

Die „Data Intelligence Offensive“ (DIO) zielt auf die Forcierung und Förderung der Datenwirtschaft und des optimierten Einsatzes von Technologien ab, vorrangig der Künstlichen Intelligenz (KI). DIO fördert Perspektivenwechsel von einem Austauschkonzept zu einem Handelskonzept, von Data Exchange zu Data Sharing und Trading. DIO bezweckt die Implementierung von europäischen Werten wie Datenschutz und Datensouveränität im technischen Grunddesign und in den regulatorischen Vorschriften.

BMK EU-Datenstrategie Dialoge

Zusammenfassung und Positionen der österreichischen Data Community

Data Intelligence Offensive

25. Februar 2021

Bericht in inhaltlicher Erfüllung des Beauftragungsvertrags VERTRAG: BW000021315
Gegenstand des Vertrags: „Wie werden datengetriebene Organisationen die europäische
Datenstrategie umsetzen und was wird dafür benötigt?“
(Stellungnahme EU-Datenstrategie).

Inhalte

| | |
|--|----|
| 0. Präambel – Datenstrategie und Regulierungsbedarf..... | 1 |
| 1. Methode | 4 |
| TEIL 1: Erhebung des Status Quo – die EU-Datenstrategie in Österreich | 5 |
| 1.1. Förderung des Ausbaus eines Daten-Service-Ökosystems..... | 6 |
| 1.2. Weiterentwicklung und Betreuung des Daten-Service-Ökosystems | 7 |
| 1.3. Themen (sektorspezifisch)..... | 7 |
| 1.3.1. Datenräume..... | 7 |
| 1.3.2. Datenkreise..... | 7 |
| 1.4. Mehrwert durch (Wieder-)Verwendung und Verknüpfung von Daten..... | 8 |
| 1.5. Beschreibung der Schnittstellen- Wichtigkeit der High-Value-Datasets | 9 |
| 1.6. Wertschöpfung mit Innovation – durch Verschränkung von Schnittstellen verschiedener Datenräume | 11 |
| 1.7. Use Cases | 11 |
| 1.8. Awareness – Bewusstsein und Vertrauen schaffen..... | 15 |
| 1.9. Forschung. Technologie. Innovation. | 16 |
| TEIL 2: Themen aus Live-Diskussionen der Data Community..... | 19 |
| 2.1. Datenmanagement..... | 19 |
| 2.2. Datenanalyse und -visualisierung | 19 |
| 2.3. Verarbeitung natürlicher Sprachen..... | 19 |
| 2.4. Gesundheitsbereich | 20 |
| 2.5. Technologiegestütztes Lernen..... | 20 |
| 2.6. Gemeinschaftliche Nutzung von Daten über Unternehmensgrenzen hinweg | 20 |
| 2.7. Effiziente Handhabung produktionsrelevanter Daten | 20 |
| 2.8. Anwendung moderner KI-Verfahren auf Produktionsdaten | 21 |
| 2.9. Teilnahme an GAIA-X..... | 21 |

0. Präambel – Datenstrategie und Regulierungsbedarf

Informations- und Kommunikationstechnologien, sowie die damit verbundene Datenverarbeitung sind zum digitalen Nervensystem unserer Gesellschaft geworden. Es gibt keinen Bereich unserer modernen Lebenswelt mehr, der nicht direkt oder indirekt von IKT abhängig ist. Gleichzeitig ereignen sich regelmäßig schwere Security-Vorfälle, Datenverluste, Missbräuche personenbezogener Daten und immer wieder auch massive Hacking-Angriffe, die weite Teile von IT-Infrastrukturen kompromittieren. Diese Vorfälle sind Indikatoren dafür, dass das Qualitätsniveau der heutigen Infrastruktur weit entfernt davon ist, was die Gesellschaft von einer für ihr Funktionieren und Überleben essentiellen Infrastruktur verlangen kann und muss.

Die Wirkmächtigkeit von Internet und digitaler Telekommunikation baut auf den Grundprinzipien einer hochgradig verteilten, redundanten und resilienten Infrastruktur auf. Systemisch ist es jedoch verstärkt zu einer Zentralisierung durch die Marktmechanismen der globalisierten Plattformökonomie gekommen.

Dieser Trend hat auf der Basis einer anfänglich noch resilienten Infrastruktur einerseits zu einem zunehmend hohen Maß an Fragilität und Single-Points-of-Failure eingeführt, andererseits zu einer oligopolistischen Marktmachtkonzentration von Akteur*innen außerhalb der Europäischen Union. Daher finden sich die Anwender*innen in der EU heute in der problematischen Situation, in Sachen Hardware von asiatischen (chinesischen) Anbieter*innen und in Sachen Software und Plattformen von US-amerikanischen Anbieter*innen enorm abhängig zu sein.

Diese Abhängigkeit besteht aus wenigstens drei Dimensionen: einer politischen, einer ökonomischen und nicht zuletzt der genannten technischen. Die politischen und ökonomischen Dimensionen werden zunehmend erdrückend, da die fortschreitende Zentralisierung monopolisierende Machtzentren außerhalb der Europäischen Union schafft und damit lokale Risiken dramatisch erhöht.

Moderne IKT und Datenverarbeitung beruhen zunehmend auf diesen Plattformen und Cloud-Lösungen. Damit rücken diese ins Zentrum aller strategischen Betrachtungen für Nationen und die europäische Union als Gemeinschaft. Kurzfristige »Effizienzgewinne« und Einfachheit in einzelnen Projektumsetzungen und partikulären Geschäftstätigkeiten erzeugen im derzeitigen Modell langfristigen Schaden für alle. Diese Situation lässt sich nicht mehr alleine auf technischer Ebene

oder aus Sicht einzelner Akteur*innen verbessern. Politik und Regulation in Form einer Datenstrategie werden notwendig, um langfristige und existentielle Aspekte zu berücksichtigen und steuernd einzugreifen.

Ethische Rahmenbedingungen und gesellschaftliche Wertvorstellungen bestimmen alle drei Dimensionen einer komplexen IT-Infrastruktur und lassen sich nach einer technischen Implementierung auch nicht schnell und einfach verändern. Die Unternehmen in den USA agieren im Wesentlichen nach dem Prinzip: Machen, was (technisch) machbar ist und sich global skalieren lässt. Respekt der Privatsphäre und Datenschutz haben keine nennenswerte Relevanz. Nicht-US-Interessen und Konzepte sind in den Strategien der Anbieter*innen von sehr untergeordneter Bedeutung. Dies betrifft unter anderem auch die bekannten Zugriffe von geheimdienstlicher Seite auf Unternehmen, Personen und Daten aller Art. Dazu kommt die Abhängigkeit von Hardware (im Besonderen von asiatischen Anbieter*innen), beziehungsweise von ganzen Systemlösungen. Sie ist aus Sicherheitsgründen problematisch, da bei der heutigen Komplexität der Systeme nicht garantiert werden kann, dass diese nicht Funktionen implementieren, die nicht im Interesse der Kund*innen liegen. Außerdem kommen die genannten schwerwiegenden Qualitätsprobleme bei nahezu allen Anbieter*innen dazu: Datenverlust, Security-Vorfälle und technische Ausfälle, sowie schwere Fehler in der Software, haben aktuell keine nennenswerten Folgen für Anbieter*innen, sehr wohl aber für Kund*innen und Gesellschaft. Mangelnde Offenheit von Protokollen, Diensten und Nutzungsbedingungen führt zu einer Verschärfung dieser Lock-In Situation.

Die Europäische Union folgt aus guten Gründen in vielen Bereichen anderen Prinzipien, wie z.B. in der Umsetzung der DSGVO gezeigt wurde. Datenschutz und Schutz der Privatsphäre sowie langfristiges Denken im Sinne des Vorsorgeprinzips sind in der EU verankerte Prinzipien. Datensparsamkeit und Resilienz sollten daher als zentrale Prinzipien jeder modernen Cloud-Plattform und IT-Infrastruktur gelten. Derzeit wird aber, beispielsweise, trotz DSGVO kaum Rücksicht genommen, wo Daten abgespeichert werden und wer de facto (und auf die USA bezogen, im Grunde auch de jure) Zugriff auf diese hat. Auch die Regulierung von Inhalten (und Diensten, die auf Plattformen betrieben werden, siehe etwa die aktuelle Diskussion um soziale Netzwerke oder App-Stores) sollte nach rechtsstaatlichen und demokratischen Prinzipien erfolgen und nicht alleine privaten Akteur*innen und deren singulären Firmeninteressen überlassen werden – jedenfalls nicht unter der aktuellen monopolistischen Situation, in der einzelne Unternehmen als weltweite Kommunikations-Infrastruktur verstanden werden müssen.

Die DSGVO erscheint daher als ein erster, sehr wichtiger Schritt, auf den weiter aufgebaut werden sollte. Sie war auch eine außerhalb der EU – inklusive den USA – deutlich wahrgenommene Initiative, hat vielfach auch bemerkenswert positive Resonanz ausgelöst und gezeigt, dass die EU in entscheidenden Bereichen der emergenten Datenwirtschaft Vorreiterin sein kann.

Der Regulierung von Plattformen, Services und Datenverarbeitungsprinzipien muss eine Definition gemeinsam vertretener ethischer und gesellschaftlicher Werte vorangehen, welche die oben genannten Prinzipien berücksichtigt. Datensparsamkeit, Resilienz, demokratische Werte, sowie offene Standards (Protokolle) und Software, sollten dabei Leitlinien einer Datenstrategie bilden. Resilienz lässt sich nur erzielen, wenn übermäßige Zentralisierung und Monopolisierung verhindert bzw. regulatorisch rückabgewickelt werden. Regulierung sollte daher auf systemischen aber wenigen und einfachen Prinzipien beruhen, die es einerseits Klein- und Mittelbetrieben ohne größere administrative Schwierigkeiten erlaubt, Services auf Basis dieser Regularien anzubieten und andererseits zu große Konglomerate und Monopole strategisch benachteiligt.

Von großer Bedeutung ist in diesem Kontext auch die Definition von Qualitätsstandards, denen IKT-Services zu folgen haben. Zu lange wird es schon gesellschaftlich akzeptiert, dass die Folgen mangelhafter Softwarequalität vertraglich auf die Nutzer*innen und Gesellschaft abgewälzt wurden. Die EU kann auf Basis resilienter und ethischer IKT-Richtlinien und -Services auch wieder eine Vorreiterrolle auf internationaler Ebene bilden. Teil einer Regulierungsbestrebung kann die Definition eines Gütesiegels und damit verbundene Audits sein. Staatliche Services, EU-Projekte, wissenschaftliche Projekte und sensitive Dienste (»kritische Infrastruktur«) sollten mittelfristig ausschließlich auf derartig zertifizierten Plattformen betrieben werden dürfen. Aber auch für europäische Unternehmen und Organisationen darf es keine Option mehr sein, kritische Daten und Verarbeitungslogik bei nicht zertifizierten Betreibern auszulagern. Eine derartige Standardisierung und Regulierung ermöglicht den Aufbau europäischer Cloud-Lösungen und IT-Plattformen, von denen eine Digitalisierung zu erwarten ist, die tatsächlich als Fortschritt für die Bürger*innen Europas gelten kann. Wenn diese europäischen Standards und Wertvorstellungen im Kern berücksichtigen, öffnen sie neue ökonomische Perspektiven für lokale Anbieter*innen, die damit eine tatsächliche Chance bekommen würden, aktuell dominierenden Anbieter*innen Konkurrenz zu machen. Folgt man der ausstrahlenden Wirkung der DSGVO werden sich diese Perspektiven auch nicht auf die EU alleine beschränken.

1. Methode

Die Erläuterungen im folgenden Dokument wurden im Auftrag des Bundesministeriums für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (BMK) von der Data Intelligence Offensive (DIO) in drei **Workshops** mit Vertreter*innen aus Forschung, Industrie, bzw. der österreichischen Data Community erhoben.

In einem weiteren Schritt wurden diese Inhalte zusammengeführt und eine konkrete Expert*inneneinschätzung aus dem DIO-Netzwerk formuliert, um detailliertere Empfehlungen aufzuschlüsseln.

Von den drei Workshop Dialoge wurden zwei Online und einer in Präsenz abgehalten, wobei in beiden Formaten großer Wert auf die Interaktivität zwischen den Teilnehmer*innen, sowie auf deren proaktive Einbindung in die Erarbeitung der Inhalte gelegt wurde.

Diese Prozessstruktur wurde gewählt, da die DIO als Vermittler der Ansichten der Data Community fungiert und den Austausch mit staatlichen Behörden ermöglicht. Trotz bzw. gerade aufgrund der Diversität der Teilnehmenden, konnte eine diverse und fundierte Expert*innenmeinung erarbeitet werden, die auf vielfältigen Erfahrungen und Knowhow basiert.

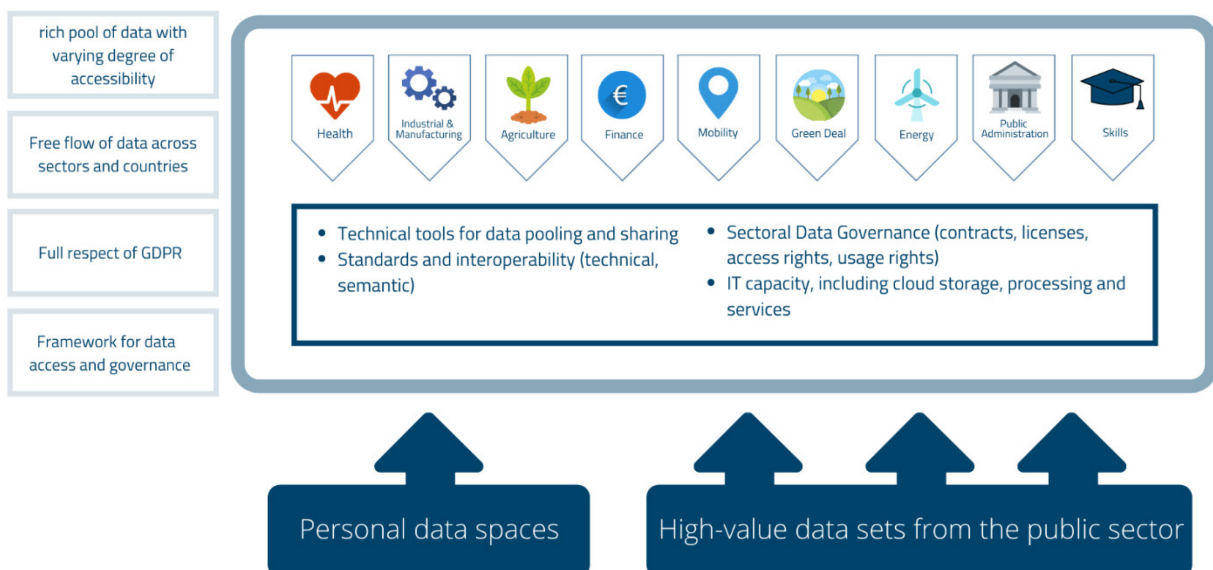
TEIL 1: Erhebung des Status Quo – die EU-Datenstrategie in Österreich

Im Kontext der laufenden Diskussionen und Konsultationen wurde in den Workshops von den Teilnehmer*innen deren Ansichten und Beurteilungen zur EU-Datenstrategie abgefragt und die unterschiedlichen Sichtweisen zur Diskussion gebracht.

Die europäische Datenstrategie zielt auf die Schaffung eines Binnenmarkts für Daten ab, damit diese innerhalb der EU branchenübergreifend zum Nutzen von Unternehmen, Forschern und öffentlichen Verwaltungen verwendet werden können. Der Zugang zu ständig wachsenden Datenmengen und die Fähigkeit, diese zu nutzen, werden als Schlüssel zu Innovation und Wachstum gesehen.

Die europäische Datenstrategie dient dazu, das Ziel eines echten Binnenmarkts für Daten zu verwirklichen und die festgestellten Probleme mit politischen Maßnahmen und finanzieller Förderung anzugehen. Dabei sind Anschubinitiativen in einer Reihe von Datenräumen geplant.

Common European data spaces



© Originaldarstellung: Yvo Volman, Europäische Kommission / adaptierte Darstellung durch DIO

Die EU-Datenstrategie ist auf vier Säulen aufgebaut:

- Säule A: Ein sektorübergreifender Governance-Rahmen für Datenzugang und Datennutzung.

- Säule B: Voraussetzungen – Investitionen in Daten und in die Stärkung der europäischen Kapazitäten und Infrastrukturen für das Hosting, die Verarbeitung und die Nutzung von Daten sowie der Interoperabilität.
- Säule C: Kompetenzen: Stärkung der Handlungskompetenz von Einzelnen, Investitionen in Kompetenzen und in KMU.
- Säule D: Gemeinsame europäische Datenräume in strategischen Sektoren und Bereichen von öffentlichem Interesse.

Zur strategischen Einordnung sind die folgenden Ausführungen der Ansichten und Einschätzungen der Experten-Teilnehmer*innen den jeweiligen Säulen mit einer Kennzeichnung (A, B, C, D) in ihrer Relevanz zugeteilt.

1.1. Förderung des Ausbaus eines Daten-Service-Ökosystems

Relevante Säule(n) der EU-Datenstrategie: C,D

Aktuelle Probleme österreichischer Unternehmen:

- Das Sammeln und das nachhaltige Anbieten von Daten ist kostenintensiv. Ein Lösungsvorschlag dazu ist, dass diese Infrastrukturinvestitionen vom Staat getragen werden – sinnvolle Alternativen dazu sind noch unklar.
- Insgesamt arbeiten privatwirtschaftliche Firmen noch mit Mikrosystemen, da der Wille, Daten auszulagern, nach wie vor gering ist.
- Es ist weitgehend unklar, welche Daten der Öffentlichkeit gehören. Dies bezieht sich insbesondere auf Daten, die für die Öffentlichkeit – z.B. für wertvolle Statistiken – relevant sind.
- Standardisierung hat Priorität: Ein Datenmarktplatz kann nur bestehen, wenn hohe Qualität gewährleistet wird und Mängel rasch behoben werden können. Wichtig ist dabei, dass nicht alle personenbezogenen Daten auch schützenswert sind. Die Einwilligung zum Teilen solcher Daten sollte unter einer klaren gesetzlichen Regelung erfolgen.

Zusammengefasst sind für den Aufbau eines Daten-Service-Ökosystems folgende Faktoren besonders relevant:

- Anonymisierung
- Federated Machine Learning auf Grund von dezentralen Datensilos

- Datensynthetisierung

1.2. Weiterentwicklung und Betreuung des Daten-Service-Ökosystems

Relevante Säule(n) der EU-Datenstrategie: D

Die österreichische Data Community ist sich einig: Technisch ist sehr viel mehr möglich als organisatorisch. Ein gemeinschaftlicher Public-Private-Ansatz könnte Abhilfe schaffen – erste Schritte wären, sich einer Vision von gemeinsamer Struktur und breit akzeptierten Taxonomien bzw. Katalogen dadurch anzunähern, indem die Interoperabilität und Portabilität von Infrastrukturen, Daten und Diensten auf der Basis von europäischen Standards, smarten Schnittstellen, umfassenden Regeln zum Identitätsmanagement und integrierten Abrechnungssystemen gesichert werden.

Bedeutsam wären auch verstärkte Bemühungen, um zu klären, welche öffentlichen Daten unter welchen Konditionen verfügbar sind.

1.3. Themen (sektorspezifisch)

1.3.1. Datenräume

Relevante Säule(n) der EU-Datenstrategie: D

Daten sind in den letzten Jahren zu einem Schlüsselfaktor für die Entwicklung von Innovationen und neuen Geschäftsmodellen geworden. Daher ist das klare Ziel der europäischen Datenstrategie, einen einheitlichen europäischen Datenraum zu schaffen, um auf Basis von qualitativen und sicheren Daten bessere Entscheidungen treffen zu können.

Dabei ist die Vernetzung zahlreicher Akteure und die Schaffung von Schnittstellen ebenso entscheidend, wie die Performanz und Ausfallsicherheit der Infrastruktur (siehe GAIA-X als European Cloud Computing Projekt). Ziel ist es, einen sicheren, domänenübergreifenden Datenraum zu schaffen, der Unternehmen verschiedener Branchen und aller Größen die souveräne Bewirtschaftung ihrer Datengüter ermöglicht (siehe IDSA als Initiative).

1.3.2. Datenkreise

Relevante Säule(n) der EU-Datenstrategie: D

Die Umsetzung des österreichischen Konzepts der Datenkreise identifiziert das Potential, den notwendigen Daten- und damit Wissensaustausch souverän, transparent und sicher zu

ermöglichen. Der souveräne und sichere Datenaustausch soll in Datenkreisen zu Wertschöpfungsnetzwerken führen, die smarte Services, innovative Leistungsangebote und automatisierte Geschäftsprozesse ermöglichen.

Der damit entstehende enorme Mehrwert für alle Stakeholder bedarf aber einer gezielten Unterstützung. Da Daten oft sehr unterschiedlicher Natur sind (z.B. Verkehrsdaten vs. Firmendaten), ist es noch unklar, wie die technologische und organisatorische Umsetzung der Datenkreise aussehen wird.

Das Konzept der Datenkreise findet große Resonanz und Akzeptanz in der Data Community. Zur tatsächlichen Umsetzung der Datenkreise ist es erforderlich, dass vorab Wertgemeinschaften geschaffen werden (Definition gemeinsamer Ethikrichtlinien, Standards, Ziele, Mehrwert für alle Beteiligten, etc.).

1.4. Mehrwert durch (Wieder-)Verwendung und Verknüpfung von Daten

Relevante Säule(n) der EU-Datenstrategie: B, D

Bevor Daten wiederverwendet und verknüpft werden können, sollte Klarheit geschaffen werden, woher die Daten kommen, welche Qualität sie haben, wer die Rechte darauf hat, etc. Datenhandhabung, Datenverwendungsstrategie und Ähnliches, müssen von der Leitungsebene innerhalb von Organisationen geregelt werden.

Eine bedeutende Hürde stellen sachliche, technische und rechtliche Unklarheiten in Bezug auf Daten-Eigentümer*innenschaft dar. Daran hängt jedoch auch die Umsetzung von Datensouveränität.

Wenn die Verfügbarkeit bei personenbezogenen Daten auf expliziter, individueller Einwilligung basiert, leiden sowohl die Datenqualität eines ganzen Datensets wie auch seine (wirtschaftlich, technisch notwendige) Größe.

Auch der Rückzug aus der Datenwirtschaft erzeugt Kosten und Österreich wäre in einem solchen Szenario im Nachteil gegenüber jenen, die gesamthafte Datensätze haben (also den großen Plattformbetreiber*innen). Deren Nutzungsbedingungen sind jedoch für die End-User*innen weitgehend nicht plausibel, wodurch die europäische Prämisse des Datenschutzes und der souveränen Datennutzung oft (unbewusst) freiwillig außer Acht gelassen wird.

1.5. Beschreibung der Schnittstellen- Wichtigkeit der High-Value-Datasets

Relevante Säule(n) der EU-Datenstrategie: D

Der Austausch von Daten muss breit ausgerichtet sein. Anhaltspunkte für einen solchen Austausch sind z.B. Data Market Austria (DMA), International Data Spaces (IDS), GAIA-X, Cooperation OGD und Ähnliches.

Die relevanten Institutionen und Organisationen sollten intensiver zusammenarbeiten und sich austauschen, damit sogenannte „Silos“ aufgebrochen werden – dabei ist die Priorisierung von High-Value-Datasets (HVD) essentiell. Deren Bereitstellung ist von entscheidender Bedeutung, um ihr Potenzial für die europäische Gesellschaft und Wirtschaft zu nutzen, beispielsweise, um die Forschung zu bereichern, Entscheidungen zu treffen oder neue Produkte und Dienstleistungen zu entwickeln.

Die Bewertung von Datensätzen hängt von der Marktposition, der Spezifikation oder auch dem (geografischen, sektoralen) Wirkungsbereich ab. Datensatz-Download-Statistiken werden häufig verwendet, reichen jedoch nicht aus, um deren Wert und ihre möglichen Auswirkungen zu bewerten. Klarheit über standardisierte Grundlage für die Wertbewertung sollte geschaffen werden.

Rollen und Verantwortlichkeiten bei der Festlegung, Implementierung und Wartung von HVDs sind häufig weder klar noch werden sie durch ein Mandat oder bestimmte Ressourcen unterstützt.

Um die Auswirkungen von HDV weiter zu erhöhen und Markteintrittsbarrieren für Start-ups und KMU abzubauen, müssten Bemühungen auf diejenigen Datensätze abzielen, die das größte Potenzial für Gesellschaft und Wirtschaft haben, beispielsweise Wetterdaten, Umweltdaten, Geodaten, Finanzdaten, anonymisierte Gesundheitsdaten für die Forschung oder Registerforschungsdaten.

Die Beschreibung und Implementierung von Schnittstellen sollte von Akteur*innen, die als *Data Intermediaries* fungieren, geleistet werden.

Das EU Projekt TRUSTS (www.trusts-data.eu) bietet hier einen sinnvollen Ansatz, um adäquate Rahmenbedingungen herzustellen. In Österreich werden mehr sinnvolle Initiativen benötigt, die

datengetriebene Innovationen vorantreiben können (z.B. auch für ein Zusammenspiel von Standardisierung und Regulierung).

Die Bereitschaft zum Data Sharing hängt wesentlich davon ab, ob für die Akteur*innen absehbar ist, welche Art von Daten sie erhalten und wie deren Qualität und Wert ist.

Öffentliche Stellen sollten Unternehmen außenwirksam anerkennen, wenn diese zum Datenaustausch bereit sind. Ihre Bereitschaft sollte für andere sichtbar sein.

Auch das Thema Anonymisierung braucht klare Regeln. Ein öffentlicher Datentresor mit Verwaltungsdaten, klaren Schnittstellen und klaren Nutzungsregeln erscheint für die Data Community wünschenswert. Der Zugang zur Offenlegung in Österreich muss klar verbessert werden, denn Datenschutz geht vor Veröffentlichung in PSI. Wichtig ist hierbei:

- PSI Benutzerfreundlichkeit – differenzierte Informationen
- Bewerbung des Datenangebots
- Vertrauensbasis / TRUST in der Nutzung

Die Data Community spricht sich dafür aus, dass die öffentliche Hand eigene Datenmärkte aufbaut und diese mit Daten aus der öffentlichen Hand befüllt, z.B. Mobilitätsdaten, Verkehrsdaten (VAO > Leuchtturm). Die Ängste und Bedenken können durch Best-Practice-Beispiele abgebaut werden, um dadurch die Kollaboration der Player zu fördern. Diese erfolgreichen Use-Cases (Vorschläge siehe Kapitel 2.7.) müssen sichtbar gemacht und beworben werden, Metadaten müssen publiziert und personenbezogene Daten durch Datensynthetisierung nutzbar gemacht werden.

In allen Stakeholder-Communities sollte mehr Awareness zu datengetriebener Innovation kreiert werden (siehe Kapitel 2.8.). Außerdem bedarf es einer Lösung einiger technologischer Fragen zu geschützten Datenbereichen.

Österreich hat z.B. keine Cloudstrategie, welche aus Sicht der Data Community absolut notwendig wäre. Ein Lösungsansatz könnte die Zusammenarbeit von Universitäten, dem Bundesrechenzentrum (BRZ) und privatwirtschaftliche Firmen sein.

Cloudprodukte wie dieses sollten automatische DSGVO-Konformität aufweisen, aber auch so einfach zu verwenden sein wie z.B. Google, Microsoft Azure, etc. Eine vielversprechende, relevante Überlegung ist, diese Idee in das Konzept und die Association von GAIA-X einzubringen.

Solche Initiativen sollten unter eine gemeinsame Schirmherrschaft bzw. auf eine Kooperationsplattform gestellt werden.

Die Stakeholderdefinition würden dabei umfassen:

- Fokus öffentliche Hand + Unternehmen,
- Unternehmen + Unternehmen,
- Unternehmen + Privatpersonen,
- Öffentliche Hand + Privatpersonen,
- Privatpersonen + Privatpersonen.

1.6. Wertschöpfung mit Innovation – durch Verschränkung von Schnittstellen verschiedener Datenräume

Relevante Säule(n) der EU-Datenstrategie: D

Ein Vorschlag aus der Data Community ist, dass öffentliche Daten europaweit in Datenpools inkl. den damit einhergehenden Regeln verfügbar gemacht werden. Auf dieser Basis würde mehr Domänenwissen auf Grund von gemeinsamen Datenräumen geschaffen, wie beispielsweise bei folgenden Anwendungsfällen: Verkehrsplanung, balanciertes Stromnetz, Mobilität, Migrationsbewegungen oder Klimaschutz (Bepflanzungsdaten, z.B. der unterschiedliche Regionen typischen Baumwuchs). Der Frage, ob öffentliche Daten überhaupt dem Urheberrecht unterliegen, muss erst auf den Grund gegangen werden.

Derzeit gibt es zu wenig Wissen über die Einsparungspotentiale in der Produktion durch smarte Datennutzung. Eine finale Aussage, ab welchem Stadium Daten rechtlich gesehen als anonymisiert gelten, bleibt weiterhin aus. Es herrscht also ein Zwiespalt zwischen innovativen Data-Science-Gedanken und dem Datenschutzthema.

1.7. Use Cases

Relevante Säule(n) der EU-Datenstrategie: D

Use Cases sollten vollständig durchgespielt werden, damit eine Erkenntnis darüber entsteht, wie Daten verwendet werden dürfen und welche rechtlichen Probleme dabei auftauchen könnten. Daraus können sich Best-Practices ableiten.

Wünschenswert und aktuell sinnvoll sind Use Cases zu den Themen Landwirtschaft, Logistik und Mobilität.

Folgende Fragen sind hierbei zentral:

- Wie kann der Wissensaustausch stattfinden?
- Welche Plattformen existieren bereits?
- Wie können meine Daten angereichert werden?

Beispiele für mögliche Use Cases im Bereich Mobilität:

Verkehr & Logistik - Öffentliche Verkehrsmittel

Use Case „Öffentliche Verkehrsnetzoptimierung in Wiener Industriezonen“

- Ziele: Erhebung des Bedarfs von zusätzlichen Buslinien in den Wiener Industriezonen (21., 22. und 23. Bezirk)
- Schlagworte: Nachhaltigkeit, Klimafreundlichkeit, öffentlicher Verkehr
- Beschreibung: Durch Quelle-Ziel-Matrizen zwischen Wohngebiet der Arbeitnehmer*innen und deren Arbeitsplätzen wird festgestellt, ob neue Buslinien gebraucht werden, welche Route sie nehmen und zu welchen Zeiten sie vermehrt fahren
- Potentielle Partner*innen: Mobilfunker, WKO, ZAMG, SpotOn Statistics, Wiener Linien, Upstream Mobility, ISTmobil

Use Case „Dynamisches Ticket Pricing öffentlicher Verkehr“

- Ziele: Erhöhung der Auslastung innerhalb der Personenzüge zu relativ schwach gebuchten Zeiten durch dynamisches Ticket Pricing, das sowohl auf den Ticketverkäufen der Verkehrsbetriebe als auch auf einer vorhergesagten Reisemobilität und der derzeitigen Luftverschmutzung basiert. Preislich flexibler als derzeitige Sparschiene
- Schlagworte: Nachhaltigkeit, Klimafreundlichkeit, öffentlicher Verkehr, Entlastung Straßenverkehr
- Beschreibung: Zu den Stoßzeiten an den Werktagen (6-8 und 16-19 Uhr) sind die Züge des öffentlichen Verkehrs stark belegt. In der restlichen Zeit gibt es eine niedrigere Belegung. Damit vermehrt Personen außerhalb der Stoßzeiten auf den Zug umsteigen, können günstigere Tickets in dieser Zeit angeboten werden. Dies kann auch österreichische

Urlauber*innen, je nach Destination, dazu motivieren, bei der Anreise auf den Zug umzusteigen, wenn Auslastung und Preis niedriger sind. Dadurch wird verhindert, dass Züge "halb leer" fahren und es wird dem erhöhten Straßenverkehrsaufkommen entgegengesteuert.

- Potentielle Partner*innen: ÖBB, WESTbahn, Verkehrsbetriebe, ZAMG, Reiseveranstalter*innen, Reisebüros

Straßenbau und -verwaltung - Navigation

Use Case „Stärkung der Position einer Echtzeit-Navigations-App“

- Ziele: Etablierung der ÖAMTC-App als österreichische Lösung einer Echtzeit-Navigations-App durch relevante Zusatzfeatures
- Schlagworte: österreichische Lösung, Verkehrsoptimierung
- Beschreibung: Die aktuell marktführenden Navigations-Apps ermöglichen eine gute Routenplanung, um von Standort A nach Standort B zu gelangen. Dabei sind im Regelfall mehrere Parameter relevant:
 - Art des Transportmittels (Fußweg, Fahrrad, öffentlich, KFZ)
 - Art der Route, z.B. Meidung von Mautstraßen
 - Schnellste (Zeit) vs. kürzeste (Distanz) Routenführung
 - Realistische Fahrt/Ankunftszeitabschätzung basierend auf der aktuellen und prognostizierten Verkehrslage
- Bei den Anbieter*innen handelt es sich meist um große, international agierende Navigationsdienstleister*innen (Google Maps, Waze, TomTom, etc.), die auch eine hohe Reichweite bei den Smartphone-Nutzer*innen haben. Damit sich eine österreichische Lösung dagegen durchsetzen kann, muss sie Features anbieten, die von den bisherigen Anbieter*innen nicht berücksichtigt wurden.
- Folgende USP wären hierbei anhand eines ersten Brainstormings denkbar:
 - Alternativer Parkplatz: Vor allem Wochenendtourist*innen, aber auch Pendler*innen, benötigen bei der Ankunft im Stadtgebiet einen Parkplatz, der sich nahe des Ziels befindet. (Arbeitsplatz, U-Bahn-Station, Veranstaltungszentrum, Stadion, etc.). Bei der Abfahrt kann es sein, dass am ausgewählten Standort noch freie Parkplätze zu finden sind, sich die Situation aber während der Fahrt ändert und alle Parkplätze bei Ankunft belegt sind. In einem solchen Fall können mehrere Alternativen angeboten werden, wobei dies immer vom Ziel der Reise abhängt. Einer*m Tagestourist*in ist es

vor allem wichtig, dass sie*er in der Nähe einer U-Bahn-Station oder des Veranstaltungsortes einen Parkplatz findet. Die App kann die*den Nutzer*in dann zum nächstgelegenen Parkhaus, welches noch nicht ausgelastet ist, leiten. Pendler*innen könnte die aktuelle bzw. zur Ankunftszeit prognostizierte Auslastung von P+R (Park & Ride)-Anlagen oder anderen privaten Garagen in Zielnähe angezeigt werden.

- Services entlang der Route: Bei Abfahrt ist der*dem Fahrer*in bereits bewusst, wie lange sie*er ohne Pause fahren kann und wie viel Treibstoff noch im Tank ist. Die App kann in solchen Fällen nach einer bestimmten, von der*dem Nutzer*in eingegebenen Zeit oder mittels einer „Pausenfunktion“, eine Navigation zur nächstgelegenen Raststation vornehmen. Der Treibstoff an Autobahn-Tankstellen ist meist teurer, als bei jenen, die sich erst bei Auf- oder Abfahrten befinden. Die App kann bei Bedarf auch zur günstigsten Tankstelle mit dem geringsten Umweg navigieren.
 - Wetterereignisse: Neben Staus werden auch Wettervorhersagen (z.B. der ZAMG) entlang der Route berücksichtigt. Es wird vor Extremwetterereignissen wie Regen, Schneefall, Sturm oder Hagel gewarnt.
 - Maut: Buchung bzw. Abrechnung von Streckenmaut, die mittels APNR (Kennzeichenerkennung/Toll by Plate) z.B. über die ÖAMTC-App abgerechnet werden.
- Mehrwert:
 - Nutzer*innen profitieren von einer österreichischen Navigationslösung, die mehr kann, als nur von A nach B zu navigieren. Ihnen werden länderspezifische Zusatzleistungen angeboten, welche bestehende Anbieter*innen noch nicht eingebunden haben.
 - Sofern die App eine entsprechende Reichweite erzielt, können die Daten der Anwender*innen zur Verbesserung der jeweiligen Funktionen einfließen. So könnte beispielsweise bei der kurzfristigen Prognose die P+R-Auslastung berücksichtigt werden, falls mehr ÖAMTC-App-Nutzer*innen als gewöhnlich zu einem bestimmten P+R navigieren.
 - Der ÖAMTC generiert durch zusätzliche App-Nutzer*innen eine höhere Reichweite. Es kann langfristig auch sein, dass die sich daraus ergebende Datenbasis von anderen Navigationsdienstleister*innen zugekauft wird. Die App sollte zugunsten der Verkehrssicherheit weitgehend auf Werbung verzichten. Als Entschädigung werden

Daten zu den Routen und zum Fahrverhalten gesammelt, die in ein übergreifendes Angebot mit einfließen.

- Die Parkhausbetreiber*innen können ebenso von den Daten aus der App und zum Fahrverhalten profitieren, um deren Standortplanung zu optimieren. Außerdem wird durch die Navigation zum nächstgelegenen Parkhaus sichergestellt, dass die Fahrer*innen dorthin und nicht in die Kurzparkzone fahren.
- Potentielle Partner*innen:
 - ÖAMTC: Entwickelt die App und stellt die daraus gewonnenen Erkenntnisse den anderen Partner*innen zur Verfügung. Profitiert von der Datensammlung, der größeren Reichweite der App und der höheren Zufriedenheit der Kund*innen.
 - Parkhausbetreiber*innen: Bereitstellung der Parkhausauslastung in Echtzeit und der Standorte. Nutzen die aufbereiteten Daten aus der App zur Standortoptimierung und profitieren von der besseren Auslastung der Parkhäuser.
 - E-Control: Bereitstellung der Treibstoffpreise in Echtzeit. Nutzt die aufbereiteten Daten aus der App zur Standortoptimierung und zur Feststellung des Fahrverhaltens.
 - SpotOn Statistics: Steuert die Kommunikation und das Anforderungsmanagement zwischen den Partner*innen. Strebt eine Lösung an, von der alle Partner*innen gleichermaßen und zeitnah profitieren.
 - ZAMG: Vorhersage der Wettersituation entlang der Route zwecks Wetterwarnungen.
 - Mautstraßenbetreiber*innen: Nutzen einer einheitlichen Bezahlschnittstelle zur Abrechnung von kennzeichenbasierter Maut über den ÖAMTC User*innen-Account.
 - KFZ-Versicherer*innen: indirekte Stakeholder, da diese z.B. von der Vermeidung von Unwetterschäden profitieren.

1.8. Awareness – Bewusstsein und Vertrauen schaffen

Relevante Säule(n) der EU-Datenstrategie: C

In privatwirtschaftlichen Unternehmen fehlen noch immer häufig das Wissen und die Klarheit, welche Daten vorhanden sind, wie diese verwendet werden können, wie man von Data Sharing profitieren kann, was ein Value und ein Preis wäre, etc. HVD sollten zur Verfügung gestellt werden und dadurch auch eine Bewusstseinsbildung fördern.

Die Berichterstattung über Künstliche Intelligenz (KI) und ähnliche Themen ist häufig sehr negativ besetzt und so entstehen Vorurteile in Bezug auf Verletzung des Datenschutzes, Jobverlust durch

KI, etc. Es muss an einer positiven Darstellung gearbeitet werden, um eine allgemeine Toleranz gegenüber neuen Herausforderungen in Bezug auf Daten zu schaffen. Das könnte z.B. durch Success-Stories vorgenommen werden, mit dem Fokus: Was kann KI wirklich?

Für ein breites Bewusstsein braucht es eine klare Definition aus Sicht der Bürger*innen, der Verwaltung und der Unternehmen. Auch der Begriff der Datensouveränität ist noch weitgehend unbekannt. Zusätzlich ist in diesem Kontext ein Abgleich zwischen den folgenden Konzepten notwendig:

- Datennationalismus vs. Data Sharing
- Datenoffenheit vs. Datensicherheit
- Datenprivatheit vs. Datenhandel

Die erwünschte Awareness für Datenkataloge kann auch durch die oben erwähnten Best-Practices geschaffen werden. Es sollte auch den Unternehmen klarwerden, dass das Teilen von Daten zwischen zwei Wettbewerber*innen besser ist, als wenn die Daten gänzlich unverwendet bleiben. Wichtig ist es, Vertrauen zu schaffen, Möglichkeiten zur Zusammenarbeit aufzuzeigen und das auf relevanten Ebenen.

Eine Bewusstseinsbildung über Datenmärkte wird die nächsten Jahre essentiell prägen. Noch fehlt es massiv an Knowhow, insbesondere bei politischen Entscheidungsträger*innen. Auch die Position der*des Chief Digital Officer (CDO) bedarf einer besseren öffentlichen Aufklärung. Die Aufgaben und Einbindung in die Hierarchie werden sehr unterschiedlich ausgelegt – es braucht eine bessere Positionierung. Die Verwaltung sollte hier als Vorbild vorangehen und der CDO in der Verwaltung als Vorbild mit Außenwirkung fungieren.

1.9. Forschung. Technologie. Innovation.

Relevante Säule(n) der EU-Datenstrategie: B

Hauptanliegen der österreichischen Data Community: Standardisierung ist wichtig, aber aktuell in klassischen Forschungsprojekten kaum finanzierbar. Innerhalb des Marktes wird diese ebenfalls nicht umgesetzt – warum das so ist, sollte von der öffentlichen Hand als Forschungsfrage zugelassen werden.

Aktuelle Unklarheiten im Forschungsbereich sind:

- Müssen Modelle neu berechnet werden, wenn User*innen bei den bereits berechneten Modellen abspringen, weil sie ihre Daten nicht mehr teilen möchten? (vorgeschlagene Lösung: Federated Learning)
- Was sind denn eigentlich personenbezogene Daten? Mit welchem Aufwand können personenbezogene Daten aus Modellen wieder herausgerechnet werden?
- Wie müssen Forschungsprojekte angelegt werden, damit die erworbene Finanzierung auch eingesetzt werden kann, um mit den Plattformen anderer Nationen zusammenzuarbeiten?
- Wie kann unter Verwendung des eigenen Forschungsgeldes an ein bestehendes Forschungsprojekt eines anderen Landes angeknüpft werden?
- Wie können hochsensible Daten der Forschung zur Verfügung gestellt werden?
- Wie ist mit fehlerhaften Daten und weiterführende Datenschäden umzugehen?

Folgende Kernfragen stellen sich für den Gesamtkomplex von Forschung, Technologie, und Innovation:

Generell:

- *Wem gehören welche Daten?* (Bsp. Sensordaten: Gehören die Daten denen, die beobachtet werden, den Beobachter*innen oder den Equipmenthersteller*innen?)
- *Was ist der konkrete Mehrwert von Open-Data-Ansätzen für Unternehmen, öffentliche Hand und Gesellschaft?*
- *Wie kann Datenqualität hergestellt werden und wie kann man Datenqualität konkret messen?*
- *Welche Mehrheit bringen Data Governance und Standardisierung für Unternehmen und wie können diese Aspekte sinnvoll und wertstiftend in die Firmenstrategie implementiert werden?*
- *Wie kann eine sinnvolle Nutzer*innen-Remuneration für Data Sharing umgesetzt werden? Können vorübergehende Membership-Modelle Abhilfe schaffen?*
- Etwaige Lizenzfragen
- Haftungsfragen bei Datenqualität

EU-Datenstrategie:

- *Trägt die europäische Datenstrategie dazu bei, dass Wettbewerbsvorteile für europäische Unternehmen entstehen oder profitieren primär Regierungen davon?*

Der Verein zur Förderung der Datenwirtschaft und der Optimierung von Datentechnologien – gegründet 2018

- *Welche regionalen Unterschiede gibt es in Bezug auf die europäische Datenstrategie innerhalb der Europäischen Union? Wie wirken sich diese auf die Datenforschung und die jeweiligen Datenökonomien aus?*

Öffentliche Hand:

- *Warum wird seitens der öffentlichen Hand nicht mehr publiziert, um Open-Data-Ansätze damit zu unterstützen? Was verhindert die extensive Publikation von Daten?*
- *Können öffentliche Daten per se als Allgemeingut angesehen werden? Warum, bzw. warum nicht?*
- *Welche Vorteile bzw. welche Kosten bringt eine nationalstaatliche Cloud?*
- *Unterliegen öffentliche Daten überhaupt dem Urheberrecht? Warum, bzw. warum nicht?*
- *Wie kann Anonymisierung bei einer digitalen Bürger*innen-App DSGVO-konform und gemäß europäischen Richtlinien umgesetzt werden?*

TEIL 2: Themen aus Live-Diskussionen der Data Community

2.1. Datenmanagement

Erfolgreiches Datenmanagement ist von höchster Wichtigkeit für Unternehmen, damit diese Mehrwert aus ihren Daten generieren können. Dies umfasst all jene Technologien und Methoden, die zum erfolgreichen Abschluss der einzelnen Phasen eines Data-Science-Projekts benötigt werden. Dazu gehören die Datenakquise, -bereinigung und Qualitätsevaluierung, Datenmanagementplattformen zum Datenaustausch, -handel (Interoperabilität) und die -bereitstellung für Dritte (Data Sharing), notwendige Protokolle, die Datenarchivierung und -aufbewahrung, Datenverarbeitung und -analyse, sowie die Aspekte der Datensouveränität, Eigentumsrechte, und Selbstkontrolle der Daten.

2.2. Datenanalyse und -visualisierung

Mit den Methoden der Datenanalyse werden Muster und Gemeinsamkeiten aus Rohdaten extrahiert und daraus Schlussfolgerungen und Implikationen gezogen. Diese können in umsetzbare Handlungsanweisungen transferiert werden und Unternehmen und Organisationen bei der Optimierung ihrer Prozesse unterstützen. Die Datenvisualisierung dient dazu, der*dem Anwender*in ein Bild der zugrundeliegenden Eigenheiten der Daten zu vermitteln und verschafft Übersicht sowie Fokus auf das Wesentliche.

2.3. Verarbeitung natürlicher Sprachen

Der Bereich der Verarbeitung natürlicher Sprachen hat in den letzten Jahren durch die Innovationen in neuronalen Netzen und hoch performanten Sprachmodellen (BERT, GPT-3) gewaltige Fortschritte gemacht. Diese erlauben Anwendungen wie beispielsweise die automatische Texterstellung mit bisher ungekannter Genauigkeit und auf beinahe menschlichem Niveau. Sprachmodelle existieren mittlerweile für unterschiedliche Sprachen. Mögliche

Anwendungsgebiete sind der Einsatz in Technologien für die Moderation von Online-Kommunikation, beispielsweise für die Erkennung demokratiegefährdender Inhalte in sozialen Medien, Erkennung von polarisierender und diskriminierender Sprache, Frühwarnmechanismen beim Auftauchen subversiver Inhalte. Weiters gibt es Anwendungen im Bereich der Texterzeugung, beispielsweise für Conversational Agents (Chatbots) sowie im Bereich des automatischen Journalismus, wo einfache, repetitive und wenig kreative Texte wie Wetterberichte oder Fußballergebnisse durchaus auch von automatischen Systemen geschrieben werden können.

2.4. Gesundheitsbereich

Hier werden unterschiedliche gesundheitsrelevante Bereiche abgedeckt, wie beispielsweise die Verarbeitung von Patient*innenakten, Gendaten, und das Verständnis über die Effektivität von Behandlungsmethoden.

2.5. Technologiegestütztes Lernen

Hierunter fallen Aspekte wie (Social) Micro-Learning oder die Analyse des bisherigen Lernerfolgs mit nachfolgender Methodenoptimierung. Speziell in einer von Covid-19 geprägten Ära mit daraus resultierender Notwendigkeit für virtuellen Unterricht ist dieser Aspekt von großer Bedeutung.

2.6. Gemeinschaftliche Nutzung von Daten über Unternehmensgrenzen hinweg

Hierbei geht es darum, die Wettbewerbsfähigkeit von Unternehmen durch die Optimierung von Prozessen zum Datenaustausch und -handel über Unternehmensgrenzen hinweg zu erhöhen. Dies umfasst einerseits die Erforschung von Technologien zum Datenaustausch, aber auch die Entwicklung von Best-Practices, um operative Prozesse in Unternehmen anzuleiten und Ineffizienzen aus dem Weg zu räumen. Durch die Kombination unternehmensinterner Daten mit neuen und neuartigen Daten steigert sich das Innovationspotential von Datenanwendungen.

2.7. Effiziente Handhabung produktionsrelevanter Daten

Unternehmen fehlt oft das Wissen über den effizienten Umgang mit produktionsrelevanten Daten. Dadurch entsteht im internationalen Vergleich ein Wettbewerbsnachteil. Es sollen Kompetenzen und Expertise dafür aufgebaut werden, welche Technologien und Methoden am geeignetsten für den Umgang mit großen Datenmengen sind. Weiters soll vermittelt werden, wie stark differenzierte Technologie-Stacks homogenisiert und miteinander verbunden werden können. Dies steigert die Effizienz des Datenaustausches zwischen Datensilos. Weiters schafft es die Voraussetzung für

Verknüpfungen über Unternehmensgrenzen hinweg und legt letztlich auch die Grundlage für eine effektive Kommunikation mit GAIA-X.

2.8. Anwendung moderner KI-Verfahren auf Produktionsdaten

Die letzten Jahre haben gewaltige Verbesserungen im Bereich der KI-Methoden gebracht, insbesondere im Bereich der im Moment sehr geläufigen neuronalen Netze (Stichwort: Deep-Learning). Diese Methoden haben auch im Bereich produktionsrelevanter Daten enormes Potential, das noch ausgeschöpft werden sollte.

Zusätzlich dazu sind die Methoden *Transfer Learning* und *Explainable AI* hochrelevant. Bei ersterem handelt es sich um Methoden, um die auf einer Domäne erstellten Modelle auf eine andere Domäne zu übertragen. Dies hat den Vorteil, dass auch Domänen mit geringen vorhandenen Datenmengen bearbeitet werden könnten. Aus Unternehmenssicht ist diese Methode relevant, da die in einem Unternehmen trainierten Modelle anderen Unternehmen zur Verfügung gestellt werden können, auch, wenn diese nicht aus demselben Fachbereich kommen. *Explainable AI* hat das Potential, das Vertrauen von Skeptiker*innen in KI-Methoden zu erhöhen. Es stellt Methoden zur Verfügung, die eine Erklärbarkeit von Entscheidungen von KI-Modellen ermöglichen. Dadurch wird der oft verurteilte "Black-Box"-Charakter von KI-Modellen in eine transparente Entscheidungsrepräsentation umgewandelt. Die Entscheidung der Maschine kann damit nachverfolgt und verständlich gemacht werden.

2.9. Teilnahme an GAIA-X

Die Initiative GAIA-X hat es sich zum Ziel gesetzt, eine pan-europäische Infrastruktur zum Datenaustausch aufzubauen. Dabei wird großes Augenmerk auf eine Vereinbarkeit mit den grundlegenden Werten der Europäischen Union, wie beispielsweise Datensouveränität, Privatsphäre und Datensicherheit in Abgrenzung zum Umgang anderer Nationen mit Daten gezogen. Es ist wichtig, dass Unternehmen die Kompetenzen, Expertisen und technologischen Methoden zur Verfügung gestellt werden, um an dieser Initiative teilnehmen zu können und davon zu profitieren. GAIA-X befindet sich noch im Konzeptstadium, gerade deswegen ist es wichtig, möglichst rechtzeitig hier anzusetzen, um eine wegweisende Funktion übernehmen zu können.