ein Datensatz überhaupt als relevant erkannt und für eine mögliche Nutzung in Betracht gezogen wird, kann sich daher bereits beim Blick auf die Metadaten entscheiden. Vor diesem Hintergrund sind vor allem sektorübergreifend anwendbare Metadatenstandards, wie beispielsweise DCAT-AP<sup>97</sup> und Dublin Core<sup>98</sup>, und Handreichungen zur Bereitstellung hochqualitativer Metadaten relevant.<sup>99</sup> Datenstrategien sollten daher die verständliche Bereitstellung hochwertiger Metadaten als essenzielles Kriterium für die tatsächliche Nutzbarkeit von Daten hervorheben.

Von ebenso entscheidender Bedeutung ist darüber hinaus die Interoperabilität: Damit ist vordergründig die technische Interoperabilität im Sinne der Nutzung nicht-proprietärer offener Datei- und Schnittstellentypen gemeint. Kritischer, weil nicht selten mit latenten Interessenkonflikten behaftet, ist jedoch die semantische Interoperabilität von Schnittstellen, Datenformaten und Datenstrukturen (vor allem Taxonomien und Ontologien). Im Kern geht es dabei um die Frage, welches Wissen über einen Sachzustand ein Datensatz oder eine Datenquelle abbilden soll, wie dieses Wissen über spezifische Datenformate und -strukturen repräsentiert werden soll und welche Art von Datenabfragen als besonders relevant gelten. Gerade bei der Umsetzung von Dateninfrastrukturen für unterschiedliche Nutzendengruppen, wie z. B. spezifischen Datenräumen, bieten

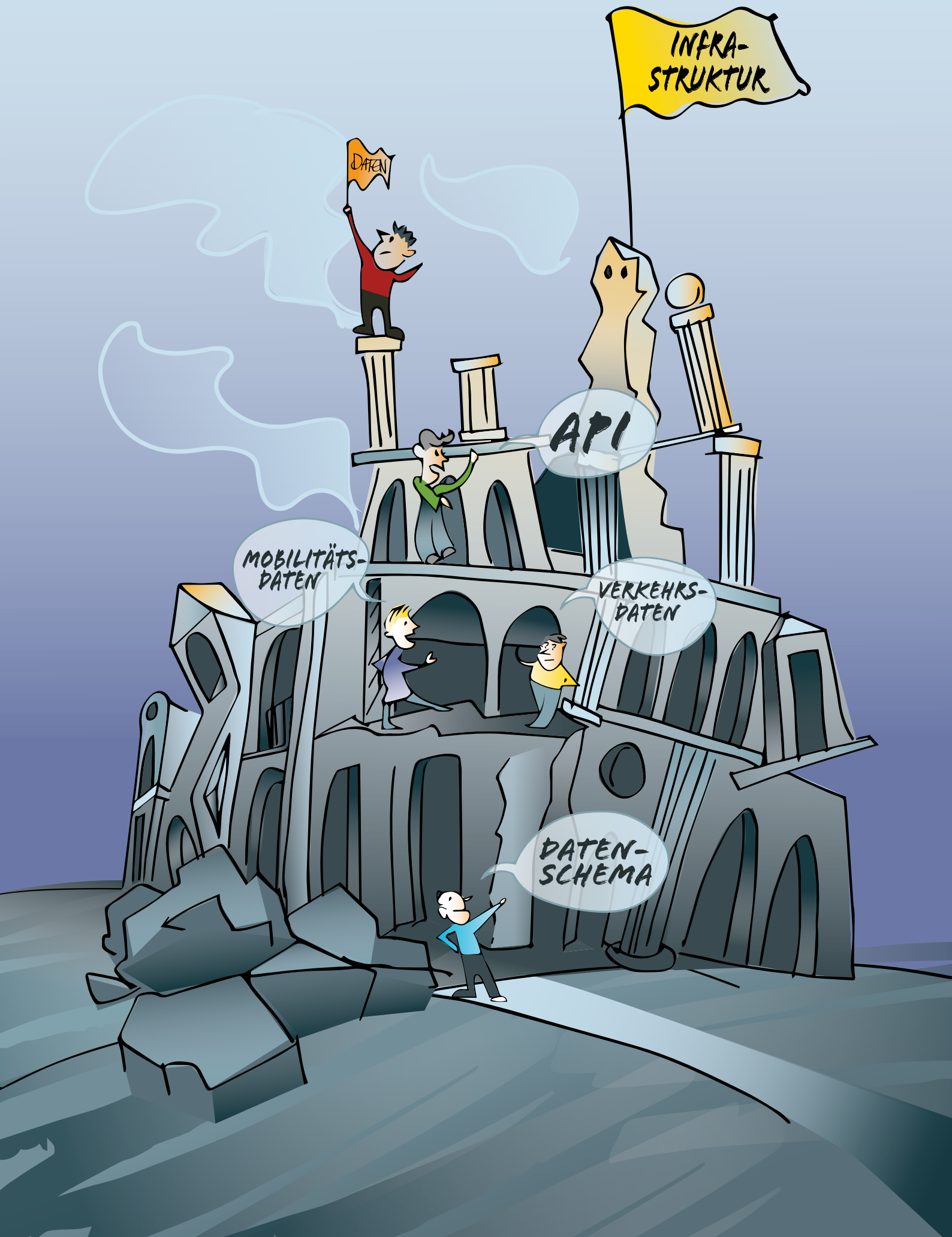
<sup>95</sup> Neben Daten kann gerade die Auffindbarkeit von Programmierschnittstellen durch deren Integration in Dateninfrastrukturen erheblich profitieren.

<sup>96</sup> In der Praxis bedeutet dies beispielsweise, dass die Spalten einer Tabelle mit Titeln versehen sind und Formatierungen, z. B. von Datumsangaben, einem konsistenten Format folgen.

<sup>97</sup> [https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe\\_en](https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en). DCAT-AP basiert dabei auf dem Data Catalog Vocabulary (DCAT) von Riccardo Albertoni (2020): »Data Catalog Vocabulary (DCAT) - Version 2«. <https://www.w3.org/TR/vocab-dcat-2/>.

<sup>98</sup> <https://www.dublincore.org/>.

<sup>99</sup> Lina Bruns, Benjamin Dittwald, Fritz Meiners (2019): »Leitfaden für qualitativ hochwertige Daten und Metadaten«. [https://cdn0.scrvt.com/fokus/e472f1bf447f370f/32c99a36d8b3/NQDM\\_Leitfaden-f-r-qualitativ-hochwertige-Daten-und-Metadaten\\_2019.pdf](https://cdn0.scrvt.com/fokus/e472f1bf447f370f/32c99a36d8b3/NQDM_Leitfaden-f-r-qualitativ-hochwertige-Daten-und-Metadaten_2019.pdf).



DIE SEMANTISCHE INTEROPERABILITÄT  
VON DATENINFRASTRUKTUREN IST  
EINE ENTSCHIEDENDE  
INNOVATIONSPOLITISCHE FRAGE  
FÜR DATENSTRATEGIEN.

diese Fragen erhebliches Konfliktpotenzial, denn je nach gewählter Lösung können spezifische Datennutzungen leichter oder schwieriger umsetzbar werden. Folglich stehen sich in diesen Fällen regelmäßig die Interessen zweier Nutzengruppen entgegen: Auf der einen Seite etablierte Datennutzende, oftmals Domänenexpert:innen, die ihre bestehenden Datenanwendungen optimieren und inkrementell, d.h. zumeist unter Verwendung weiterer domänenspezifischer Daten, weiterentwickeln möchten. Dafür fordern sie, den Fokus auf etablierte, domänenspezifische Lösungen zu legen und diese, wo nötig, schrittweise weiterzuentwickeln. Für etablierte Datennutzende bietet dieses Vorgehen den unmittelbaren Vorteil, dass Daten für ihre teils sehr spezifischen Anwendungen in einer passgenauen Struktur vorliegen. Weniger angesprochen werden davon jedoch oftmals domänen- oder sektorfremde Datennutzende, häufig Markt- oder Domänenneulinge, die Anwendungen basierend auf Daten aus teils sehr unterschiedlichen Domänen entwickeln möchten. Entscheidender als hochspezialisierte Datenformate und Schnittstellen sind für diese Nutzenden meist Lösungen, die ein schnelles Verständnis, den einheitlichen Abruf sowie eine einfache Kombinierbarkeit von Daten aus unterschiedlichen Domänen und Dateninfrastrukturen (oder -räumen) ermöglichen.

Diesem Verlangen folgte übrigens die Londoner Verkehrsbehörde TfL beim Aufbau ihrer Unified API. Über sehr unterschiedliche Datenquellen und Systeme wurden dabei verwendete Vokabulare über eine Taxonomie harmonisiert, redundante Daten aus Backendsystemen bereinigt und Schnittstellenabrufe vereinheitlicht. Im Mittelpunkt stand dabei, die Verständlichkeit und Nutzbarkeit der Schnittstelle und der Daten nicht nur für vertraute Expert:innen, sondern auch für Neulinge sicherzustellen.<sup>100</sup>

Die semantische Interoperabilität von Dateninfrastrukturen ist damit eine entscheidende innovationspolitische Frage für Datenstrategien: Wenn diese eine erweiterte Datennutzung und domänenübergreifende Innovationen befördern sollen, können sie nicht nur auf bestehende domänenspezifische Datenformate, -strukturen und Schnittstellen setzen. Mindestens ebenso wichtig ist, dass für Taxonomien, Ontologien und Schnittstellen auch domänenagnostische Lösungen entwickelt und als Standards etabliert werden. Der DCAT-AP-Metadatenstandard<sup>101</sup>, Dublin Core<sup>102</sup>, die W3C-Prov-Ontologie<sup>103</sup> zur Erfassung von Provenienzinformatoren oder die Open Digital Rights Language<sup>104</sup> sind Beispiele für Lösungen, die dabei helfen, dass Daten eine gemeinsame Sprache sprechen. Standardisierte Schnittstellen mit parametrisierten hierarchischen Schnittstellendefinitionen können zudem eine einheitliche Funktionsweise von Schnittstellen sicherstellen. Sie ermöglichen, dass nicht nur die Daten mehrerer Schnittstellen standardisiert sind, sondern auch, dass grundlegende Datenabrufe einheitlich sind. Damit können Abrufe problemlos, d.h. ohne detaillierte Kenntnis etwaiger Schnittstellenspezifika, durch sektorfremde Systeme und Nutzende durchgeführt werden.

Für ambitionierte strategische Visionen wie den digitalen Binnenmarkt der EU sollten agnostische Lösungen statt sektoraler Speziallösungen sogar eine der höchsten Prioritäten darstellen: Denn erst wenn Nutzende vergleichsweise einfach über zahlreiche Plattformen und Infrastrukturen Daten finden und beziehen können, können sich Netzwerkeffekte für die Nützlichkeit

<sup>100</sup> Transport for London (2020): »Unified API«. <https://tfl.gov.uk/info-for/open-data-users/unified-api>.

<sup>101</sup> [https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe\\_en](https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en).

<sup>102</sup> <https://www.dublincore.org/>.

<sup>103</sup> Paul Groth, Luc Moreau (2013): »An Overview of the PROV Family of Documents«. <https://www.w3.org/TR/prov-overview/>.

<sup>104</sup> Renato Iannella, Serena Vilata (2018): »ODRL Information Model 2.2«. <https://www.w3.org/TR/odrl-model/>.

OHNE ANGEPASSTE ORGANISATIONEN

LASSEN SICH DIE MEHRWERTE

NEUER TECHNOLOGIEN NUR

SCHLECHT REALISIEREN.

und Potenziale der kompatiblen Dateninfrastrukturen einstellen. Beides trägt letztlich dazu bei, dass das Gesamtsystem von Dateninfrastrukturen mehr ist als die Summe seiner Einzelteile.<sup>105</sup>

### Handlungsempfehlungen

Neben geeigneten Lizenzbedingungen und für offene bzw. geteilte Daten geeigneten Distributionskanälen ist die Zugänglichkeit von Daten in der Realität auch von deren technischer und semantischer Verständlichkeit abhängig. Datenstrategien sollten folglich, in Anlehnung an die FAIR-Prinzipien, die Entwicklung und Anwendung von, idealerweise standardisierten, Taxonomien, Ontologien und Schnittstellen priorisieren. Diese sollten für eine möglichst große Anzahl von Datennutzenden verständlich und praktisch nutzbar sein. Sektorspezifische Lösungen, insbesondere in Form von Datenformaten, -strukturen und Schnittstellen, sind zwar in der Regel ideal auf bereits existierende Nutzungsszenarien und Nutzendengruppen ausgerichtet. Neue Anwendungsfelder lassen sich aber mit diesen mitunter nur schwer erschließen, weil sie für neue Nutzendengruppen oftmals schwer verständlich sind. Um eine breitere Datennutzung zu ermöglichen, sollte daher die Entwicklung und Umsetzung sektorübergreifender Lösungen Priorität besitzen. Die Eignung sektoraler Regelungen und spezifischer Formen der Datenbereitstellung für breitere Nutzendengruppen sollte kritisch geprüft werden. Einen Konsens für vereinfachende bzw. abstrahierende Lösungen zu erreichen, beispielsweise für Vokabulare oder Metadatenformate, kann unter Sektorexpert:innen

schwierig sein. In entsprechende Beratungen sollten daher auch Repräsentant:innen prospektiver Nutzendengruppen einbezogen werden. Ultimatives Ziel einer strategischen Koordinierung von Dateninfrastrukturprojekten sollte es schließlich sein, mehr und diversere Datennutzende anzuziehen. Das gelingt schwerlich, wenn Dateninfrastrukturen und Daten für diese Nutzenden unverständlich bleiben.

## 4.6 DATENGETRIEBEN ZU SEIN, VERÄNDERT ORGANISATIONEN

Nach den vorangehenden systemischen Betrachtungen fragen sich einige Lesende vielleicht: Was ist die Rolle von Organisationen für Datenstrategien? Unternehmen, Verwaltungen und zivilgesellschaftliche Initiativen bündeln Wissensressourcen, Arbeitskräfte und Kapital. Mit dem richtigen Rüstzeug sind sie auch im Datenzeitalter das zentrale Instrument, um Innovationen voranzutreiben und deren Produktivitätsversprechen einzulösen. Was es dazu braucht, ist aus datenstrategischer Sicht vor allem die Fähigkeit zur Veränderung von Arbeitsweisen und zur flexiblen Anpassung von Zielsetzungen. Denn ohne angepasste Organisationen lassen sich die Mehrwerte neuer Technologien nur schlecht realisieren.

Warum dieser scheinbare Gemeinplatz so bedeutend ist, verdeutlicht der Blick auf eine heute kaum wegzudenkende, aber anfänglich enttäuschende Technologie: die Elektrizität. Bereits im letzten Drittel des 19. Jahrhunderts entwickelten Thomas Edison, Werner von Siemens und Sir Charles Wheatstone mit elektrischen Dynamos, Glühlampen und den ersten Elektrizitätskraftwerken eine Reihe an Innovationen, mit denen die Elektrizität rein technisch betrachtet massentauglich war. Enthusiastische Industriepioniere investierten ab den 1880er Jahren in elektrische Antriebe, um die Dampfmaschinen als Antriebsquelle in ihren Fabriken abzulösen. Der erhoffte Produktivitätsaufschwung blieb jedoch aus. Die meisten Industrieunternehmen waren daher zunächst skeptisch gegenüber der neuen

<sup>105</sup> Zu verschiedenen Aspekten von Open Data siehe folgende Publikationen: Lisa Bieker et al. (2019): »Open Data zwischen Wunsch und Wirklichkeit«. <https://cdn0.scrvt.com/fokus/cab5baab79c99a8cf1ca63bda8d5/Open-Data--zwischen-Wunsch-und-Wirklichkeit.pdf>. Mario Wiedemann, Mila Frerichs (2020): »Der Musterdatenkatalog: Ein Beitrag für mehr Open Data in Kommunen«. <https://www.oeffentliche-it.de/-/der-musterdatenkatalog-ein-beitrag-fuer-mehr-open-data-in-kommunen>. Jens Klessmann, Torsten Staab (2018): »Strategische Bereitstellung offener Verwaltungsdaten«. <https://www.oeffentliche-it.de/documents/10181/14412/Strategische+Bereitstellung+offener+Verwaltungsdaten>.

ZUR VORLETZTEN JAHRHUNDERTWENDE  
MUSSTE ERST EINE GANZE  
GENERATION VON MANAGERN  
ABGELÖST WERDEN, DAMIT DIE  
VORTEILE ELEKTRISCHER ANTRIEBE  
BESTMÖGLICH GENUTZT WERDEN KONNTEN.

Technologie und setzten weiter auf Dampfkraft. So kam es, dass noch im Jahr 1900 elektrische Motoren nur fünf Prozent der in US-amerikanischen Fabriken benötigten mechanischen Energie lieferten – und es noch bis 1910 dauerte, bis sich diese Situation nennenswert zu verändern begann.<sup>106</sup> Spürbar entwickelte sich das Produktivitätswachstum in den damals führenden Industrieländern sogar erst in den 1920er und 1930er Jahren, um sich dann allerdings gleich mit Raten um fünf Prozent zu verdoppeln oder sogar zu verdreifachen.<sup>107</sup>

Ein praktischer Blick zeigt, was in der Zwischenzeit passiert war. Der bereits erwähnte Wirtschaftshistoriker Paul David analysierte die Aufzeichnungen amerikanischer Fabriken, um festzustellen, wie elektrische Motoren dort bis kurz nach der Jahrhundertwende eingesetzt wurden.<sup>108</sup> Die Antwort ist rückblickend verblüffend: wie Dampfmaschinen.<sup>109</sup> Fabrikanten kauften regelmäßig die leistungsfähigsten Elektromotoren, platzierten sie an die Stelle der Dampfmaschine und verbanden sie mit dem bereits bestehenden labyrinthischen System von Antriebswellen, Schäften und Riemen aus Fabriken der Jahrhundertwende.

<sup>106</sup> Tim Harford (2017): »Why didn't electricity immediately change manufacturing?«. <https://www.bbc.com/news/business-40673694>.

<sup>107</sup> Harald Edquist, Magnus Henrekson (2004): »Technological breakthroughs and productivity growth«. SSE/EFI Working Paper Series in Economics and Finance No. 562. <https://www.econstor.eu/bitstream/10419/56150/1/39295690X.pdf>.

<sup>108</sup> Paul A. David (1989): »The Modern Productivity Paradox in a Not-Too Distant Mirror«. <https://warwick.ac.uk/fac/soc/economics/research/workingpapers/1989-1994/twerp339.pdf>.

<sup>109</sup> Der Einwand, dass diese Art des Einsatzes von Elektromotoren aufgrund deren anfänglicher Kosten, Größe oder Verfügbarkeit wenig verwunderlich sei, überzeugt vor dem Hintergrund der Dauer des Transformationsprozesses nicht. Dieser zog sich selbst in führenden Industrien über ca. 30 Jahre und bei Nachzügler bis zu 50 Jahre hin. Dementgegen sanken die Kosten für Elektromotoren deutlich schneller, deren Miniaturisierung ging zügig voran und auch die Verfügbarkeit stieg stetig. Als alleinige Hinderungsgründe für eine massenhafte und produktivitätsfördernde Adaption von Elektromotoren taugen die genannten Faktoren daher in der Regel bis etwa um die Jahrhundertwende und nur in Ausnahmefällen länger, beispielsweise aufgrund hoher pfadabhängiger Investitionen in dampfenergiegetriebene Produktionsstätten und -abläufe.

Harald Edquist, Magnus Henrekson (2004): »Technological breakthroughs and productivity growth«. SSE/EFI Working Paper Series in Economics and Finance No. 562. <https://www.econstor.eu/bitstream/10419/56150/1/39295690X.pdf>.

Dieses System einer einzigen Antriebsquelle erforderte unter anderem, dass Maschinen mit einem hohen Bedarf mechanischer Energie nah an der (Dampfmaschine als) Antriebsquelle platziert wurden. Andernfalls bestand die Gefahr, dass überlange Antriebswellen den Verwindungskräften nicht standhielten oder aber der Energieverlust bei langen Übertragungswegen zu groß war. In der Folge wurden Produktionsmaschinen nicht nur neben, sondern auch ober- und unterhalb der Dampfmaschine aufgestellt. Fabriken der Zeit hatten daher oft mehrere Stockwerke, über die teils schwere Einzelteile im Verlaufe des Produktionsprozesses immer wieder hinauf und hinunter bewegt werden mussten. So eingesetzt waren Elektromotoren zwar immer noch weniger laut und dreckig als Dampfmaschinen, die Effizienz steigerten sie aber ebenso wenig wie den Fabrikausstoß.<sup>110</sup>

Laut Erik Brynjolfsson und Andrew McAfee musste erst eine ganze Generation von Managern abgelöst werden, damit 30 Jahre später Fabriken so gebaut wurden, wie wir sie heute kennen – und wie die Vorteile elektrischer Antriebe am besten nutzbar sind<sup>111</sup>: Als Gebäude, die eine nach dem Materialfluss aufgereichte, meist einstöckige Produktionslinie behausen, die wiederum gespickt ist mit vielen kleineren stromgetriebenen Maschinen statt mit einem großen Zentralmotor.<sup>112</sup> Noch entscheidender war jedoch, dass im Schatten der architektonischen Veränderung und des Personalwechsels auch neue Produktionsprozesse im Fabrikalltag einzogen. Bereits ab ca. 1914

<sup>110</sup> Harald Edquist, Magnus Henrekson (2004): »Technological breakthroughs and productivity growth«. SSE/EFI Working Paper Series in Economics and Finance No. 562. <https://www.econstor.eu/bitstream/10419/56150/1/39295690X.pdf>.

<sup>111</sup> Erik Brynjolfsson, Andrew McAfee (2016): »The second machine age. Work, progress, and prosperity in a time of brilliant technologies«. Norton & Company (New York / London).

<sup>112</sup> Vor allem die Miniaturisierung von Elektromotoren sowie deren ab Beginn der 1920er Jahre besonders rapide fallenden Kosten bereiteten den Boden für diese Veränderungen. Letztlich entscheidend für die massenhafte Adaption neuer Fabrikgrundrisse und -designs war jedoch die Reorganisation von Produktionsprozessen basierend auf modernen Managementtheorien.

ORGANISATIONEN MÜSSEN

STRUKTURELL SOWIE AUF

TEAM- UND MITARBEITENDENEBENE

FLEXIBLER WERDEN.

zählten dazu die bei Ford eingesetzten tayloristischen bzw. fordistischen Produktionsprozesse sowie, ab Mitte der 1930er Jahre, von Toyota eingeführte Produktionsmethoden, die später die Namen Just-in-Time bzw. Lean-Manufacturing erhalten sollten. Dass die Lehre vom Erfolg der angepassten Organisation auch im digitalen Zeitalter trägt, verdeutlichen zahlreiche Firmen, die um die Jahrtausendwende bereits bekannte Informationstechnologien und Innovationen in ihren Geschäftsprozessen rekombinierten. Sie stellten dabei wesentliche Prozesse zwischen Lieferanten, Produzenten und Endkund:innen neu auf und erzielten damit Produktivitätszuwächse, die jene ihrer Mitbewerber deutlich übertrafen.<sup>113</sup> Beispiele sind die Anwendung von Vendor-managed Inventories (lieferantengesteuerte Bestandssysteme) und Enterprise-Resource-Planning-Software sowie deren Verknüpfung mit internetbasierten Vertriebsfrontends als direkten Kundenzugängen. Die US-Firmen Walmart und CVS etwa setzten diese Systeme frühzeitig ein und etablierten dafür im Zuge der 1990er Jahre erfolgreich eigene Datenteilungsprozesse mit Lieferanten.<sup>114</sup>

Die seit gut einer Dekade sprunghaft wachsenden globalen Datenbestände und die Verfügbarkeit von Technologien zu deren Analyse und Verwertung stellen Organisationen vor ähnliche strategische Herausforderungen wie die Elektrifizierung und die erste Welle der Digitalisierung zwischen den 1980er und frühen 2000er Jahren. Welche Prozess- und Organisationsformen sich für datengetriebene Innovationen am Ende als Produktivitätsbooster erweisen, ist bisher noch offen. Es zeichnen sich aber

bereits drei Entwicklungen ab, auf die Datenstrategien einen besonderen Fokus legen sollten:

Erstens werden Technologien immer wichtiger, die den Datenaustausch und die kollaborative Datennutzung erleichtern, ob zwischen unterschiedlichen Teams in einer Organisation oder mit externen Partnern. Der Erfolg von oftmals Cloud-basierten Datenplattformen sowie von Wikisystemen verdeutlicht diese Entwicklung. Für die optimale Arbeitsweise und Struktur von datengetriebenen Organisationen hat dies Folgen. In Zeiten, in denen vor allem die Grenzkosten zur Verbreitung von Daten innerhalb von Organisationen gegen null gehen, ist es sinnlos, an funktional oder organisatorisch begründeten Datensilos und Zugangshierarchien festzuhalten.<sup>115</sup> Für schlecht skalierende Datenverbreitungs-, Zugangs- und Analyseprozesse im Archiv- oder Großrechnerzeitalter stellten Silos und Hierarchien vertretbare Kosten-Nutzen-Abwägungen dar. Im Cloudzeitalter sind sie aber nicht mehr angemessen. Stattdessen müssen, zumindest innerhalb von Organisationen, Datenzugänge und Analysekapazitäten so weit wie möglich demokratisiert, d. h. vereinfacht und für mehr Nutzende freigegeben werden. Wie im Rahmen des Forschungsprojektes Qualifica Digitalis des IT-Planungsrates diagnostiziert wurde, benötigen gerade öffentliche Verwaltungen dazu technische Werkzeuge, vor allem um Daten kollaborativ auszuwerten und zu kombinieren.<sup>116</sup> Um die Datennutzung tatsächlich zu fördern, ist es also essenziell, dass Nutzende in Organisationen, die ihre Datennutzung intensivieren wollen, ein (gegebenenfalls nutzendengruppenspezifisches) Gesamtpaket aus verfügbaren Daten und Analyseinstrumenten vorfinden.

<sup>113</sup> Erik Brynjolfsson, Lorin M. Hitt (2003): »Computing Productivity: Firm-Level Evidence«. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=290325](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=290325).

Zu ähnlichen Ergebnissen kommen auch Nicholas Bloom (Stanford University), Raffaella Sadun (Harvard University) und John van Reenen (London School of Economics); sie führen dabei aber die höhere Performanz US-amerikanischer Unternehmen im Vergleich zu europäischen Konkurrenten auf ein strafferes Personalmanagement zurück: Nicholas Bloom, Raffaella Sadun, John van Reenen (2012): »Americans Do IT Better: US Multinationals and the Productivity Miracle«. *American Economic Review* Vol. 102 Issue 1, S 167 – 201. <https://www.hbs.edu/ris/download.aspx?name=Americans%20Do%20IT%20Better.pdf>.

<sup>114</sup> »Welcome to CVS Suppliers«. <https://cvssuppliers.com/>.

<sup>115</sup> Ausgenommen hiervon sind sensible Datenbestände, die beispielsweise aus Privatsphäregründen oder weil es sich um Datenbestände mit Geschäftsgeheimnissen handelt nur sehr begrenzt verfügbar sein sollen.

<sup>116</sup> Juliane Schmeling, Lina Bruns (2020): »Kompetenzen, Perspektiven und Lernmethoden im digitalisierten öffentlichen Sektor«. [https://qualifica-digitalis.de/wp-content/uploads/QD\\_Metastudie\\_20201005\\_barrierefrei\\_v5.pdf](https://qualifica-digitalis.de/wp-content/uploads/QD_Metastudie_20201005_barrierefrei_v5.pdf).



RELEVANTE INNOVATIONEN,  
OB AUF DATEN ODER AUF  
TECHNOLOGIEN ZUR  
DATENVERWERTUNG BASIEREND,  
KÖNNEN AUS UNTERSCHIEDLICHSTEN  
ECKEN KOMMEN.

Zweitens müssen Organisationen strukturell sowie auf Team- und Mitarbeitenebene flexibler werden. Dazu benötigen sie einerseits Arbeitsstrukturen, die es Teams ermöglichen, verfügbare Datenressourcen möglichst dezentral und schnell zu nutzen, um damit Entwicklungs- und Produktideen zügiger als bisher zu testen. Die breite Umsetzung von SCRUM- oder Lean-Start-up-Prozessen in privaten und öffentlichen Organisationen spricht für die sektoragnostische, universelle Anwendbarkeit dieser Managementtechniken. Organisationen selbst werden dadurch nicht unbedingt dezentral, aber aus geschlossenen werden offene Hierarchien, die ein flexibleres Arbeiten ermöglichen. Damit solche Strukturen andererseits auch tatsächlich tragbar werden, müssen auf Team- und Mitarbeitenebene aber auch erweiterte Fähigkeiten zur Interaktion und interdisziplinären Zusammenarbeit gefördert werden. Gerade für öffentliche Verwaltungen stellt der Wandel von vergleichsweise starren, hierarchischen Strukturen zu flexibilisierten Teams eine deutliche Veränderung dar. Neben formalen Kompetenzen im agilen Management müssen bei Mitarbeitenden Transformations- und Orientierungskompetenzen entwickelt werden. Dazu gehören ebenso die Fähigkeit, Veränderungsprozesse zu verstehen und erlernte Kenntnisse auf andere Kontexte anzuwenden, wie auch Anpassungsfähigkeit, Veränderungsbereitschaft und Weiterbildungswille.<sup>117</sup>

Drittens zeigen vor allem radikale Produktentwicklungen, dass Innovationen zunehmend aus verschiedenen Sektoren kommen können. Das wohl prominenteste und zugleich nicht umwegfreie Beispiel hierfür ist das iPhone. Als dieses 2007 von Apple eingeführt wurde, handelte es sich dabei zunächst um eine eher inkrementelle Innovation: Es präsentierte eine in vielen Belangen bessere Alternative zu anderen, bereits auf dem Markt befindlichen Smartphones. Der anfängliche Erfolg des iPhones

ist folglich wohl auf die Überlegenheit des iPhones im Hinblick auf »klassische« Smartphone-Funktionen zurückzuführen. Zur disruptiven Innovation wurde das iPhone schließlich nicht durch sein äußerliches Design, eine intuitivere Bedienoberfläche oder eine performantere Hardware. Disruptive Wirkung entfaltete das iPhone erst in Kombination mit Apples App Store. Dieser stellte erstmals eine Plattform zur Verfügung, die Smartphone-nutzende und App-Entwickelnde quasi direkt zusammenbrachte und beiden einen ständig wachsenden Applikationsmarkt bot. Disruptiv wirkte diese Innovation dann auch nicht primär auf den Smartphonemarkt, sondern vor allem auf den Laptop- und Computermarkt. Plötzlich war für Smartphones eine Menge an

Anwendungen verfügbar, die es so bisher nur für Computer gegeben hatte. So aufgestellt hatte mit dem iPhone ein Smartphone erstmals tatsächlich das Potenzial, Laptops und Computer als primäre Internetzugangsmittel für Nutzende abzulösen.<sup>118</sup> Mit der Entwicklung von Tablets wurde diese Marktentwicklung nur konsequent weitergeführt. In der heutigen Datenökonomie wiederholt sich dieses Muster: Relevante Innovationen, ob auf Daten oder auf Technologien zur Datenverwertung basierend, können aus unterschiedlichsten Ecken kommen, die nicht unbedingt im angestammten Industrie- oder Wissenschaftssektor des Ausgangsproduktes oder Herstellers liegen. Organisationen

<sup>117</sup> Juliane Schmeling, Lina Bruns (2020): »Kompetenzen, Perspektiven und Lernmethoden im digitalisierten öffentlichen Sektor«. [https://qualifica-digitalis.de/wp-content/uploads/QD\\_Metastudie\\_20201005\\_barrierefrei\\_v5.pdf](https://qualifica-digitalis.de/wp-content/uploads/QD_Metastudie_20201005_barrierefrei_v5.pdf).

<sup>118</sup> Interessant ist das iPhone als Beispiel, weil es sowohl Elemente einer inkrementellen bzw. erhaltenden als auch einer disruptiven Innovation enthält. Ob eine Innovation disruptiv wirkt, entscheidet sich nicht einfach anhand besonders neuartiger Eigenschaften, sondern durch die wettbewerbliche Wirkung auf Konkurrenzprodukte in relevanten Märkten. In diesem Sinne verdrängte das iPhone die Konkurrenzprodukte von Nokia und Blackberry primär durch seine überlegenen Funktionen und Produkteigenschaften, die aber mit einem deutlich höheren Preis einhergingen. Der oft zitierte Niedergang von Nokia und Blackberry wurde aus diesem Blickwinkel ausgerechnet von einem Produkt eingeleitet, das entlang bekannter Parameter entwickelt worden war. Einfach ausgedrückt: Das iPhone war eine inkrementelle, »absehbare« Innovation, weil es im Vergleich zu anderen Mobiltelefonen und Smartphones schlichtweg vieles besser, aber auch teurer machte. Disruptiv wirkte das iPhone hingegen tatsächlich eher auf den Computermarkt: Hier unterbot es bestehende Software- und Applikationsangebote durch ein einfacheres, günstigeres und für viele Nutzende ausreichendes Angebot. Clayton M. Christensen, Michael E. Raynor, and Rory McDonald (2015): »What Is Disruptive Innovation?«. <https://hbr.org/2015/12/what-is-disruptive-innovation>.

DATENSTRATEGIEN SOLLTEN  
AUCH DIE ENTWICKLUNG VON  
ORGANISATIONEN UND DEREN  
FÄHIGKEIT ZUR DATENNUTZUNG  
BERÜCKSICHTIGEN.

müssen daher konsequent Möglichkeiten verfolgen, die ihnen den Zugang zu neuen Kollaborationspartnern und deren Ideen eröffnen.<sup>119</sup> Insbesondere öffentliche Verwaltungen sollten dazu auf der Ebene der Mitarbeitenden Kompetenzen entwickeln und kombinieren, die nicht unbedingt zum traditionellen Repertoire des Sektors gehören: Kreativität, unternehmerisches Denken und Innovationskompetenz. So benötigt die digitale Umgestaltung von Geschäftsprozessen kreatives Denken ebenso wie eine lösungsorientierte Reflektion über verwaltungs- oder unternehmerische Zielsetzungen, damit neu gefasste Aufgaben organisationsgerecht priorisiert werden.<sup>120</sup> Offen ist dabei, in welchem Umfang auch im öffentlichen Sektor diese Veränderungsprozesse vollends mit internen Ressourcen initiiert und getragen werden müssen: Sollten Verwaltungen beispielsweise im Idealfall eigene Innovationsteams unterhalten? Gerade im Sinne staatlicher Gestaltungsperspektiven ist bei dieser Frage zu bedenken, dass auch unabhängige institutionelle Matchmaker eine Lösung sein können, um für Unternehmen, Verwaltungen, Wissenschaft und die Zivilgesellschaft ein offeneres, besser vernetztes kollaborativeres Ökosystem für Innovationen zu schaffen. Institutionen wie das Open Data Institute und das System der Catapults<sup>121</sup> dienen beispielsweise in Großbritannien als Innovationsstewards, indem sie in Beratungs-, Entwicklungs- und Forschungsprojekten Vernetzungsmöglichkeiten zwischen unterschiedlichen Partnern fördern und damit sogar den sektorübergreifenden Ideenaustausch besonders vorantreiben.

### Handlungsempfehlungen

Neben Maßnahmen zur Steigerung der Zugänglichkeit von Daten sollten Datenstrategien auch die Entwicklung von Organisationen und deren Fähigkeiten zur Datennutzung berücksichtigen. Dies ist nicht zuletzt auch deshalb wichtig, damit die Organisationen von der Verfügbarkeit ihrer eigenen Daten und deren auf die Organisationsziele hin ausgerichteten Auswertung profitieren können. Relevant sind dazu vor allem die Schaffung von Kollaborations- und Förderprogrammen, die den, auch sektorübergreifenden, Wissens- und Ideentransfer für neue Geschäftsmodelle beschleunigen und neue Partnerschaften zwischen Unternehmen, Verwaltungen und Zivilgesellschaft schaffen. Gezielte Programme wie das britische Catapult-Netzwerk, das Kollaborationen zwischen Industrieunternehmen, Start-ups und der Wissenschaft gezielt unterstützt, können diese Aufgabe realisieren. Ergänzend sollten zudem laufend neue Erkenntnisse zur datengetriebenen Transformation von Unternehmen, zivilgesellschaftlichen Organisationen, politischen Organisationen und Verwaltungen erhoben werden. Denn den Weg der Transformation zu verfolgen und zu erkennen, fällt ohne strukturiertes Wissen schwer. Mit dem Open Data Institute<sup>122</sup> wurde in Großbritannien eine Einrichtung geschaffen, die Organisationen aus verschiedensten Sektoren gezielt dabei unterstützt, diesen Weg zu beschreiten. Das Institut hilft, den allgemeinen Wissensstand zur datengetriebenen Transformation von Unternehmen und des öffentlichen Sektors weiterzuentwickeln. Mit seinen sektorübergreifenden Unterstützungs- und Beratungsleistungen wirkt es der Fragmentierung des britischen Datenökosystems entgegen und hilft diesem stattdessen beim Zusammenwachsen. Aus strategisch-politischer Sicht kann diese Einrichtung auch ein Vorbild für die Weiterentwicklung des deutschen Datenökosystems und seiner Akteurslandschaft sein.

<sup>119</sup> Nathan Furr, Kate O’Keeffe, Jeffrey H. Dyer (2016): »Managing Multiparty Innovation«. <https://hbr.org/2016/11/managing-multiparty-innovation>.

<sup>120</sup> Juliane Schmeling, Lina Bruns (2020) : »Kompetenzen, Perspektiven und Lernmethoden im digitalisierten öffentlichen Sektor«. [https://qualifica-digitalis.de/wp-content/uploads/QD\\_Metastudie\\_20201005\\_barrierefrei\\_v5.pdf](https://qualifica-digitalis.de/wp-content/uploads/QD_Metastudie_20201005_barrierefrei_v5.pdf).

<sup>121</sup> »Welcome to the CATAPULT Network«. <https://catapult.org.uk/>.

<sup>122</sup> »ODI – Open Data Institute«. <https://theodi.org/>.



## 5. AUSBLICK

Wie zuvor argumentiert folgen fortschrittliche Datenstrategien einer vergleichsweise einfachen Formel: Sie erkennen Daten als gesellschaftlich tief verankerte Hebel zum Erreichen von mehr Innovationen und Produktivität. Damit Daten die in sie gesetzten Hoffnungen einlösen können, streben fortschrittliche Datenstrategien zudem eine möglichst intensive und vielfältige Datennutzung an. Als Schlüssel dazu dienen ihnen ein hürdenfreier Datenzugang, Wettbewerb und befähigte Organisationen.

Aus politischer Sicht markieren die aktuellen Diskussionen und Schritte zur Entwicklung von Datenstrategien einen Wendepunkt in der Datenpolitik. Nach einer Dekade primär angebotsgetriebener Daten(-wirtschafts-)politik rückt die Nachfrageseite der Datennutzung in den Vordergrund. Die im Sommer 2021 erfolgte Einführung des Datennutzungsgesetzes<sup>123</sup> als Teil des zweiten Open-Data-Gesetzes verdeutlicht diesen Schwerpunktwechsel in Deutschland. Die strategisch relevanten Arbeiten sind damit aber noch längst nicht abgeschlossen: Je mehr Daten ins Zentrum politischer, gesellschaftlicher und wirtschaftlicher Entwicklungen vorrücken, desto universeller werden Inhalt und Anspruch strategischer Maßnahmen zur Förderung der Nutzung von Daten. Die breite Datendefinition des von der EU-Kommission vorgelegten Data Governance Act unterstreicht etwa, dass Daten nicht mehr nur Tabellen oder Dateien sind, sondern »jede digitale Darstellung von Handlungen, Tatsachen oder Informationen«<sup>124</sup> sein können. Im Angesicht eines solch umfänglichen Datenverständnisses werden vor allem politische Entscheider definieren müssen, wie für den gesamten Datenlebenszyklus und über unterschiedliche Sektoren hinweg einheitliche Regeln für Datennutzung und -zugang geschaffen werden können. Die Herausforderungen liegen vor diesem Hintergrund in der Reduktion von Komplexität und in der Zusammenführung eines fragmentierten Datenökosystems. Das Erbe

sektorspezifischer Infrastrukturen und Regelungen zu vereinfachen und kompatibel mit einer universellen sektorübergreifenden und domänenagnostischen Datennutzung zu machen, dürfte dabei ein wichtiger Schlüssel zum Erfolg sein.

---

<sup>123</sup> »Gesetz für die Nutzung von Daten des öffentlichen Sektors (Datennutzungsgesetz – DNG) vom 16. Juli 2021«. BGBl. I S. 2941, 2942, 4114.

<sup>124</sup> Artikel 2 der »Verordnung (EU) 2022/868 des Europäischen Parlaments und des Rates vom 30. Mai 2022 über europäische Daten-Governance und zur Änderung der Verordnung (EU) 2018/1724 (Daten-Governance-Rechtsakt)«. <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A32022R0868>.

## KONTAKT

Resa Mohabbat Kar  
Kompetenzzentrum Öffentliche IT (ÖFIT)  
Tel.: +49 30 3463-7173  
Fax: +49 30 3463-99-7173  
info@oeffentliche-it.de

Fraunhofer-Institut für  
Offene Kommunikationssysteme FOKUS  
Kaiserin-Augusta-Allee 31  
10589 Berlin

[www.fokus.fraunhofer.de](http://www.fokus.fraunhofer.de)  
[www.oeffentliche-it.de](http://www.oeffentliche-it.de)  
Twitter: @OeffentlicheIT

ISBN: 978-3-948582-09-8

