HMD Praxis der Wirtschaftsinformatik (2026) 63:826–840
https://doi.org/10.1365/s40702-026-01275-y

SCHWERPUNKT

Vom Datenmonopol zur Datensouveränität – Selbstbestimmte Datenflüsse als Treiber digitaler Verantwortung

Léon Dankert · Tobias Dreesbach · Dennis Maecker · Aaron Witzki · K. Valerie Carl · Oliver Thomas

Eingegangen: 18. Januar 2026 / Angenommen: 26. April 2026 / Online publiziert: 20. Mai 2026
© The Author(s) 2026

Zusammenfassung Die Verfügbarkeit qualitativ hochwertiger Daten über Unternehmensgrenzen hinweg ist eine Grundvoraussetzung für die Entwicklung fairer und zuverlässiger KI-basierter Systeme. Bestehende Datenökosysteme fokussieren jedoch primär den Austausch zwischen Unternehmen, während Endnutzende als Datenerzeugende weitgehend passive Akteure bleiben. Um Endnutzenden eine faire Datensouveränität zu ermöglichen, ergibt sich die Notwendigkeit neuer Ansätze. Dieses Paper untersucht, wie Governance-Mechanismen gestaltet werden müssen, um Datensouveränität, Datenverfügbarkeit, Transparenz und Nachvollziehbarkeit zu vereinen und somit aus Sicht von Corporate Digital Responsibility (CDR) verantwortungsvolle Systeme sicherzustellen. Basierend auf konzeptionellen Überlegungen zu dezentraler Datenhaltung und einer Analyse von Anwendungsfällen aus dem Bereich Smart Living wird ein Governance-Framework erarbeitet. Das Framework strukturiert zentrale Rollen, Prozesse und Mechanismen entlang des Datenlebenszyklus und ermöglicht eine systematische Bewertung hinsichtlich Umsetzbarkeit, Wirksamkeit und Beitrag zur Datenverfügbarkeit. Zudem werden praxisnahe Ansätze zur Umsetzung von Datensouveränität über reine regulatorische Mindestanforderungen der DSGVO hinaus aufgezeigt. Das Paper liefert damit einen Beitrag zu CDR für Datenökosysteme sowie zur Gestaltung fairer Datenakquise in KI-Kontexten.

Schlüsselwörter Datensouveränität · Dezentrale Datenhaltung · Corporate Digital Responsibility · Datenmanagement · Data-Governance

Léon Dankert · Tobias Dreesbach · Dennis Maecker · Oliver Thomas
Deutsches Forschungszentrum für Künstliche Intelligenz, Smart Enterprise Engineering, Hamburger Straße 24, 49084 Osnabrück, Deutschland

Aaron Witzki · ✉ K. Valerie Carl
Professur für Wirtschaftsinformatik und Informationsmanagement, Goethe Universität Frankfurt/Main, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt/Main, Deutschland
E-Mail: kcarl@wiwi.uni-frankfurt.de

Abstract The availability of high-quality data across company boundaries is a fundamental prerequisite for the development of fair and reliable AI-based systems. Existing data ecosystems, however, primarily focus on exchange between companies, while end users, as data generators, largely remain passive actors. To enable fair data sovereignty for end users, new approaches are therefore required. This paper examines how governance mechanisms must be designed in order to combine data sovereignty, data availability, transparency, and traceability, and thus ensure responsible systems from the perspective of Corporate Digital Responsibility (CDR). Based on conceptual considerations regarding decentralized data storage and an analysis of use cases in the field of Smart Living, a governance framework is developed. The framework structures central roles, processes, and mechanisms along the data lifecycle and enables a systematic assessment with regard to feasibility, effectiveness, and contribution to data availability. In addition, practical approaches are presented for implementing data sovereignty beyond the mere minimum regulatory requirements of the GDPR. The paper thereby contributes to CDR for data ecosystems as well as to the design of fair data acquisition in AI contexts.

Keywords Data sovereignty · Decentralized data storage · Coporate digital responsibility · Data management · Data governance

1 Einleitung

Die fortschreitende Digitalisierung und der umfassende Einsatz künstlicher Intelligenz (KI) erfordern den Zugang zu qualitativ hochwertigen Daten über Unternehmensgrenzen hinweg. Während etablierte Datenökosysteme primär den Datenaustausch zwischen Organisationen fokussieren (Oliveira et al. 2019), bleiben Privatpersonen (als Endnutzende bspw. von intelligenten Haushaltsgeräten, -steuerungen und Applikationen) als Datenerzeugende weitgehend passive Akteure ohne echte Kontrolle über die unternehmensübergreifende Weitergabe ihrer Daten (Tallon 2013). Die aktive Einbindung von Endnutzenden als souveräne Entitäten in Datenökosystemen bietet das Potenzial, Transparenz über Datenflüsse zu schaffen, Vertrauen zu stärken und gleichzeitig die Datenverfügbarkeit für innovative KI-basierte Services zu erhöhen (Dreesbach et al., 2025b).

Derzeit verbleiben jedoch viele in Anwendungen und Internet of Things (IoT)-Geräten erzeugte Daten in isolierten Systemen einzelner Unternehmen, sodass ihre Verfügbarkeit für andere Unternehmen und die Entwicklung und den Betrieb KI-basierter Services stark eingeschränkt ist (Tallon 2013). Wo Daten weitergegeben werden, geschieht dies teils intransparent und ohne dass Endnutzende eine echte, fortlaufende Souveränität über Zwecke, Empfänger und Dauer der Nutzung ausüben können. Nur selten wird transparent kommuniziert, inwiefern die Nutzung generierter, teils sogar personenbezogener Daten für Endnutzende unerwartet über Unternehmensgrenzen hinweg erfolgt und ob sie tatsächlich Kontrolle über die Weitergabe ihrer Daten ausüben können. Diese Problematik gewinnt an Bedeutung, da fehlende Einwilligungen bei der Datenerhebung für Modelltraining immer stärker

in den regulatorischen Fokus rückt und Fairness zunehmend wichtiger wird (Christ et al. 2025).

Um diesen Herausforderungen zu begegnen, untersucht dieser Beitrag, wie ein Governance-Framework die divergenten Anforderungen von Datensouveränität, -transparenz und Nachverfolgbarkeit auf der einen und Datenverfügbarkeit auf der anderen Seite integrieren kann. Im Fokus steht, unter welchen Bedingungen verschiedene Governance-Ansätze für Unternehmen strategisch relevant werden und zur Corporate Digital Responsibility (CDR) beitragen. Dabei bezieht der Beitrag sich auf den Anwendungskontext von Smart Living, da es sich bei diesen Daten zum einen um sehr private, persönliche Daten handelt und sie andererseits für Endnutzende sehr präsent sind. Die zugrunde liegende Forschungsfrage lautet daher: *Welche Mechanismen enthält ein Governance-Framework im Sinne von CDR, um die teilweise divergenten Anforderungen von Datensouveränität, Datentransparenz, Nachverfolgbarkeit und Datenverfügbarkeit zusammenzuführen und ihren Aufwand entlang der Datenerhebung und -verarbeitung abzuschätzen?*

Durch die systematische Betrachtung unterschiedlicher Unternehmensinteressen und Datencharakteristika wird analysiert, wie sich Anforderungen an Governance-Mechanismen unterscheiden und welche Implikationen dies für die Gestaltung fairer Datenakquise hat. Das aufgestellte Governance-Framework soll Endnutzenden echte Entscheidungsfreiheit eröffnen (z. B. über verständliche, standardisierte Einwilligungsprozesse) und Datenflüsse transparent und nachvollziehbar machen. Gleichzeitig berücksichtigt das Framework den Aspekt, dass eine verantwortungsvolle Datensouveränität nicht im Widerspruch zu unternehmerischen Interessen an Datenverfügbarkeit stehen sollte, sondern vielmehr langfristige Wettbewerbsvorteile durch den Aufbau von Vertrauen und die Ermöglichung fairer Datenakquise generiert.

2 Grundlagen dezentraler Governance-Mechanismen

In Bezug auf Daten umfasst der Begriff Governance Richtlinien und Strukturen, die den Umgang mit Datenbeständen im Unternehmen und darüber hinaus vorgeben (Tallon 2013). Eine Governance bewertet den Schutz von Datenbeständen vor verschiedenen Risikofaktoren (Tallon 2013) und verteilt Pflichten und Rechte unter den beteiligten Stakeholdern (Oliveira et al. 2019). Sowohl einzelne Unternehmen als auch unternehmensübergreifende Datenökosysteme schaffen durch eine klar definierte Governance strukturelle Rahmenbedingungen für den sicheren Datenaustausch zwischen Organisationen.

Der Forschungsstand zu Datenökosystemen sowie zu deren konkreten Ausprägungen, etwa in Form sogenannter Datenräume als föderierte Infrastrukturen für den organisationsübergreifenden Datenaustausch, ist trotz wachsender Relevanz noch durch stark variierende Definitionen geprägt. Die Infrastruktur für Datenökosysteme kann von unterschiedlichen Akteuren bereitgestellt und koordiniert werden, etwa durch dominierende Teilnehmer oder durch spezialisierte, potenziell neutrale Intermediäre (z. B. Datenraum-Betreiber). Beispiele für föderierte Datenräume sind konkrete Umsetzungen wie Catena-X oder Smart-Living-NEXT, während Initiativen wie GAIA-X die zugrunde liegenden Konzepte und Infrastrukturbausteine

entwickeln. Governance-Mechanismen gelten jedoch übereinstimmend als essenzieller Bestandteil solcher Datenökosysteme (Oliveira et al. 2019). Sie stärken die Datensouveränität einzelner Parteien und bieten Schutzmechanismen (Tallon 2013). Governance-Mechanismen gehen über den Zugriff auf Daten hinaus und umfassen Nutzungsregeln, die u. a. Zweckbindungen, Weitergabeverbote, Rollenbeschränkungen sowie Pflichten wie Löschfristen explizit und nachvollziehbar abbilden (Zrenner et al. 2019). Besonders wichtig ist dies im Kontext von nutzergenerierten Daten, da dort neben direkt eingebundenen auch indirekt Betroffene ohne Vertragsbeziehung in den Verantwortungsbereich fallen können, u. a. Privatpersonen (Christ et al. 2025). Nutzergenerierte Daten erweitern die Governance-Anforderungen damit über klassische Vertrags- und business-to-business (B2B)-Grenzen hinaus (Christ et al. 2025).

Einzelpersonen als Endnutzende und somit Datengenerierende werden in Datenökosystemen bislang häufig vernachlässigt, obwohl es vielfach ihre Daten sind, die die zugrunde liegende Wertschöpfung ermöglichen. Auch in Referenzarchitekturen von Datenökosystemen mit Datensouveränität stehen bislang vor allem Beziehungen zwischen Unternehmen im Vordergrund (Scheider et al. 2024). Eine Perspektive unter Ausschluss von Endnutzenden greift jedoch zu kurz, wenn es im Sinne von CDR um die Bewertung gesamtgesellschaftlicher Auswirkungen datenbezogener Praktiken geht. Koskinen et al. (2019) stellen einen ersten Ansatz für ein Ökosystem vor, das auf den Menschen ausgerichtet ist und in dem die Interessen der Nutzenden über klare Regeln und eine starke Governance-Repräsentation systematisch verankert werden. Eine weitere passende Erweiterung des Fokus auf Endnutzende in Datenökosystemen eröffnet die Verbindung zu Konzepten wie CDR. CDR verdeutlicht, dass der reine Fokus auf den Austausch von Daten zwischen Unternehmen nicht ausreicht. „CDR steht für ganzheitliches verantwortungsbewusstes Handeln von Unternehmen in Bezug auf die Auswirkungen digitaler Technologien und Daten“ (Carl et al. 2025) und damit die freiwillige Verantwortung von Unternehmen für die gesellschaftlichen und ethischen Folgen digitaler Technologien und datengetriebener Geschäftsmodelle (Lobschat et al. 2021; Mueller 2022). Während Aspekte wie Fairness und Transparenz algorithmischer Entscheidungen Bestandteil von CDR sind, ist eine für Endnutzende faire Datensammlung bislang weniger adressiert (Carl und Hinz 2024).

Im Sinne einer verantwortungsvollen Datensammlung und -nutzung, stellen Endnutzende eine eigenständige Akteursgruppe dar, deren Rolle über die reine Zustimmung hinausgehen sollte und die aktiv in die Ausgestaltung von Datenflüssen einbezogen werden muss. Damit verschiebt sich Governance auf eine neue Perspektive: First Party (Service- oder App-Anbieter), Second Party (Endnutzende) und Third Party (weitere Unternehmen) interagieren innerhalb eines gemeinsamen Datenökosystems. Im Sinne einer fairen Datensammlung sollen Endnutzende als Second Party bewusst, informiert und freiwillig über die Nutzung ihrer Daten entscheiden können (Carl 2025). Datensouveränität meint dabei nicht nur die Möglichkeit, einmalig in AGB einzuwilligen, sondern die kontinuierliche Kontrolle über konkrete Datenzugriffe und -weitergaben und damit eine aktive, ggf. situative Zustimmung anstatt einer umfassenden einmaligen und oftmals nicht bewusst informierten Einwilligung.

Aus Unternehmensperspektive adressiert faire Datensammlung das Henne-Ei-Problem von Datenknappheit und Bedarf an Trainingsdaten und ermöglicht sichtba-

res CDR-Engagement. Dies gilt für Unternehmen aller Größen, ist für KMUs jedoch besonders relevant, da sie meist über weniger Daten und Ressourcen verfügen und stärker auf faire Datenzugänge angewiesen sind.

3 Gewährleisten von Datensouveränität am Beispiel eines Smart Living Ökosystems

Datenökosysteme ermöglichen den kontrollierten Austausch von Daten zwischen Organisationen, wobei jede Partei über eigene Konnektoren entscheidet, welche Datenausschnitte zu welchem Zweck freigegeben werden (Strnadl und Schöning 2023). Im Smart-Living-Ökosystem Smart-Living-NEXT wird dieser Ansatz herstellerübergreifend umgesetzt: Die Infrastruktur ist föderiert, die Datenhoheit verbleibt bei den teilnehmenden Unternehmen, und gleichzeitig wird ein datenschutzkonformer, zweckgebundener Zugriff als Grundlage datenbasierter Services ermöglicht.

Das Projekt COMET ist Teil von Smart-Living-NEXT und zielt darauf ab, Endnutzende von passiven Datenlieferanten zu souveränen Gestaltenden ihrer Datenflüsse zu machen, während gleichzeitig Unternehmen verlässlichen Zugang zu qualitativ hochwertigen Daten erhalten sollen (Carl 2025; Dreesbach et al. 2025b). In diesem Kontext ergeben sich drei zentrale Anforderungen an Datensouveränität, für die jeweils Governance-Mechanismen abgeleitet werden können.

Als erste Anforderung muss Datennutzung transparent sein, etwa durch verständliche Visualisierungen und Nutzungstracking, sodass Datengebende Zweck, Umfang und Nutzung nachvollziehen können.

Als zweite Anforderung ist eine Balance zwischen Kontrolle und Interaktionsaufwand erforderlich. Die im Projekt erarbeiteten Governance-Mechanismen folgen dabei überwiegend einem dezentralen Ansatz, bei dem Daten möglichst nahe an ihrer Quelle gespeichert und verarbeitet werden, idealerweise auf den Endgeräten. Dies stärkt die Datensouveränität, bringt jedoch auch technische Herausforderungen mit sich, insbesondere hinsichtlich erhöhtem lokalem Speichervolumen sowie Anforderungen für Backups und Wiederherstellungen. Sollen Daten von der First Party oder einem Unternehmen/Datenökosystemteilnehmer (Third Party) verwendet werden, müssen die Daten zunächst bei Second Party angefragt werden. Die Second Party soll durch diese erhöhte Kontrollmöglichkeit in die Lage versetzt werden, selbst zu entscheiden, ob und wann sie Daten bereitstellt, ohne dabei durch eine hohe Frequenz an Anfragen und den damit verbundenen Interaktionsaufwand überlastet zu werden.

Als dritte Anforderung müssen freigegebene Daten strikt zweckgebunden verarbeitet werden: Daten sollen jeweils nur einer konkreten Anfrage zugeordnet und nicht langfristig gespeichert werden. Damit andere Ökosystemteilnehmer erkennen können, welche Daten prinzipiell erfragbar sind, ist ein Metadatenkatalog erforderlich, auf den sie Zugriff haben. Die Metadaten beschreiben lediglich das Vorhandensein potenziell verfügbarer Datensätze auf der Seite der Second Party, ohne die Daten selbst dauerhaft im System zu halten.

4 Governance-Mechanismen im Detail

4.1 Metadatenverwaltung und Datenkatalog

Metadatenkataloge stellen eine strukturierte Übersicht über verfügbare Daten und deren Eigenschaften bereit und werden in der Literatur als zentrale Bausteine datengetriebener Organisationen diskutiert (Petrik et al. 2024; Scheider und Mallick 2025). Sie unterstützen Transparenz und Datensouveränität, da aus dem Metadatenkatalog ersichtlich wird, ob für einen bestimmten Zweck prinzipiell geeignete Daten vorliegen, ohne dass bereits Rohdaten oder konkrete personenbezogene Datensätze offengelegt werden müssen (Petrik et al. 2024). Zugriffsrichtlinien und Prüfungen können über ein Regelwerk vorgegeben werden (Ehrlinger et al. 2021).

Um Datenzugang von Anwendungen transparent zu gestalten, können App- und Service-Anbieter nutzungsbezogene Metadaten in einem solchen Metadatenkatalog bereitstellen, etwa zu Datentyp, Struktur, zulässigen Nutzungen und regulatorischen Einschränkungen (Ehrlinger et al. 2021; Jahnke und Otto 2023). Besonders für KMU ist ein Metadatenkatalog bei Datenanfragen relevant, da diese häufig über begrenzte Ressourcen für die Identifikation, Dokumentation und Abstimmung von Datenzugängen verfügen. Ein Metadatenkatalog kann die Verfügbarkeit von Daten sichtbar machen und Such- sowie Abstimmungsaufwände über Unternehmensgrenzen hinweg reduzieren. In der Praxis liegen Herausforderung insbesondere bei einem hohen initialen Erstellungsaufwand sowie den Anforderungen an eine konsistente und automatisierte Pflege des Metadatenkatalogs.

4.2 Standardisierte Datenanfrage und KI-gestützte Vorfilterung

Die Beziehung zwischen First Parties (Service- bzw. App-Anbieter), beispielsweise Anbietern digitaler Smart-Home-Anwendungen, und Second Parties (Endnutzende) bildet die Grundlage für vertrauenswürdige datenbasierte Services (Koskinen et al. 2019). Third Parties (Drittunternehmen) entwickeln typischerweise ergänzende Auswertungen, datenbasierte Mehrwertdienste oder KI-gestützte Funktionen, die über den originären Leistungsumfang der First Party hinausgehen und so zusätzliche Nutzenpotenziale für die First oder Second Party erschließen.

Das übergreifende Ziel standardisierter Datenanfragen besteht darin, die Beziehung zwischen den Parteien durch eine konsistente Data Governance zu stärken und transparente, nachvollziehbare Entscheidungsprozesse bei Datenanfragen zu etablieren (Geisler et al. 2022; Scheider et al. 2024). Eine Standardisierung des Anfrage- und Prüfprozesses setzt ein einheitliches, maschinenlesbares Format voraus, das eine automatisierte Verarbeitung sowie eine systematische Vergleichbarkeit und konsistente Bewertung unterstützt. Dies reduziert Interpretationsspielräume und entlastet Second Parties bei informierten Entscheidungen über die Freigabe ihrer Daten, indem alle entscheidungsrelevanten Informationen wie Datentypen, Zweck, Nutzungsdauer und anfragende Partei strukturiert gebündelt werden (Carl 2025, Witzki et al. 2025). Third Parties sollten ausschließlich auf klar begrenzte, zweckgebundene Datenausschnitte zugreifen können (Carl 2025). Bei wiederkehrenden Anfragen können Zustimmungen zur Datenfreigabe von der Second Party lokal verwaltet und defi-

nierte Präferenzen automatisiert auf entsprechende Anfragekategorien angewendet werden. Gleichzeitig kann ein optional vorgeschalteter Filter unzulässige Anfragen erkennen oder fehlende Angaben markieren. Beispielsweise kann ein Large Language Model (LLM) als Teil des Vorfilters Freitextfelder auswerten und Datentypen mit dem angegebenen Zweck vergleichen.

Eine mögliche Implementierungsvariante sieht vor, dass Datenanfragen zwar serverseitig spezifiziert werden, aber ausschließlich durch das Endgerät der nutzenden Person aktiviert werden (z. B. durch regelmäßigen Abruf oder lokale Bestätigung). Dadurch verbleibt der Kontrollpunkt für den Datenzugriff bei der Second Party und stärkt die Wahrnehmung souveräner, nicht ferngesteuerter Entscheidungen.

4.3 Granulare Einwilligungssteuerung

Die granulare Einwilligungssteuerung ist ein zentraler Baustein der Datensouveränität (Munoz-Arcentales et al. 2019). Anstatt generischer Einmalzustimmungen, die dauerhaft gelten, erhalten Second Parties (sofern gewünscht) individualisierte Anfragen für jeden beabsichtigten Datenzugriff. Eine granulare Einwilligungssteuerung ist über eine Verwaltung der Einzelanfragen in einem Einwilligungsregister abbildbar (Witzki et al. 2025). Über ein Einwilligungsportal können Second Parties das Register einsehen, Einwilligungen erteilen, anpassen oder widerrufen.

Dabei ist nicht vorgesehen, dass jede einzelne Anfrage zwingend manuell bestätigt werden muss. Stattdessen kann ein abgestuftes Modell implementiert werden, das zwischen einmaligen Grundzustimmungen, blockweisen Freigaben (z. B. nach Zweck oder anfragender Organisation) und individuellen Einzelentscheidungen differenziert. First Parties benötigen zu Beginn häufig breitere Einwilligungen, während neue Zwecke im Verlauf der Anwendung zusätzliche Anfragen auslösen. LLM-basierte Komponenten können die Anfragen für die Second Party verständlich zusammenfassen und präferenzbasiert bei der Einwilligung unterstützen. Für Third Parties können pauschale oder individuelle Prüfungen definiert und bestehende Einwilligungen regelmäßig überprüft werden. Herausforderungen entstehen vor allem durch Integrationsaufwand und die Verknüpfung von Einwilligungslogik und Backend-Verarbeitung. Ein konzeptionell denkbarer, separat angebundener „Authenticator“ könnte die Notwendigkeit eigener Einwilligungssysteme reduzieren. Vergleichbar mit einer Authenticator-App könnte ein solcher Ansatz den Implementierungsaufwand senken und auch kleineren Unternehmen eine strukturierte Governance für ihre Anwendungen ermöglichen.

4.4 Evaluierung der Datenqualität

Qualitätsstandards und eine automatisierte Bewertung von Datensätzen ermöglichen Unternehmen nicht nur eine Grundlage für Entscheidungen über die Nutzbarkeit der Daten, sie helfen auch fehlerhafte oder unvollständige Datensätze früh auszufiltern und die Risiken schlechter Datenqualität zu reduzieren. Eine automatisierte Bewertung und Qualitätsanzeige ermöglicht Second Parties den Zustand ihrer Daten einzuschätzen. Dies trägt wesentlich dazu bei, das Vertrauen der Second Parties in den Umgang mit Daten zu stärken. Für First Parties stellen qualitätsgeprüfte Daten-

sätze eine verlässliche Grundlage dar, um ihre datenbasierten Dienste konsistent und sicher zu betreiben und neue Angebote zu entwickeln. Datenqualität ist besonders bei der Entwicklung KI-basierter Modelle von entscheidender Bedeutung (Cai und Zhu 2015). Third Parties wiederum profitieren von den Qualitätsanforderungen, da sie gezielt auf geprüfte und eindeutig abgegrenzte Datenausschnitte zugreifen können. Neben dem semantischen Kontext sollte Datenqualität entlang standardisierter Dimensionen, wie Vollständigkeit, Aktualität und Plausibilität geprüft werden (Cai und Zhu 2015; Dressbach et al. 2025b). Auch ein semantischer Abgleich zwischen der Beschreibung des angeforderten Datensatzes und des tatsächlich gelieferten Inhalts ist von Bedeutung. Herausfordernd sind dabei insbesondere die Entwicklung und Etablierung solcher Qualitätsstandards und Automatisierungen, wie in (Dressbach et al. 2025a) umgesetzt.

4.5 Visualisierung und Nutzungstracking

Um Datensouveränität über den Zeitpunkt der Einwilligung hinaus zu erhöhen, sollten sich Second Parties auch nach der Datenanfrage weiterhin in einer grafischen Oberfläche informieren können, wer (welches Unternehmen) ihre Daten warum (welcher Zweck und Service) und für welchen Zeitraum verwendet hat. Diese Informationen können über das in Kapitel 4.3 aufgezeigte Einwilligungsportal bereitgestellt werden.

In einer dezentralen Umsetzung verlässt lediglich ein reduzierter Informationsanteil das Endgerät der Second Party, der für die Visualisierung der Datennutzung erforderlich ist (z. B. Dienst, Zweck, Zugriffshäufigkeit). Der dezentrale Ansatz stärkt somit Privacy-by-Design, da zugrunde liegende Daten so nicht vollständig zentral bei der First Party vorgehalten werden müssen. Die Wahl zwischen zentralen und dezentralen Architekturen ist dabei aber als Trade-off zu verstehen, bei dem Datensouveränität, Systemeffizienz und technische Umsetzbarkeit gegeneinander abgewogen werden müssen. So können zum Beispiel bei dem empfohlenen dezentralen Ansatz lediglich Metadaten von der First Party erhoben werden, die allgemeine Aussagen über die voraussichtlichen Daten treffen. Feingranulare oder nachträgliche Analysen über alle Second Parties hinweg sind nur eingeschränkt möglich und nur wenn die First Party die Daten erst selbst zweckgebunden anfragen.

Ein nächster konzeptioneller Ausbauschritt wäre ein durchgängiges Tracking der einmal freigegebenen und übertragenen Daten. Um das höchste Maß an Datensouveränität zu ermöglichen, müsste die spätere Nutzung der Daten vollständig nachvollziehbar und kontrollierbar sein (Zrenner et al. 2019). Ein durchgängiges, verlässliches Tracking der realen Datennutzung ist in der Praxis jedoch schwer realisierbar, da First Parties, genauso wie Third Parties, verschiedene Systemarchitekturen, Standards und Prozesse nutzen. Eine strikte Nachvollziehbarkeit von Datenflüssen könnte zukünftig zwar Vertrauen stärken, gleichzeitig würde der Druck steigen, jede Nutzung in Bezug auf die hinterlegte Einwilligung zu begründen. Gerade für KMUs ist daher eine ausgewogene Gestaltung relevant, die Transparenz schafft, ohne Innovationsspielräume vollständig zu blockieren.

4.6 Aggregation und Anonymisierung von Daten

Standardisierte Anonymisierungs- und Aggregationsmechanismen bieten das Potenzial, Re-Identifikationsrisiken zu reduzieren und zugleich mehr Spielraum für eine weitergehende, rechtlich und ethisch vertretbare Datennutzung zu schaffen. LLMs können zur Anonymisierung beitragen und DSGVO-Konformität zusätzlich absichern, indem sie sensible Attribute erkennen und Vorschläge zur Transformation machen. Beispielweise kann eine lokale LLM-Instanz als vorgelagerte Anonymisierungsschwelle vor einer Weitergabe an Dritte prüfen, ob Daten ausreichend anonymisiert sind und mittels einer Sicherheitsschwelle Übermittlungen blockieren oder mit Warnhinweisen versehen. Insbesondere für Privatpersonen erhöht dies die Sicherheit, dass nur die im Vorfeld klar definierten Daten und erst nach expliziter Freigabe von ihrem Gerät an einen Service übertragen werden. Wird Endnutzenden außerdem transparent vermittelt, zu welchen Zwecken Daten angefragt werden und dass diese nur in anonymisierter Form weitergegeben werden (insbesondere über Unternehmensgrenzen hinweg), sinken wahrgenommene Risiken gegenüber einer Freigabe.

4.7 Implementieren der Mechanismen im Kontext des Projekts COMET

Um die zuvor beschriebenen Mechanismen modular und skalierbar für Anfragen an Second Parties umzusetzen, wurden etablierte DevOps- und Cloud-Native-Prinzipien eingesetzt. Diese agieren bei Datenanfragen als Middleware zwischen First und Second Parties. Da einzelne Mechanismen in sich geschlossen sind und als eigenständige Komponenten betrieben werden können, wurden sie als separate Services implementiert und containerisiert. Der Betrieb erfolgt auf einem Kubernetes-Cluster, was eine unabhängige Weiterentwicklung einzelner Komponenten sowie horizontale Skalierung und Lastverteilung ermöglicht. Jeder implementierte Service innerhalb der Middleware übernimmt eine klar abgegrenzte Aufgabe und übergibt die Anfrage anschließend an den nächsten Mechanismus, beispielsweise zur Prüfung einer Datenanfrage.

Die Services sind in Python umgesetzt und stellen jeweils eine schlanke RESTfull API auf Basis von Flask bereit. Diese Schnittstelle dient der Ansteuerung des jeweiligen Mechanismus und hält die Services bewusst leichtgewichtig, ohne zusätzliche Framework-Komplexität einzuführen. Um den betrieblichen Overhead gering zu halten, werden keine umfangreichen Datenbankstrukturen eingesetzt. Persistente Speicherung ist auf notwendige Funktionalitäten beschränkt, lediglich Metadatenverwaltung und Nutzungstracking verwenden einfache SQL-basierte Datenbanken. Die eigentlichen Datensätze werden hingegen nicht dauerhaft gespeichert, sondern über ein zentrales Redis-basiertes In-Memory-System transportiert. Dadurch wird sichergestellt, dass sensitive Daten nicht als dauerhafte Einträge in einer Datenbank abgelegt werden, sondern nur temporär während der jeweiligen Verarbeitungsschritte verfügbar sind und danach direkt überschrieben werden.

5 Governance-Framework

5.1 Zielsetzung des Governance-Frameworks

Die im vorherigen Kapitel ausgearbeiteten Ergebnisse lassen sich zu einem umfassenden Governance-Framework zusammenführen. Dieses Framework ermöglicht es, beteiligte Rollen und Verantwortlichkeiten den einzelnen Mechanismen systematisch zuzuordnen, deren Einbettung in Prozessketten einzuordnen und die Mechanismen im Hinblick auf Datenverfügbarkeit, Umsetzbarkeit und Wirksamkeit vergleichend zu bewerten. Damit erhalten First Parties eine Grundlage, CDR abhängig von den jeweiligen Rahmenbedingungen zielgerichtet und effizient umzusetzen. Gleichzeitig unterstützt das Framework ein realistisches Abwägen von Kosten und Nutzen bei der Implementierung der vorgestellten Mechanismen, da Integrationsaufwand und betriebliche Auswirkungen stark vom jeweiligen Systembestand abhängen. Die beschriebenen Mechanismen sind nicht als abschließender Katalog zu verstehen, sondern als Referenzrahmen und Ausgangspunkt. Sie sollen ein möglichst vollständiges Grundbild zentraler Governance-Funktionen liefern und können je nach Domäne, Regulierung und Systemkontext erweitert oder angepasst werden.

5.2 Klassifikation der beteiligten Rollen

Das Framework unterscheidet drei zentrale Rollen, die entlang der Wertschöpfung und der Datenflüsse unterschiedliche Verantwortlichkeiten übernehmen:

1. *First Party:* Dazu zählen beispielsweise App-Anbieter oder Smart-Home-Plattformen, die den primären Kontakt zu Endnutzenden halten und typischerweise als Data Controller die Verantwortung für die Datennutzung und -verarbeitung tragen.
2. *Second Party*: In der Regel ist dies die endnutzende Person, deren Daten erhoben und verarbeitet werden und die als Data Owner die Entscheidungshoheit über ihre Daten ausübt.
3. *Third Party:* Hierzu zählen externe Unternehmen, die ergänzende Services bereitstellen oder Daten auswerten und als Data Processor die Datenverarbeitung im Auftrag und nach Vorgaben der First Party übernehmen.

Die Einordnung spiegelt damit zugleich die Abfolge der Mechanismen wider: von der Bereitstellung strukturierender Informationen (First Party), über die souveräne Freigabe (Second Party), bis zu einer externen Weiterverarbeitung (Third Par-



Abb. 1 Beteiligte Rollen und deren Verantwortlichkeiten am Beispiel einer Anfrage einer Third Party zur Verarbeitung von Smart-Living-Daten (s. Schritte 1–8)

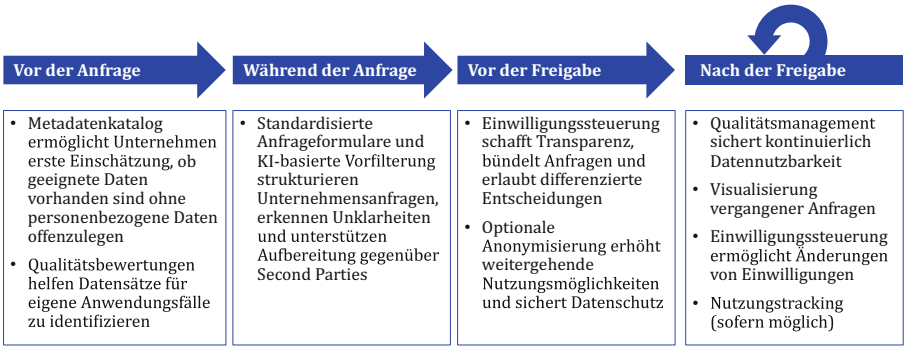


Abb. 2 Prozessuale Einordung der verschiedenen Mechanismen im Lebenszyklus einer Anfrage

ty). Diese Klassifikation bildet die Grundlage, um Zuständigkeiten im Governance-Framework klar zuzuordnen und Schnittstellen zwischen den Akteuren konsistent zu gestalten. Während die Party-Begriffe die Beziehung der Akteure zueinander beschreiben, konkretisieren die Rollen Data Controller, Data Owner und Data Processor ihre jeweilige Funktion im Umgang mit Daten, indem sie Verantwortlichkeit, Entscheidungshoheit und Verarbeitung zuordnen: Der Data Controller verantwortet die Datennutzung, der Data Owner entscheidet über die Freigabe der eigenen Daten und der Data Processor übernimmt deren Verarbeitung im Auftrag. Das Modell in Abb. 1 stellt die Zusammenhänge zwischen den Parteien bei Datenanfragen dar.

5.3 Prozessuale Einordnung der Mechanismen

Die Governance-Mechanismen greifen entlang des gesamten Datenlebenszyklus ineinander. Abb. 2 zeigt ihr prozessuale Einordnung im Lebenszyklus einer Anfrage.

Aus Unternehmenssicht entsteht durch die Einordnung ein klar strukturierter Governance-Prozess, der die Implementierung vereinfacht und modulare Bausteine liefert. KMUs können gezielt dort einsteigen, wo der erwartete Nutzen den Implementierungsaufwand rechtfertigt (z. B. beginnend mit einem Metadatenkatalog und standardisierten Anfragen, gefolgt von einem Einwilligungsportal und Nutzungstracking).

5.4 Bewertung der Mechanismen

Zur Bewertung der Mechanismen dient eine Einordnung entlang der drei Dimensionen Umsetzbarkeit, Wirksamkeit und Datenverfügbarkeit (s. Abb. 3). Während die Umsetzbarkeit den technischen und organisatorischen Implementierungsaufwand darstellt, spiegelt die Wirksamkeit den Beitrag zu Datensouveränität, Datentransparenz und Nachvollziehbarkeit und somit den Beitrag zu CDR dar. Das Potenzial zur Steigerung der Datenverfügbarkeit stellt die Erhöhung der nutzbaren Datenbasis für u. a. KI-Anwendungen dar und ist für Unternehmen somit unmittelbar marktwirtschaftlich relevant.

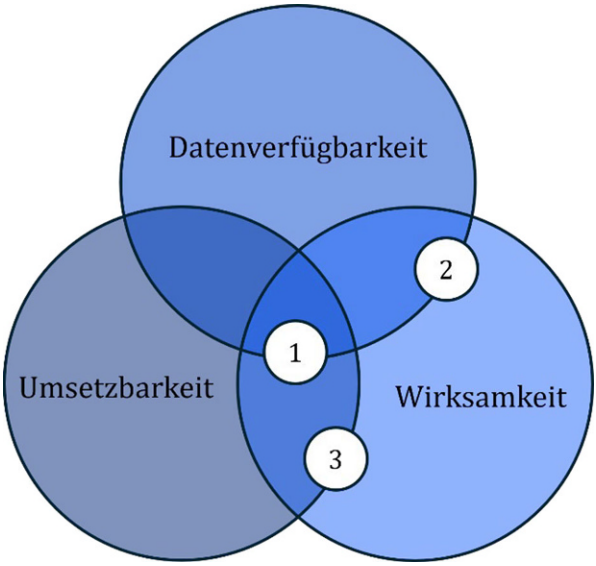


Abb. 3 Bewertungsmatrix der Governance-Mechanismen entlang von drei Dimensionen und mit drei Beispielen

Die Einordnung der Mechanismen entlang der Dimensionen ermöglicht Unternehmen eine Priorisierung und kann zugleich als Orientierung für eine schrittweise Umsetzung der Mechanismen dienen. Mechanismen wie die Umsetzung und Visualisierung standardisierter Datenanfragen, Metadatenkataloge und eine automatisierte Vorfilterung von Datenanfragen weisen in der Regel ein günstiges Verhältnis von Umsetzbarkeit und Nutzen auf (vgl. Nr. 1, Abb. 3). Granulare Einwilligungssteuerung und Nutzungstracking sind besonders wirksam für Datensouveränität, aber technisch anspruchsvoller (vgl. Nr. 2, Abb. 3). Datenqualitätsbewertung und Anonymisierung sind für den produktiven Einsatz von KI essenziell, erfordern jedoch eine stärkere Integration in bestehende Datenlandschaften (vgl. Nr. 3, Abb. 3).

Die Dimensionen dienen damit als Entscheidungsinstrument im Sinne von CDR. Unternehmen können transparent machen, welche Governance-Mechanismen sie bereits umgesetzt haben, welche geplant sind und wie diese zur fairen Datenakquise beitragen.

6 Diskussion und Fazit

Durch die vorgestellten Mechanismen wird aus CDR-Perspektive deutlich, dass Verantwortungsübernahme und unternehmerischer Nutzen kein Widerspruch sind. Metadatenkataloge, standardisierte Datenanfragen, granulare Einwilligungssteuerung und Nutzungstracking können Vertrauen bei Second Parties stärken und zugleich rechtliche Risiken reduzieren. Langfristig kann dies Wettbewerbsvorteile unterstützen, etwa durch eine glaubwürdige Positionierung und höhere Kundenbindung. Gleichzeitig erleichtert eine solche Governance den Zugang zu hochwertigen und kontextreichen Daten. Das ist insbesondere für KMUs relevant, da es dem „Henne-Ei-Problem“ datengetriebener Innovationen entgegenwirkt: Ohne Daten keine

leistungsfähigen Modelle, ohne Vertrauen und klaren Nutzen aber auch keine ausreichende Datenbasis.

Als Einschränkung des Frameworks ist zu berücksichtigen, dass die Ableitung aus Anwendungsfällen im Bereich Smart Living erfolgte. Eine Überprüfung und Weiterentwicklung in weiteren Domänen (z. B. Mobilität, Gesundheit, Energie oder öffentliche Verwaltung) ist notwendig, um Übertragbarkeit, Anpassungsbedarfe und Wirksamkeit der Mechanismen belastbarer zu bewerten. Langfristig kann das Governance-Framework dennoch dazu beitragen, faire und menschenzentrierte Datenökosysteme zu fördern, in denen Second Parties, insbesondere Privatpersonen, als souveräne Akteure auftreten und First und Third Parties verantwortungsvoll, aber effizient auf Daten zugreifen können, die für KI-Anwendungen erforderlich sind.

Eine kritische Frage jedoch bleibt: Wer stellt die Mechanismen bereit und trägt die damit verbundenen Implementierungs- und Betriebskosten? Datenökosysteme bilden einen geeigneten Rahmen für Infrastruktur und Koordination, da diese durch unterschiedliche Akteurskonstellationen bereitgestellt und organisiert werden können (vgl. Kapitel 2). Insbesondere können Akteure wie dominierende Teilnehmer oder Datenraum-Betreiber Governance-Mechanismen als geteilte Services bereitstellen, indem sie beispielsweise Metadatenkataloge, standardisierte Datenanfragen oder Qualitätsbewertungen (Dreesbach et al. 2025a) bündeln und interoperabel anbieten. Dadurch ließe sich der Integrationsaufwand für einzelne Unternehmen senken und eine breitere Beteiligung – gerade auch kleinerer Akteure – erreichen.

Das Framework impliziert für die Praxis einen konkreten Schritt in Richtung einer transformierten Data Governance: weg von rein organisationsinternen Prozessen und Regelungen, hin zu interoperablen, nachvollziehbaren Mechanismen entlang von Ökosystemgrenzen. Dabei ist ein niederschwelliger Einstieg, etwa durch priorisierte, modulare Pilotanwendungen und eine schrittweise Umsetzung der entsprechenden Governance-Ansätze, entscheidend, um Akzeptanz bei Unternehmen und Endnutzenden zu fördern und erste Implementierungshürden zu reduzieren. Wissenschaftlich eröffnet der Beitrag Ansatzpunkte, um die bislang wenig betrachtete Rolle von Endnutzenden in Datenräumen systematischer zu integrieren und CDR im Ökosystemkontext stärker zu operationalisieren.

Zusammengefasst ist transparente Datenerhebung nicht nur eine Ergänzung zu Fairness in KI-Modellen, sondern eine grundlegende Voraussetzung für verantwortungsvolle Datennutzung im Sinne von CDR. Das in diesem Beitrag vorgestellte Governance-Framework geht über reine regulatorische Mindestanforderungen der DSGVO hinaus und zeigt, wie Unternehmen Datensouveränität und Datenverfügbarkeit gemeinsam adressieren können. Es versteht transparente und souveräne Datenerhebung als integralen Bestandteil von CDR: Unternehmen übernehmen Verantwortung, indem sie Mechanismen implementieren, die Endnutzenden echte Entscheidungsfreiheit ermöglichen und Datenflüsse nachvollziehbar machen.

Funding Open Access funding enabled and organized by Projekt DEAL.

Datenverfügbarkeit Alle dieser Arbeit zugrunde liegenden Daten sind in diesem Artikel enthalten.

Einhaltung ethischer Richtlinien

Interessenkonflikt L. Dankert, T. Dreesbach, D. Maecker, A. Witzki, K. V. Carl und O. Thomas geben an, dass kein Interessenkonflikt besteht.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf http://creativecommons.org/licenses/by/4.0/deed.de.

Literatur

Cai L, Zhu Y (2015) The challenges of data quality and data quality assessment in the big data era. Data Sci J. https://doi.org/10.5334/DSJ-2015-002

Carl, K. V. (2025). Von Datenknappheit zur Datenhoheit - Endkonsument*innen als aktive Gestalter*innen für mehr Fairness in der Datensammlung. In Fairness in Zeiten Künstlicher Intelligenz (pp. 107–124). Nomos Verlagsgesellschaft mbH & Co. KG. https://doi.org/10.5771/9783748965039-107

Carl, K. V., & Hinz, O. (2024). What we already know about corporate digital responsibility in IS research: A review and conceptualization of potential CDR activities. Electronic Markets, 34(27). https://doi.org/10.1007/S12525-024-00708-0

Carl, K. V., Blachetta, M., Bilfinger, B., Brink, A., Cerulli-Harms, A. K., Christ, A.-S., Dietl, K., Dörr, S., Hauer, M. P., Hinz, M., Jänig, J.-R., Keilhofer, K., Kurtz, C., Mihale-Wilson, C., Mueller, B., Przybylek, B. A., Schaufler, N., Schlegel, I. K., Scholz, M., ... Wößner, J. (2025). Establishing Common Ground in Corporate Digital Responsibility: Integrating Understandings, Definitions, and Expert Perspectives. 1415. https://doi.org/10.18420/INF2025_125

Christ AS, Carl KV, Kurtz C, Lobschat L, Mueller B, Zimmer MP (2025) Taking digital responsibility for data: toward a governance model for user-generated data. Schmalenbach J Bus Res 77(3):527–556. https://doi.org/10.1007/S41471-025-00225-6

Dreesbach, T., Dankert, L., Brîncoveanu, C., Carl, K. V., Hinz, O., & Thomas, O. (2025a). Toward Consistent Data Quality Assessment in Heterogeneous Smart Living Systems. 27th International Conference on Business Informatics (CBI), 1–10. https://doi.org/10.1109/CBI68102.2025.00027

Dreesbach, T., Dankert, L., Ruhe, J., Carl, K. V., Hinz, O., & Thomas, O. (2025b). Data Unlocked– Ein Modell zur anwendungsübergreifenden Nutzung persönlicher Daten im Smart-Living-Kontext. HMD Praxis Der Wirtschaftsinformatik 2025 62:5, 62(5), 1156–1173. https://doi.org/10.1365/S40702-025-01198-0

Ehrlinger L, Schrott J, Melichar M, Kirchmayr N, Wöß W (2021) Data catalogs: a systematic literature review and guidelines to implementation. In: Communications in computer and information science, 1479 CCIS, S 148–158 https://doi.org/10.1007/978-3-030-87101-7_15

Geisler S, Vidal ME, Cappiello C, Lóscio BF, Gal A, Jarke M, Lenzerini M, Missier P, Otto B, Paja E, Pernici B, Rehof J (2022) Knowledge-driven data ecosystems toward data transparency. J Data Inf Qual. https://doi.org/10.1145/3467022

Jahnke N, Otto B (2023) Data catalogs in the enterprise: applications and integration. Datenbank Spektrum 23(2):89–96. https://doi.org/10.1007/S13222-023-00445-2

Koskinen J, Knaapi-Junnila S, Rantanen MM (2019) What if we had fair, people-centred data economy ecosystems? In: Proceedings—2019 IEEE smartworld, ubiquitous intelligence and computing, advanced and trusted computing, scalable computing and communications, Internet of people and smart city innovation, S 329–334 https://doi.org/10.1109/SMARTWORLD-UIC-ATC-SCALCOM-IOP-SCI.2019.00100

Lobschat L, Mueller B, Eggers F, Brandimarte L, Diefenbach S, Kroschke M, Wirtz J (2021) Corporate digital responsibility. J Bus Res 122:875–888. https://doi.org/10.1016/J.JBUSRES.2019.10.006

Mueller B (2022) Corporate digital responsibility. Bus Inf Syst Eng 64(5):689–700. https://doi.org/10.1007/S12599-022-00760-0

Munoz-Arcentales A, López-Pernas S, Pozo A, Alonso A, Salvachua J, Huecas G (2019) An architecture for providing data usage and access control in data sharing ecosystems. Procedia Comput Sci 160:590–597. https://doi.org/10.1016/J.PROCS.2019.11.042

Oliveira MIS, de Barros Lima GF, Farias Lóscio B (2019) Investigations into data ecosystems: a systematic mapping study. Knowl Inf Syst 61(2):589–630. https://doi.org/10.1007/S10115-018-1323-6

Petrik D, Untermann A, Baars H (2024) Functional requirements for enterprise data catalogs: a systematic literature review. 500 LNBIP. Lecture Notes in Business Information Processing, S 3–18 https://doi.org/10.1007/978-3-031-53227-6_1

Scheider S, Mallick MK (2025) Exploring metadata catalogs in health care data ecosystems: taxonomy development study. JMIR Form Res. https://doi.org/10.2196/63396

Scheider S, Lauf F, Möller F, Otto B, Scheider S, Lauf F, Möller F (2024) A reference system architecture with data sovereignty for human-centric data ecosystems. Bus Inf Syst Eng 65(5):577–595. https://doi.org/10.1007/S12599-023-00816-9

Strnadl CF, Schöning H (2023) Datenplattformen, Datenräume und (Daten-)Ökosysteme – Einordnung und strategische Aspekte. In: Data Governance, S 83–103 https://doi.org/10.1007/978-3-662-67556-4_2

Tallon PP (2013) Corporate governance of big data: perspectives on value, risk, and cost. Computer 46(6):32–38. https://doi.org/10.1109/MC.2013.155

Witzki, A., Carl, K. V., Dankert, L., Thomas, O., & Hinz, O. (2025). What are your intentions with my data? A tool to enhance consumer data sovereignty in digital ecosystems. Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft Fur Informatik (GI), P-364, 13–28. https://doi.org/10.18420/OID2025_01

Zrenner J, Möller FO, Jung C, Eitel A, Otto B (2019) Usage control architecture options for data sovereignty in business ecosystems. JEIM 32(3):477–495. https://doi.org/10.1108/JEIM-03-2018-0058

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.