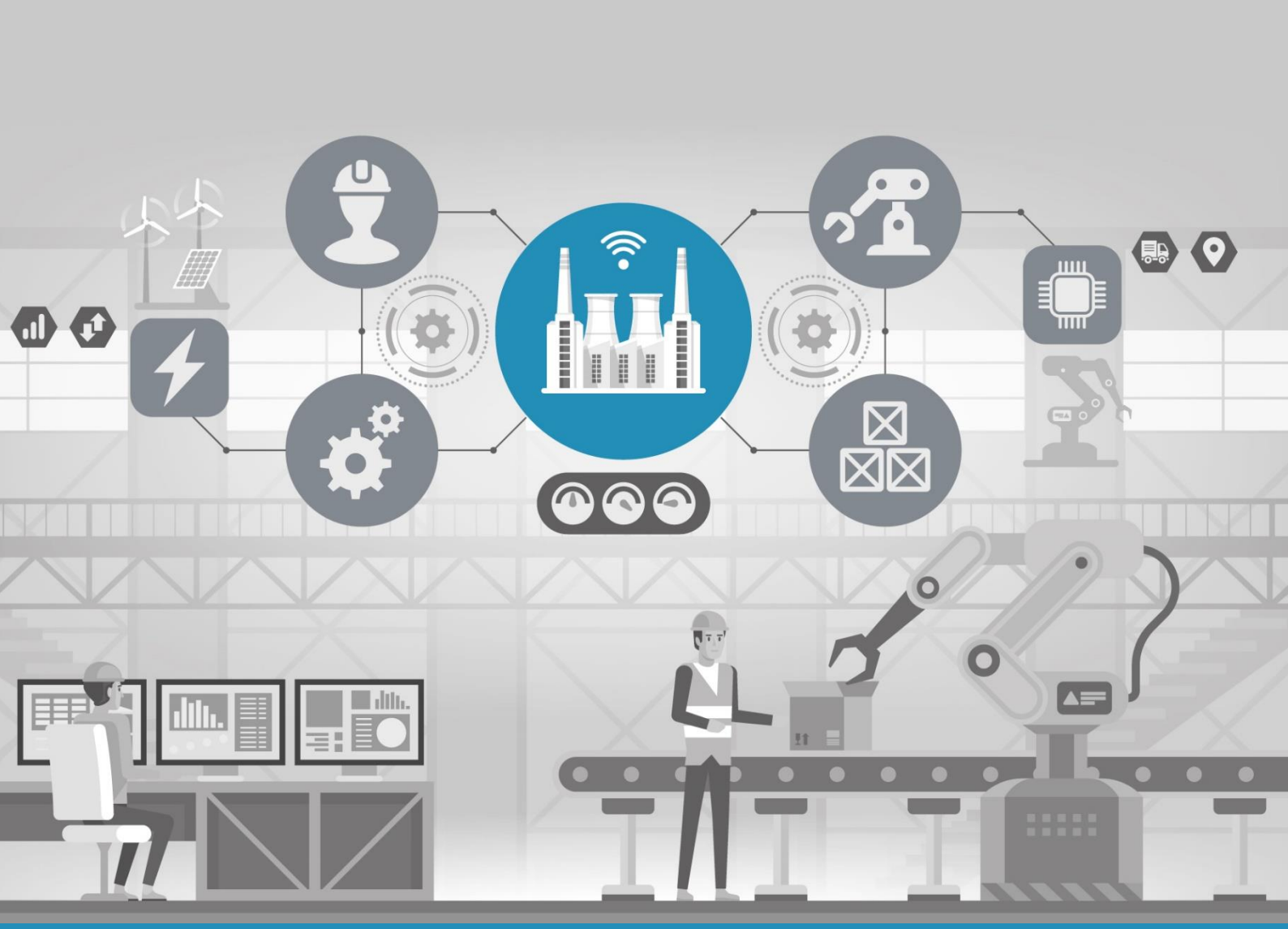


IIP-Ecosphere Whitepaper

Datenschutz und Datensicherheit in Datenökosystemen

Fabian Bruckner*, Nils Jahnke*

* *Fraunhofer-Institut für Software- und Systemtechnik ISST*

White Paper IIP-2021/003

IIP-Ecosphere
Next Level Ecosphere for Intelligent Industrial Production

Disclaimer

Die Inhalte des Dokuments wurden mit großer Sorgfalt erstellt. Obwohl die Informationen mit größtmöglicher Sorgfalt erstellt wurden, besteht kein Anspruch auf sachliche Richtigkeit, Vollständigkeit und/oder Aktualität, insbesondere kann diese Publikation nicht den besonderen Umständen des Einzelfalles Rechnung tragen.

Eine Verwendung liegt daher in der eigenen Verantwortung des Lesers. Jegliche Haftung wird ausgeschlossen. Dieses Dokument enthält Material, das dem Urheberrecht einzelner oder mehrerer IIP-Ecosphere-Konsortialparteien unterliegt. Alle Rechte, auch der auszugsweisen Vervielfältigung, liegen bei den Urhebern.

Dieses Dokument spiegelt nur die Ansicht der Autoren zum Zeitpunkt der Veröffentlichung wider. Das Bundesministerium für Wirtschaft und Energie bzw. der zuständige Projektträger haften nicht für die Verwendung der hierin enthaltenen Informationen.

Veröffentlichung: März 2021 auf https://www.iip-ecosphere.eu/

DOI: 10.5281/zenodo.4588330

Gefördert durch:

Bundesministerium
für Wirtschaft
und Energie

Executive Summary

Dieses Whitepaper stellt Probleme und Lösungsansätze im Bereich Schutz und Sicherheit von Daten in Datenökosystemen, respektive Plattformökosystemen dar. Dabei wird insbesondere der Blickwinkel des Projekts Next Level Ecosphere for Intelligent Industrial Production (IIP-Ecosphere) betrachtet, das im Rahmen des KI-Innovationswettbewerbs durch das Bundesministerium für Wirtschaft und Energie (BMWi) gefördert wird. Im Rahmen des Whitepapers werden durch eine umfassende Literaturanalyse identifizierte unmittelbare und mittelbare Probleme aus den Bereichen Datenschutz und –sicherheit in Datenökosystemen und dazugehörige Lösungsansätze präsentiert. Insbesondere im Fokus der Betrachtungen steht dabei der in IIP-Ecosphere zu gestaltende Datenmarktplatz. Die ermittelten theoretischen Erkenntnisse werden mittels einer Befragung von Partnern und Assoziierten auf ihre Praxisrelevanz geprüft und konsolidiert. Obwohl die Anzahl der Befragten keine allgemeingültigen Schlüsse zulässt, können die in diesem Whitepaper beschriebenen Befragungsergebnisse als Indikator für das Meinungsbild von Industrieunternehmen und deren Partnern hinsichtlich des immer wichtiger werdenden Phänomens der Datenökosysteme herangezogen werden.

Im Fokusbereich Datenmarktplätze und Datensharing zeigt sich, dass eine grundsätzliche Skepsis der Unternehmen gegenüber einer Teilnahme vorherrscht. Abgesehen von fehlenden Anreizmechanismen für die Teilnahme an Datenmarktplätzen ist dies insbesondere begründet in der Angst, Geschäftsgeheimnisse an die Datenempfänger weiterzugeben. Weiterhin befürchten Unternehmen einen Missbrauch der weitergegebenen Daten für nicht genehmigte Zwecke. Während erster Bereich durch die Etablierung eines umfassenden (Daten-)Risikomanagements angegangen werden kann, liefert Usage Control Lösungen für Probleme im zweiten Bereich. Die Chancen und Herausforderungen von Usage Control werden im Rahmen dieses Whitepapers erklärt und mit den Anforderungen des Konsortiums konsolidiert. Entsprechend ergeben sich wichtige Erkenntnisse für die zukünftige Weiterentwicklung von Usage Control und die Transformation von Unternehmen zur Erreichung eines Status des internen Datenmanagements, der sie zur Teilnahme an Datenökosystemen befähigt.

Weiterhin werden Teilaspekte von Datenschutz und Datensicherheit der sich in der Konzeptionierungsphase befindlichen IoT-Plattform betrachtet und wichtige Betrachtungsbereiche hervorgehoben. Dabei werden insbesondere die Themen der Architektur, Governance und das Zugriffsmanagement adressiert.

Inhaltsverzeichnis

1 Einleitung

1.1 Motivation und Ziele

Im Projekt IIP-Ecosphere entsteht ein neuartiges Ökosystem, das dabei hilft, die nächste Ebene der digitalen Produktion zu erreichen. Zu diesem Zweck werden Arbeiten durchgeführt, welche die Anwendbarkeit von Methoden künstlicher Intelligenz (KI) in der intelligenten Produktion erleichtern und in realen Anwendungsszenarien demonstrieren. Insbesondere zielen diese Aktivitäten auf eine Beseitigung der derzeit existierenden Hemmnisse und ermöglichen auch kleinen und mittelständischen Unternehmen (KMU) und Startups KI-Methoden zur intelligenten Produktion selbst zu entwickeln, zu testen und anzuwenden. Unter diese Aktivitäten fallen auch die im Projekt entstehende digitale Plattform und der Datenmarktplatz.

Abhängig von den Daten, die auf dieser Plattform und dem Datenmarktplatz verwendet werden sollen, bestehen Schutzbedürfnisse, die zwingend umgesetzt werden müssen. Diese können sich beispielsweise aus den Bereichen Datenschutz, Datensicherheit und dem Schutz des geistigen Eigentums ergeben. Zur Umsetzung der Schutzbedürfnisse sind sowohl technische als auch organisatorische Maßnahmen denkbar. Beispiele für Schutzmaßnahmen aus den verschiedenen Bereichen stellen unter anderem die Anonymisierung von Personendaten, die Verwendung von Methoden zur Datenzugriffs- und Nutzungskontrolle oder Maßnahmen des Risikomanagements dar. Beim Design und der Governance von Datenökosystemen und Datenmarkplätzen existiert dabei ein Zielkonflikt zwischen der Ausübung von Kontrolle hinsichtlich einzuhaltender Regeln oder zu nutzenden Technologien und der Offenheit gegenüber Teilnehmern und Services. Vollständige Offenheit führt dabei initial zu einer größeren Anzahl von Teilnehmern und Services, allerdings einhergehen mit geringerer Sicherheit. Technische und organisatorische Kontrollmaßnahmen führen hingegen zu einer verringerten Zugänglichkeit, aber zu besserer Servicequalität und Prozesskontrolle. Entsprechend gilt es, die gewählten Maßnahmen entsprechend der Anforderungen aus dem Umfeld anzupassen, um eine für die Zielgruppe von Unternehmen ansprechende Lösung zu entwickeln.

Ziel dieses Whitepapers stellt die Erarbeitung von Methoden und Technologien für Datenschutz und Datensicherheit in IIP-Ecosphere dar. Dazu werden mittels einer Literaturrecherche mögliche Problemstellungen bei Gewähren des Zugriffs auf und der Weitergabe von Daten in Datenökosystemen identifiziert. Darauf basierend werden unter Berücksichtigung der Erfahrung und Kompetenz im Think Tank Daten denkbare Datenschutzmaßnahmen analysiert und deren Umsetzbarkeit in IIP-Ecosphere geprüft. Zur weiteren Klärung der Umsetzbarkeit von Maßnahmen in IIP-Ecosphere, der Ermittlung von Anforderungen hinsichtlich deren Implementierung sowie zusätzlichen Präferenzen wurde eine Befragung des Konsortiums durchgeführt. Im Fokus stand dabei die Ermittlung von zu erfüllenden Schutzbedürfnissen bei den zukünftigen Nutzern der Plattform und dem Datenmarktplatz. Die Ergebnisse der Literaturanalyse und der Umfrage wurden abschließend konsolidiert, um eine gemeinsame Grundlage für die Konzeptionierung der Komponenten zu erhalten.

1.2 Interaktion mit anderen Initiativen

Der Schutz und die Sicherheit von Daten stellt einen zentralen Erfolgsfaktor von Datenökosystemen dar. Wichtige verbundene Aspekte stellen dazu einerseits die Erzeugung von Vertrauen und andererseits die Wahrung der digitalen Souveränität der einzelnen Akteure dar. Die *International Data Spaces Association (IDSA)* stellt eine Anwenderorganisation mit mehr als 120 Mitgliedern dar, die es sich zum Ziel gesetzt hat, einen internationalen Standard zum souveränen Austausch von Daten zu definieren. Die sich derzeit in einer Interimsphase befindliche Initiative *GAIA-X*, greift auf die Erkenntnisse der IDSA zurück und verfolgt darüber hinaus das Ziel, auch auf Seiten der Infrastruktur die Souveränität von Unternehmen zu gewährleisten. Dazu werden durch die GAIA-X AISBL sogenannte Federation Services, Policy Rules und eine Architecture of Standards definiert. Insbesondere auf Seiten

des Teilaspekts der Datenökosysteme werden einzelne Aspekte der IDS-Referenzarchitektur, wie die auch in diesem Whitepaper diskutierte Methode Usage Control sollen in GAIA-X Verwendung finden. Derzeit weist die Initiative GAIA-X allerdings nur einen geringen Reifegrad auf. Erste Referenzimplementierungen können frühestens mit der Bereitstellung der Federation Services erwartet werden. Weiterhin fokussieren sowohl IDSA und GAIA-X AISBL eine sektorenübergreifende Lösung. Die in diesem Whitepaper diskutierten Lösungsansätze fokussieren hingegen vielmehr den Blickwinkel des Projekts IIP-Ecosphere, dem Aufbau eines neuartigen Ökosystems für die digitale Produktion.

1.3 Struktur des Dokuments

In Anlehnung an das zuvor beschriebene Vorgehen erfolgt auch die Strukturierung dieses Dokuments, die in Abbildung 1 dargestellt ist. Zunächst erfolgt eine detaillierte Beschreibung der Problemstellung, in der durch eine Literaturanalyse einzelne Probleme im Bereich von Datenschutz und Datensicherheit bei der Gewährung des Zugriffs zu Daten oder der Datenweitergabe identifiziert und beschrieben werden. Darauf basierend werden einzelne Lösungsansätze zu Datenschutz und Datensicherheit beschrieben und aus Sicht des Think Tanks Daten in den Kontext von IIP-Ecosphere eingeordnet. Die Methodik und Ergebnisse der Befragung werden anschließen dargestellt und wichtige Gestaltungsempfehlungen aus diesen abgeleitet. Zum Abschluss des Dokuments werden die Erkenntnisse in bereichsweise eingeordnet, konsolidiert und zusammengefasst.

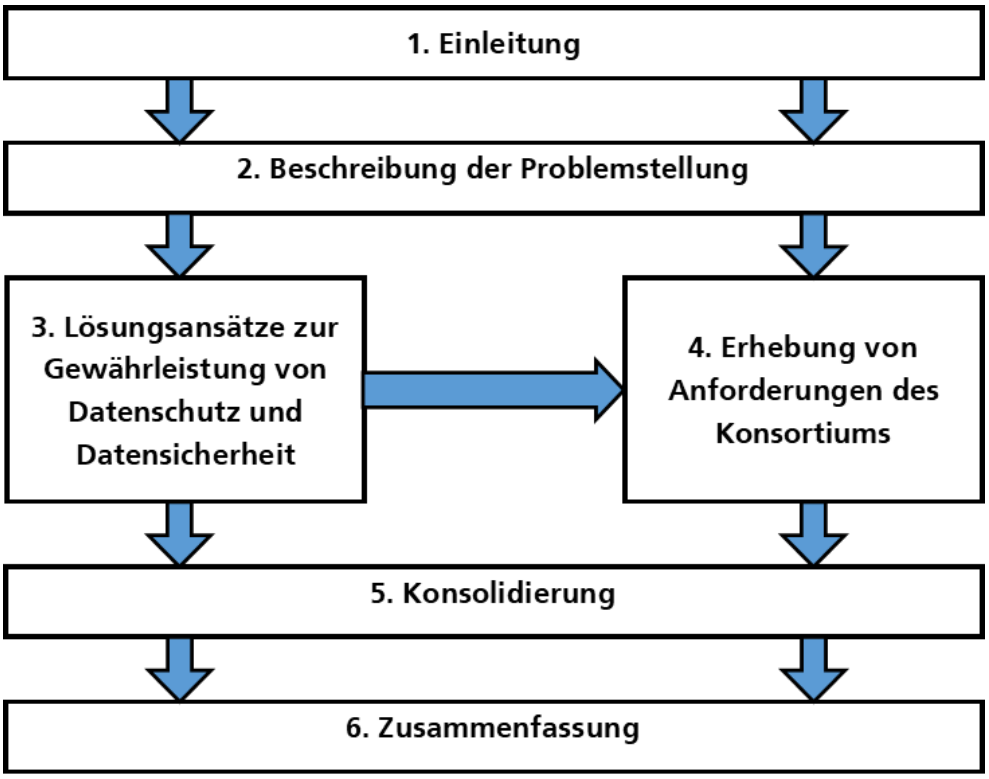


Abbildung 1: Aufbau des Whitepapers

2 Beschreibung der Problemstellung

Als Basis eines Konzepts zu Datenschutz und Datensicherheit in IIP-Ecosphere gilt es zunächst, eine Übersicht über auftretende Problemstellungen in diesem von Datenschutz und Datensicherheit in Datenökosystemen zu erlangen, sodass infolgedessen adäquate organisatorische und technische Mittel konzipiert und auf die Rahmenbedingungen von IIP-Ecosphere angepasst werden können. Mögliche Probleme wurden mittels der Betrachtung aktueller Literatur identifiziert. Dabei wurde zunächst Literatur mit explizitem Fokus auf Datenschutz und Datensicherheit von digitalen Plattformen und Datenmarktplätzen betrachtet. Weiterhin wurden allgemeine Werke zu Datenaustausch und Datenökosystemen betrachtet. Zur Ermittlung der Probleme im Bereich von Datenschutz und Datensicherheit wurde zunächst die Gesamtheit aller Herausforderungen beim Austausch von Daten ermittelt und anschließend hinsichtlich ihrer Verbindung zu Datenschutz und -sicherheit analysiert. Diese Menge an Problemen wurde abermals gefiltert, sodass ausschließlich Probleme mit Relevanz für das in IIP-Ecosphere entstehende Datenökosystem folgend diskutiert werden.

Die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) [1] liefert in ihrem, als Ergebnis eines Expertenworkshops entstandenen, Bericht „Enhancing Acces to and Sharing of Data“ eine umfassende Übersicht von Risiken und Herausforderungen im Bereich des Datenzugriffs und Datenaustausches. Die vorhandenen Herausforderungen werden dort in drei Kategorien gegliedert:

- Abwägen von Vorteilen und Risiken eines erhöhten Datenaustauschs
- Vertrauen der Teilnehmer innerhalb des Ökosystems und Befähigung der Teilnehmer zum Datenaustausch
- Anreizmechanismen zur Bereitstellung von Daten

Dabei werden innerhalb jeder Kategorie relevante Problemstellungen für Datenschutz und Datensicherheit diskutiert. Diese Problemstellungen können dabei sowohl unmittelbar, als auch mittelbar (indirekt) mit Datenschutz und Datensicherheit in Beziehung stehen. Als unmittelbares Problem werden folgend direkt auf technische oder organisatorische Rahmenbedingungen des Datensharings zurückführbare Herausforderungen diskutiert. Unter indirekten Problemen werden Rahmenbedingungen des Datensharings betrachtet, die zwar selbst kein Problem des Bereichs Datenschutz- oder Datensicherheit darstellen, allerdings etwa die Implementierung von Maßnahmen mit Datensicherheitsaspekt beeinträchtigen.

2.1 Unmittelbare Probleme im Bereich Datenschutz und Datensicherheit

Stehen Unternehmen vor der Frage, ob sie Daten mit ihren Geschäftspartnern oder sonstigen Dritten austauschen, gilt es abzuwägen, ob sich der erwartete Nutzen des Datensharings angesichts der möglichen Risiken rechtfertigen lässt. So ist u.a. bekannt, dass die Gefahr des Verrats von Geschäftsgeheimnissen Unternehmen bei dem Austausch von Daten zögern lässt oder diese sich aufgrund derer entschließen einen digitalen Service nicht zu Nutzen [1]. Im Bereich der unmittelbaren Probleme werden in diesem Abschnitt Probleme zusammengefasst, die sich direkt auf technische oder organisatorische Rahmenbedingungen des Datensharings beziehen lassen.

Ein durch die technischen Umstände hervorgerufenes Sicherheitsrisiko des Datensharings stellt die *Öffnung des Informationssystems* zum Datenzugriff für Dritte dar. Diese Öffnung setzt einer Organisation Sicherheitsrisiken aus, welche die Verfügbarkeit, Integrität und Vertraulichkeit ihrer Daten und Informationen gefährdet. Weiterhin können die Anlagen, die Reputation und die Aktivitäten der Organisation beeinträchtigt werden. Zudem ist es möglich, dass dabei, neben dem eigenen Unternehmen, weitere Unternehmen entlang der Lieferkette einer Gefahr ausgesetzt werden. Beispiele für Aktionen, die ein Unternehmen kompromittieren könnten stellen u.a. der Diebstahl von Daten durch organisierte Banden dar, welche die gestohlenen Daten wiederum auf illegalen

Marktplätzen zum Verkauf anbieten. Ebenso kann ein Abzug von Daten im Rahmen von Industriespionage erfolgen [1].

Ein besonderes Risiko eines *Datenverlustes* ergibt sich für *Personendaten* [2, 3]. Dabei resultiert nicht nur ein Schaden für die betroffene Organisation, indem ein möglicher Wettbewerbsvorteil verloren geht und das öffentliche Ansehen sinkt, sondern auf für die Individuen, die durch die Daten beschrieben werden. Obgleich die Anzahl dieser Vorfälle in der Vergangenheit relativ gering war, ergaben sich für die betroffenen Unternehmen große Schäden [1].

Neben diesen durch kriminelle Organisationen verursachten Risiken ergeben sich Risiken der *Verletzung von Datenschutz, Geschäftsgeheimnissen und anderen Interessen* auch *durch Geschäftspartner*. Die Weitergabe oder Veröffentlichung von Daten könnte bei unzureichender Governance oder technischem Schutz dem Datenempfänger erlauben, Geschäftsgeheimnisse offenzulegen und dadurch einen Wettbewerbsvorteil zu erhalten [4]. Zum anderen besteht die Gefahr, dass diese die Daten (wohlwissentlich oder versehentlich) für andere als die vereinbarten Zwecke nutzen. Die Änderung des Verwendungszwecks kann dabei unter Umständen bestehende Rechte und Pflichten untergraben. Auch aufgrund der zuvor beschriebenen Situation stellt der *Verlust von Kontrolle über Ihre Daten* für viele Organisationen ein Problem bei der Datenweitergabe dar [5]. So sind Unternehmen ohne die Bereitstellung besonderer technischer Maßnahmen nicht mehr in der Lage zu bestimmen, was mit ihren Daten geschieht. Ein Schutz durch rechtlich-organisatorische Maßnahmen reicht vielen Unternehmen nicht aus. Einige KMU haben aufgrund dieser Tatsache nicht nur auf den Austausch von Daten verzichtet, sondern auch die Nutzung von digitaler Technologien, wie etwa Cloud-Computing, abgelehnt. Ist es nicht möglich die Verwendung der Daten beim Geschäftspartner zu überwachen, so besteht zudem die Gefahr einer Identifizierung natürlicher Personen durch immer mächtigere Data Analytics-Methoden. So können vormals anonymisierte Daten durch die Verwendung von KI oder der Zusammenführung mit weiteren Datensätzen auf eine natürliche Person rückgeführt werden. Die Nutzung weiterer technischer oder rechtlicher Maßnahmen zum Schutz der Identität natürlicher Personen wird daher notwendig [1].

2.2 Indirekte Probleme mit Bezug zu Datenschutz und Datensicherheit

Als indirekte Probleme werden im Rahmen diese Whitepapers diejenigen bezeichnet, die kein unmittelbares Datensicherheits- oder Datenschutzproblem darstellen, allerdings etwa die Etablierung von Schutzmaßnahmen beeinträchtigen. Ein solches Problem repräsentiert *die Schwierigkeit einen geeigneten Risikomanagement-Ansatz* für den Datenaustausch zu etablieren. Insbesondere sind dabei zwei Faktoren hinderlich. Einerseits besitzen viele Unternehmen keinen etablierten Ansatz zum digitalen Risikomanagement innerhalb ihrer Organisation, sodass die Grundlage zum Management des Risikos in Bezug auf externe Partner fehlt. Zwar wird Risikomanagement als integraler Teil zur Entscheidungsfindung, etwa bei der Auswahl von Technologien, notwendig, jedoch fehlt den Unternehmen oftmals das notwendige Budget oder Wissen zur kontinuierlichen Umsetzung einer solchen Lösung. Andererseits existieren hohe Herausforderungen beim Management des Risikos einer Datennutzung durch Dritte. So herrscht etwa kein allgemeingültiges Verständnis über die Kategorisierung und Bewertung verschiedener Beeinträchtigungen der Privatsphäre. Weiterhin unterscheiden Unternehmen nicht zwischen Risiken für Datensicherheit und Risiken im Bereich des Datenschutzes. Werden zudem dritte Parteien in das Risikomanagement mit einbezogen, muss das Risiko in dem Maße reduziert werden, dass dieses für alle Stakeholder akzeptabel ist. Dazu müssen weitere Informationen über die Geschäftspartner in Erfahrung gebracht werden. Das Risikomanagement erfordert dazu weitere Koordination mit allen Beteiligten, was eine besonders schwierige Aufgabe darstellt [1].

Ein weiteres indirektes Problem stellt *mangelndes Vertrauen* zwischen den Teilnehmern dar. Vertrauen kann innerhalb eines Ökosystems etabliert, ausgenutzt oder abgeschwächt werden. Ein

Mangel an Vertrauen in die Mechanismen zu Datenschutz und Datensicherheit oder hinsichtlich der Aktivitäten der Geschäftspartner führt zu einer geringeren Bereitschaft zur Teilnahme an einem Plattform-Ökosystem oder am Datensharing [5]. Demgegenüber stehen Methoden zur Erhöhung des Vertrauens, wie etwa die Einbindung der Stakeholder, Etablierung gemeinsamer Standards zum Datenaustausch und zur Datennutzung oder die Sicherstellung einer hohen Qualität der im Ökosystem ausgetauschten Daten oder bereitgestellten Services, die für die Etablierung von Vertrauen sorgen [1].

Ein weiteres Hindernis bei der Etablierung von Datenschutz und Datensicherheit während des Datenaustauschs stellen entstehende *Externalitäten* dar. Unter Externalitäten werden Kosten verstanden, die nicht in den Marktpreis inkludiert sind. Auf Datenmarkplätzen stehen Datenanbieter vor der Gefahr, selbst für Maßnahmen zu Datenschutz, Datensicherheit und Datenqualität zu zahlen, ohne die Möglichkeit zu besitzen, diese einzupreisen. Aufgrund der Ungewissheit über diese Kosten könnten potentielle Datenanbieter davon abgeschreckt werden, ihre Daten anzubieten [1, 3].

Werden Daten auf einem Datenmarktplatz gehandelt, so kann es – je nach Auslegung des Datenmarktplatzes – zu *geringer Transparenz* hinsichtlich der Findung von Preisen oder dem Umgang mit Daten kommen. Während Teilnehmer auf der einen Seite durch die Services von zentralisierten Datenmarktplätzen profitieren, ergeben sich durch die geringe Transparenz schwer kalkulierbare Risiken für die teilnehmende Unternehmung. Eventuell können Datensicherheit oder der Schutz persönlicher Daten nicht gewährleistet und die Wahrscheinlichkeit von Datenlecks nicht abschließend bewertet werden. Dementsprechend stehen viele Unternehmen der Teilnahme an zentralisierten Datenmarktplätzen mit geringer Transparenz skeptisch gegenüber [1].

Zuletzt kann auch der *komplizierte Rechtsrahmen* zu Problemen im Bereich des Datenschutzes führen. Bei einer unübersichtlichen Rechtslage werden oftmals bilaterale Verträge zur Klärung der Rechte und Pflichten im Umgang mit Daten ausgehandelt. Dabei befinden sich KMU oftmals in einer schlechten Verhandlungsposition zur Klärung der Geschäftsbedingungen hinsichtlich des Zugriffs, der Nutzung und der Weitergabe von Daten. Werden die Geschäftsbedingungen durch den potentiellen Anbieter als ungerecht erachtet, agieren diese mit Zurückhaltung hinsichtlich einer Weitergabe von Daten an Dritte [1]. Weiterhin ist es für die Unternehmen schwierig, hinsichtlich der komplexen rechtlichen Situation, insbesondere bei der Übertragung von Daten in andere Rechtsräume, den Überblick zu behalten [3]. Dabei können sowohl die Momentan herrschende Rechtsunsicherheit als auch der hohe Aufwand zur Erstellung rechtlicher Gutachten hinderlich für einen Datenaustausch sein.

3 Stand der Technik - Lösungsansätze

Während im vorherigen Abschnitt aktuell herrschende Probleme im Bereich Datenschutz und Datensicherheit in Datenökosystemen und während des Daten Sharing identifiziert und beschrieben wurden, werden in diesem Abschnitt bereits existierende Lösungen und Ansätze zur Bewältigung der Problemstellung diskutiert. Die betrachteten Gegenstände werden dabei in zwei Kategorien gegliedert. Die erste Kategorie stellen Maßnahmen dar, welche durch die jeweilige Entscheidungsinstanz, die das Datenökosystem und den Datenmarktplatz entwirft oder verwaltet, definiert werden. Die zweite Kategorie behandelt Maßnahmen, die ein jedes Unternehmen intern und unabhängig von den zentral getroffenen Entscheidungen durchführen kann, um Probleme in den Bereichen von Datenschutz und Datensicherheit während der Interaktion in Datenökosystemen zu reduzieren.

3.1 Zentrale Gestaltungselemente

In diesem Abschnitt werden Elemente des Datenschutzes und der Datensicherheit diskutiert, die von der zentralen Gestaltunginstanz entworfen und implementiert werden können. Bei den traditionellen Plattform-Ökosystemen handelt es sich bei der zentralen Gestaltungsinstanz oftmals um ein einzelnes Unternehmen. So legen etwa Apple und Google allein fest, welche Services auf der Plattform zugelassen werden und wie die Einnahmen aus diesen Services verteilt werden. In IIP-Ecosphere liegt kein zentrales Unternehmen vor, dass solche Entscheidungen allein treffen kann. Stattdessen werden in den Think Tanks, unter Berücksichtigung der Anforderungen der Konsortialteilnehmer, die zentralen Mechanismen der Plattform gestaltet. Die in diesem Kapitel erarbeiteten Konzepte stellen die Basis für die Befragung des Konsortiums dar, in welcher die Präferenzen hinsichtlich einzelner Auslegungsaspekte abgefragt werden.

3.1.1 Konzeptionierung der Datenhaltung

Im Rahmen des Aufbaus eines Datenökosystems stellt die Art der Datenhaltung einen wichtigen Aspekt dar. Zur Kennzeichnung der Datenhaltungsarchitektur in Datenökosystemen wird vornehmlich eine Einteilung in zwei rivalisierende Konzepte unternommen. Einerseits existieren zentrale Lösungen, bei denen die Daten durch die verschiedenen Datenanbieter in einem zentralen Punkt angeboten werden. Andererseits ist ebenso eine dezentrale Lösung möglich, bei welcher die Daten zunächst im Besitz der Datenproduzenten verbleiben [6]. Im Folgenden werden die genannten Konzepte genauer betrachtet und diese im Hinblick auf ihre Eignung zur Nutzung im Rahmen von IIP-Ecosphere gegenübergestellt.

Der *zentralisierte Ansatz* besteht aus einer allumfassenden Datenplattform im Kern der Architektur, die durch den Plattformbetreiber bereitgestellt wird. Die Datenanbieter stellen ihre Daten und dazugehörige Metadaten über die vom Plattformbetreiber angebotenen Schnittstellen zur Verfügung, wobei die Daten damit auch physisch auf die Datenplattform ausgelagert werden. Dort können weitere Teilnehmer, wie etwa Datenmanager oder Datenanalysten, Dienstleistungen zur Verfügung stellen. Der Plattformanbieter stellt neben der Infrastruktur ebenso zusätzliche Vermittlungsservices, wie etwa Such- und Bezahlfunktionen und den Transaktionsprozess zur Verfügung [7]. Die prominentesten Anbieter solcher Datenhaltungslösungen stellen die sogenannten „Hyperscaler" Amazon, Google und Microsoft dar. Bekanntermaßen besitzen bei diesen Modellen einzelne Anbieter und Kunden nur eine geringe Möglichkeit die Rahmenbedingungen anzupassen. Stattdessen müssen sie sich nach den vorgegebenen Bedingungen richten. Dass nicht alle Plattformteilnehmer mit dieser Kräfteverteilung einverstanden sind, äußert sich etwa in dem aktuell stattfindenden Machtkampf zwischen Apple, Google und Epic Games[1].

[1] https://www.nytimes.com/2020/08/25/technology/fortnite-creator-tim-sweeney-apple-google.html

Im Rahmen des *dezentralen* Ansatzes wird vollständig auf eine intermediäre Plattform zum Austausch der Daten verzichtet und der Datentransfer erfolgt direkt von Anbieter zu Kunde. Die zentrale Plattform wird durch grundlegende Kommunikationsstrukturen zur Allokation von Angebot und Nachfrage ersetzt. Die notwendigen Informationen über die Verfügbarkeit von Daten, deren Menge und Wert sowie die Vertrauenswürdigkeit der einzelnen Teilnehmer liegen nicht mehr bei einer zentralen Instanz, sondern stehen verteilt über die einzelnen Teilnehmer zur Verfügung. Auch die für den Datenaustausch notwendigen Schnittstellen sind nicht explizit vorgegeben, sodass die Teilnehmer die Bedingungen des Handels untereinander festlegen [7]. Beispielhafte Realisierungen solch einer Architektur wurden bereits über Blockchain-Infrastrukturen getätigt. Besonders für den Bereich von IoT-Anwendungen stellten sich dabei jedoch Probleme hinsichtlich der unterstützten Latenzzeiten und der hohen Datenmenge heraus, sodass solche Architekturen in der industriellen Praxis bisher nur eine geringe Relevanz besitzen.

Während beide Konzepte im Vergleich zu bilateralen Lösungen für eine Reduktion der Transaktionskosten sorgen [7], ergeben sich bei deren Gegenüberstellung eine Reihe an Vor- und Nachteilen. Die Nutzung einer zentralen Plattform sorgt für eine Bündelung und Vereinheitlichung sämtlicher Aktivitäten, indem sie durch die zentrale Suche eine hohe Auffindbarkeit von Services gewährleistet, die Datengüter über einheitliche Schnittstellen bereitstellt und alle genutzten Formate und Distributionskanäle spezifiziert. Im Gegensatz dazu gestaltet sich die Datenfindung bei Nutzung des dezentralen Ansatzes deutlich komplexer. Weiterhin besteht dort eine hohe Varianz an Schnittstellen, Datenformaten und Preismodellen [8]. Neben der einfachen Nutzbarkeit des Datenaustausches stellen die Vertrauenswürdigkeit und Selbstbestimmtheit der Datennutzer auf einer solchen Plattform einen wichtigen Aspekt für dessen Akzeptanz dar. Aufgrund der erhöhten Notwendigkeit für die Unternehmen den Partnern auch sensible Daten zur Verfügung zu stellen, ist für die Akzeptanz der Datenaustauscharchitektur entscheidend, dass die Anbieter über die Nutzer ihres Angebots selbst, sowie den Verwendungszweck und die Bedingungen der Servicenutzung Kontrolle besitzen [9]. Dies wird bei Verwendung des dezentralen Architekturkonzepts durch den Verbleib der Daten bei den Anbietern gewährleistet [8]. Bei Nutzung einer zentralen Plattform tritt der Datenerzeuger die Entscheidungsgewalt jedoch an den Plattformbetreiber ab und verliert die Möglichkeit Entscheidungen über die Nutzung des Datengutes zu treffen [7].

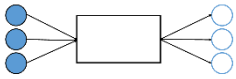
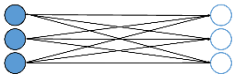
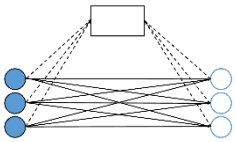
Projiziert man diesen Sachverhalt auf die Anforderungen in IIP-Ecosphere, so zeigt sich, dass hinsichtlich Benutzerfreundlichkeit und Implementierbarkeit die Nutzung einer zentralisierten Datenarchitektur sinnvoll wäre. Studien zeigen jedoch, dass Unternehmen nur an Datenaustauschen teilnehmen, wenn diese sichergehen können, dass Ihre Daten sicher vor Missbrauch geschützt werden, die einzelnen Teilnehmer sich untereinander Vertrauen und diese autonom bestimmen können, was mit ihren Daten geschieht [10, 11]. Für das zentralisierte Datenhaltungskonzept bedeutet dies folglich, dass mit hoher Wahrscheinlichkeit nur eine geringere Zahl an Unternehmen zum Datenaustausch animiert werden kann. Folglich sollte die Umsetzung einer dezentralen Plattform bevorzugt werden. Dort zeigt sich allerdings das Problem einer beschränkten Nutzbarkeit durch aufwändiges Suchen und einem geringen Standardisierungsgrad, sodass ein hoher Aufwand für die Abwicklung einzelner Transaktionen besteht. Gleichermaßen zeigt sich die Problematik von hohen Latenzen und geringen Datenmengen bei bisherigen Umsetzungen dezentraler Architekturen auf Basis von Blockchain. Somit wäre eine dezentrale, auf Blockchain-Technologie basierende Architektur für IoT-Anwendungen, wie etwa Predictive Maintenance, nur bedingt brauchbar.

Insgesamt zeigt sich somit, dass weder vollständig zentrale noch vollständig dezentrale Architekturen den Erfordernissen eines interorganisationalen Datenaustausches in Business-Ökosystemen gerecht werden. Insbesondere im Hinblick auf die Implementierung im Projekt IIP-Ecosphere wird deutlich, dass aufgrund des Trade-offs zwischen Benutzerfreundlichkeit und Währung der digitalen Souveränität

sowie den Problemen bei der technischen Realisierung keine „as-is" Implementierung eines der Ansätze stattfinden kann.

Eine gangbare Alternative ist in diesem Zusammenhang hingegen die Implementierung einer hybriden Architektur. Diese kombiniert Teilaspekte beider Ansätze, indem sie einerseits einen dezentralen Datenhandel ermöglicht, andrerseits jedoch eine zusätzliche Instanz in Anlehnung an den zentralen Ansatz bereitstellt [6]. Zweck der zentralen Instanz ist es dabei, die Schwächen dezentraler Infrastrukturen im Bereich der Vermittlung von Angebot und Nachfrage und der Standardisierung auszugleichen, ohne aber die digitale Souveränität der Datenanbieter zu gefährden. Eine vielversprechende Lösung des Allokationsproblems mittels Intermediär stellt die Implementierung eines Daten-Shops dar. Der Daten-Shop speichert und verwaltet Informationen zu den verfügbaren Datenquellen und stellt diese den Datennutzern zur Verfügung. Zur Ausführung dieser Funktion übersenden die Datenanbieter Metadaten an den Daten-Shop. Die Rohdaten verbleiben hingegen beim Anbieter. Der Daten-Shop sichert die Daten an einem internen Speicherort und erlaubt den Datennutzern Anfragen zu benötigten Daten zu senden. Zusätzlich können durch diese Komponente neben Datenart- und Ort weitere Informationen wie etwa die Datenqualität, Datengüte, der Nutzungszeitraum und die Nutzungsbedingungen transparent dargestellt werden. Der Daten-Shop übernimmt damit die Rolle eines Brokers gemäß der Definition in [12]. Tabelle 1 zeigt zusammenfassend die Gegenüberstellung der in diesem Abschnitt diskutierten Datenhaltungsmodelle.

Tabelle 1: Gegenüberstellung der Datenhaltungsprinzipien:

			
Art	**Zentral**	**Dezentral**	**Hybrid**
Vorteile	*Hohe Auffindbarkeit des Angebots durch zentrale Suche [8]* *Einheitliche Schnittstellen zur Bereitstellung von Daten oder Services [8]* *Festlegung der Austauschformate und Übertragungswege [8]*	*Anbieter behalten Kontrolle über Daten oder Services und Rahmenbedingungen von deren Nutzung [8]*	*Datenanbieter behalten Kontrolle über Daten und können Datennutzer selber autorisieren.* *Allokation von Angebot und Nachfrage wird durch zentrale Instanz ermöglicht.*
Nachteile	*Bevorzugung großer Anbieter durch Skalenvorteile [7]* *Wenig Mitbestimmung durch Plattformteilnehmer [7]* *Plattformteilnehmer verlieren Kontrolle über ihre Daten/ Services [7]*	*Hohe Anzahl an Schnittstellen, Austauschformaten und Preismodellen [8]* *Schwierige Auffindbarkeit des Angebots [8]*	*Zusätzliche Instanz erfordert Aufwand zur Implementierung, Management und Governance*

Ein erfolgreiches Beispiel für die Implementierung eines solchen hybriden Konzepts auf Basis der International Data Spaces (IDS) stellt der Advaneo Global Data Marketplace dar [9]. Auch dort werden ausschließlich Metadaten zur Beschreibung des Angebots ausgetauscht und die Rohdaten verbleiben bei den Anbietern, um digitale Souveränität zu gewährleisten. Der Austausch der Rohdaten findet nach dem Kauf auf direktem Wege zwischen Anbieter und Nachfrager statt. Den Nutzern wird ermöglicht neue datengetriebene Geschäftsmodelle aufzubauen, Kooperationen einzugehen und Innovationen zu entwickeln.

3.1.2 Etablierung von Usage Control

Nachdem die Souveränität und Selbstbestimmtheit der Teilnehmer als ein wichtiges Entscheidungskriterium bei der Auslegung des Datenhaltungskonzepts identifiziert wurden, werden im Folgenden unter dem Oberbegriff Usage Control weitere Maßnahmen diskutiert, die eine Einhaltung ebendieser Prinzipien in digitalen Ökosystemen technisch sicherstellen. Usage Control beschreibt die Möglichkeiten von Dateneigentümern über die Verwendung ihrer Daten durch Dritte zu bestimmen, indem Richtlinien zur Datennutzung entworfen, kontrolliert und durchgesetzt werden [12]. Während Access Control beschreibt, welche Teilnehmer die Daten oder Services abrufen können, legt Usage Control darüber hinaus fest, was mit den Daten nach deren Abruf durch den Datennutzer erlaubt ist [13].

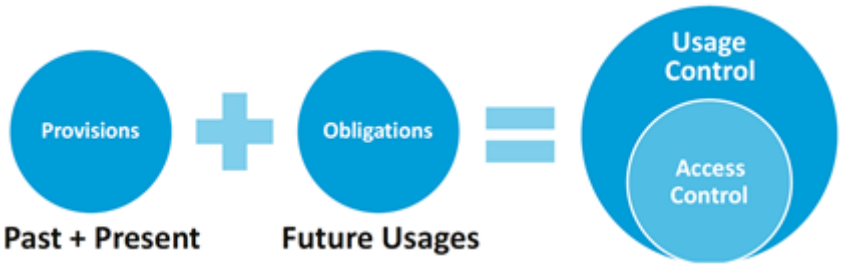


Abbildung 2: Usage Control als Erweiterung von Access Control [13]

Zur Gewährleistung der Datensouveränität innerhalb des im Rahmen von IIP-Ecosphere entstehenden Datenökosystems können Konzepte aus dem Bereich Usage Control Anwendung finden. Innerhalb dieses Abschnitts werden einzelne Konzepte aus dem Bereich Usage Control erläutert und im Anschluss im Hinblick auf die Einsatzmöglichkeiten im Projekt IIP-Ecosphere miteinander verglichen. Ebenso wird auf die notwendigen Rahmenbedingungen für einen Einsatz von Usage Control eingegangen.

Usage Control ist eine Erweiterung des klassischen Acces Control Modells. Es stehen verschiedene solcher Access Controll Modelle zur Verfügung. Am meisten genutzt werden jedoch die Modelle Role-based Access Control (RBAC) und Attribute-based Access Control (ABAC). Es ist wichtig zu betonen, dass Usage Control nur in einer bereits vertrauenswürdigen Umgebung einsetzbar ist. Denn nur dort ist garantiert, dass Usage Control tatsächlich die definierten Regeln durchsetzen kann. Usage Control alleine kann hingegen kein Vertrauen in einem System erzeugen. Ein Beispiel für solch eine vertrauenswürdige Umgebung stellt der Trusted Connector der IDS dar [14]. Zur einheitlichen Definition von Verpflichtungen, Verboten und Genehmigungen und Konsequenzen bei Verletzungen der Regeln im Umgang mit Daten sollte eine Policy Specification Language (PSL) für das gesamte Ökosystem festgelegt werden. Auch im Rahmen von IIP-Ecosphere ist es dementsprechend notwendig, diese Rahmenbedingungen zu schaffen, um ein Usage Control-Konzept zu etablieren, dass die digitale Souveränität der Nutzer gewährleisten kann.

Betrachtet man herkömmliche Lösungen aus dem Bereich Usage Control, so zeigt sich, dass diese aus einer Vielzahl von Gründen nicht in der Lage sind die benötigte Datensouveränität der Teilnehmer in Geschäftsökosystemen zu gewährleisten. So werden diese nicht im Zusammenhang mit einer

vertrauenswürdigen Umgebung eingesetzt. Weiterhin können viele verfügbare Usage Control Lösungen nicht bei einer Verletzung von Usage Policies korrigierend eingreifen und bieten lediglich eine Überwachungsfunktion. Für Nutzer bedeutet dies, dass etwa ein Datenabzug nicht verhindert werden kann, sondern nur erfasst wird. Zudem sind gegebene Lösungen oftmals nicht in der Lage einen Datenaustauschprozess zwischen mehreren Akteuren zu überwachen, sondern beschränken sich lediglich auf eins-zu-eins-Beziehungen. Schlussendlich sind viele Modelle ausschließlich für vereinzelte Bereiche, wie etwa Cloud- oder IoT-Anwendungen nutzbar und nur selten können diese Modelle generisch verwendet werden [15]. Basierend auf diesen Erkenntnissen wird ersichtlich, dass, u.a. auch für IIP-Ecosphere, sofern Usage Control Objekte verwendet werden sollen, ein Usage Control-Modell notwendig ist, das für verschiedene Bereiche, mehrere Teilhaber und in Berücksichtigung einer vertrauten Umgebung die digitale Souveränität der Akteure sichert.

Selbige Problemstellung wurde während der Entwicklung der International Data Spaces behandelt [13]. Das Ziel der IDS ist es, Akteuren zu ermöglichen, ihre eigene digitale Souveränität zu wahren. Dabei wird unter dem Begriff der digitalen Souveränität die Fähigkeit verstanden, selbstbestimmt zu entscheiden durch wen und in welchem Maße die eigenen Daten verarbeitet werden [16]. Dies ist eine zentrale Fähigkeit, welche zum erfolgreichen agieren in Daten-zentrierten Geschäftsfeldern notwendig ist.

Im Rahmen der IDS wird dabei nicht vorgeschrieben, welche Lösungen für Usage Control verwendet werden sollen. Dennoch gibt es verschiedene Komponenten, welche Usage Control Mechanismen umsetzten und im Rahmen der IDS entstanden oder für diese angepasst wurden. Diese Lösungen werden nachfolgend betrachtet.

Das übergeordnete Ziel der Umsetzung von Usage Control in den IDS besteht in einer Reduktion des organisatorischen Aufwandes, der anfällt, wenn mehrere Parteien Daten miteinander tauschen wollen. Durch ausgefeiltere und ausdrucksstärkere technische Umsetzungen von Datennutzungskontrolle soll es ermöglicht werden, an diversen Stellen auf eine organisatorische und rechtliche Umsetzung zu verzichten. Abbildung 3 zeigt, wie die IDS im Laufe der Zeit die organisatorische und rechtliche Durchsetzung von Datennutzungskontrolle anteilig durch Verfahren zur technischen Durchsetzung ablösen wollen.

Abbildung 3: Organisatorische vs. rechtliche Durchsetzung von Usage Control gemäß [13]

Connectoren:

Bevor Usage Control Lösungen in den IDS betrachtet werden, ist es für das Verständnis hilfreich, wenn zunächst kurz auf den Aufbau der IDS eingegangen wird. Für das vorliegende Dokument ist dabei eine Komponente der IDS von besonderem Interesse: Der Connector. Beim Connector handelt es sich um die Komponente von der jeder Teilnehmer (mindestens) eine Instanz betreiben muss, um mit anderen Teilnehmern der IDS zu kommunizieren. Der Connector stellt das zentrale Gateway zur Teilnahme am Ökosystem und demnach zum Austausch der Daten dar. Der Aufbau und die Struktur eines Connectors hängt stark von seinem Einsatzgebiet und der notwendigen Funktionalität ab. Beispielsweise hat ein Connector für den Einsatz in eingebetteten Systemen einen geringeren Funktionsumfang als einer, der auf dedizierten Servern betrieben wird.

Auch wenn daraus folgt, dass die Architektur unterschiedlicher Connector-Implementationen sich erheblich voneinander unterscheiden können, beschreibt das IDSA Referenzarchitekturmodell (RAM) in der Version 3.0 [13] eine Architektur für Connectoren, welche in der nachfolgenden Abbildung zu sehen ist. Diese ist nicht verpflichtend und auch nicht in allen Szenarien bzw. Implementationen umsetzbar, dient aber als Grundlage zur Umsetzung eigener Connectoren.

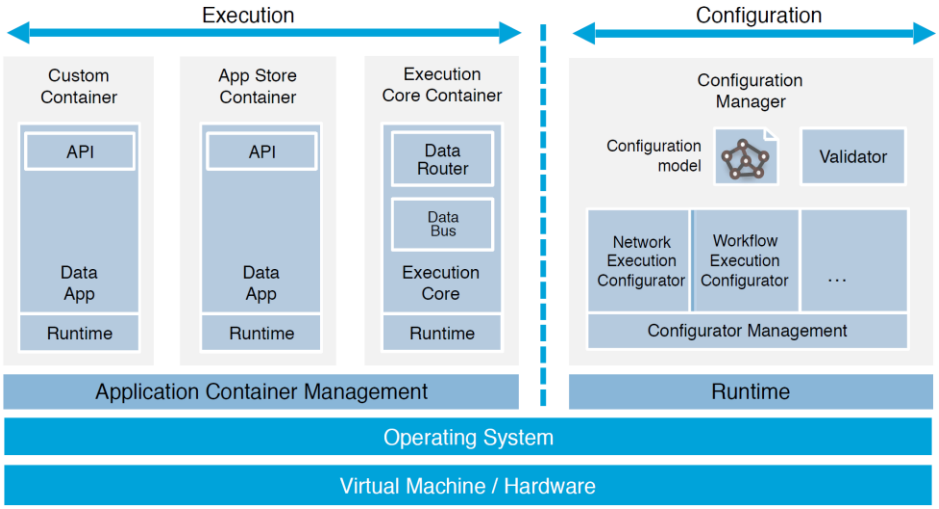


Abbildung 4: Architektur eines IDS Connector gemäß [13]

Um die unterschiedlichen Services, die in einem Connector ausgeführt werden, voneinander und von der ausführenden Hardware zu isolieren, wird auf eine Containervirtualisierung (beispielsweise Docker) gesetzt. Der Execution Core Container ist ein zentraler Service, welcher die Kommunikation mit anderen Connectoren befähigt und gleichzeitig die Weiterleitung von Nachrichten innerhalb des Connectors erlaubt. Die anderen Container enthalten sog. Data Apps, welche in den IDS Applikationen darstellen, die Daten verarbeiten (bspw. aggregieren).

IDS Informationsmodell:

Während das Konzept der Connectoren in den IDS sicherstellt, dass die Teilnehmer über eine technische Basis verfügen, welche für die Kommunikation und den Datenaustausch verwendet werden kann, ist es notwendig, sicherzustellen, dass die einzelnen Parteien ein gemeinsames Format verwenden. Nur so ist gewährleistet, dass die unterschiedlichen Teilnehmer der IDS ohne vorherigen Aufwand Nachrichten miteinander austauschen können und diese auch verstehen. Zu diesem Zweck wird das IDS Informationsmodell verwendet.

Beim IDS Informationsmodell handelt es sich um ein Metadatenmodell, welches verwendet wird um alle relevanten Elemente in den IDS zu beschreiben. Dies umfasst unter anderem:

- Teilnehmer
- Infrastrukturkomponenten
- Datenangebote
- Nachrichtentypen
- Nutzungsbedingungen

Das IDS Informationsmodell verwendet als Format Resource Description Framework (RDF). Es ist ein Open Source Projekt und die aktuellen Entwicklungen am IDS Informationsmodell können im entsprechenden github-Repository eingesehen werden[2].

[2] IDS Informationmodel https://github.com/International-Data-Spaces-Association/InformationModel

Policies:

Wenn Lösungen für Usage Control diskutiert werden, sind Policies ein untrennbarer Teil davon. Im Kontext der vorgestellten Usage Control Lösungen handelt es sich bei Policies um technische Repräsentationen der Nutzungsbedingungen, die bei der Verwendung von Daten eingehalten werden müssen. Bedingt durch unterschiedliche Einflüsse und Ursprünge der einzelnen Lösungen, verwenden die Lösungen häufig inkompatible Policysprachen.

Dabei entscheidet die verwendete Policysprache durch ihre Mächtigkeit über die Arten von Policies, die mit ihr ausgedrückt werden können. Diese hat damit auch direkten Einfluss auf die Menge an Policies, welche von einer jeweiligen Lösung umgesetzt werden kann. Dabei bedeutet es nicht zwangsläufig, dass eine Policy umgesetzt werden kann, nur, weil sie in der jeweiligen Sprache beschreibbar ist. Der Umkehrschluss dagegen ist gültig. Eine Policy, welche in einer Sprache nicht ausgedrückt werden kann, kann von der Lösung, welche die Sprache verwendet, nicht (oder nur über Umwege und Hilfskonstrukte) umgesetzt werden.

Im Rahmen der IDS ist es notwendig, dass eine gemeinsame Policysprache zum Informations- und Datenaustausch verwendet wird. Dabei muss es sich um eine möglichst mächtige Sprache handeln, die alle geforderten Policies abbilden kann. Dabei ist es weder sinnvoll noch realistisch, dass die bereits existierenden Lösungen für Usage Control ihre bisher verwendeten Sprachen verwerfen. Aus diesem Grund ist für die Verwendung der konkreten Usage Control Lösungen eine Transformation von der allgemeinen Sprache in die spezifische, von der Lösung verwendeten Sprache notwendig.

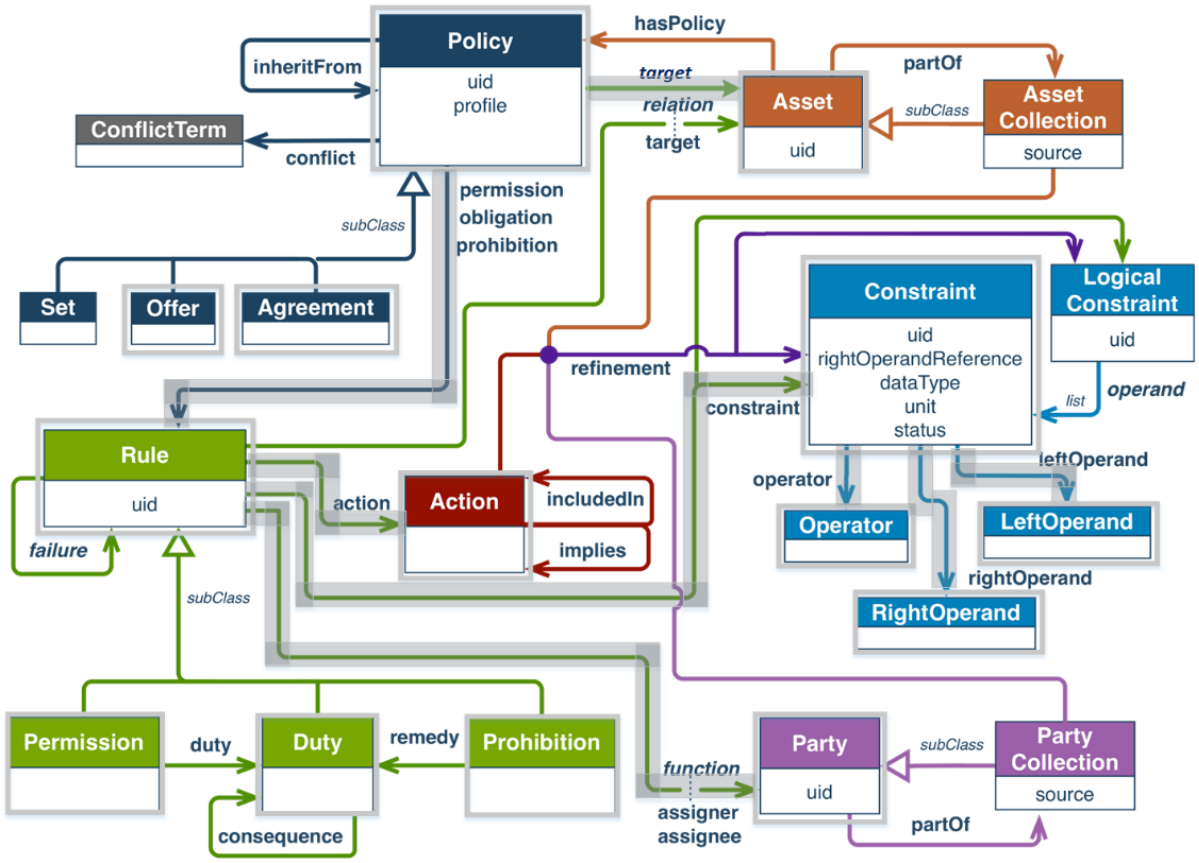


Abbildung 5: In den IDS verwendete Elemente des ORDL-Informationsmodells

In den IDS wurde sich auf die Verwendung von einer Sprache verständigt, welche an Open Digital Rights Language (ODRL) angelehnt ist und die notwendige Ausdrucksstärke besitzt. Bedingt durch die Mächtigkeit der verwendeten Sprache ist eine automatische Transformation in eine der spezifischen

Sprachen in der Regel nicht möglich. Aus diesem Grund wurden 18 Policy-Muster verabschiedet, welche in den IDS verwendet werden. Dabei muss sich jede Lösung dahingehend positionieren, welche dieser Muster sie unterstützt. Es ist vorgesehen, die Menge an Mustern in der Zukunft nach Bedarf zu erweitern.

Abbildung 5 zeigt das ODRL-Informationsmodell, welches zur Definition von ODRL-Policies verwendet wird. Grau umrandet sind die Elemente, welche in der IDS-Policysprache verwendet werden können und entsprechend im IDS Informationsmodell abgebildet sind.

Systemgrenzen:

Sollen haltbare Aussagen über Nutzungsbedingungen und deren Durchsetzung gemacht werden, ist dies nur möglich, wenn auch der Geltungsbereich dieser Aussagen definiert ist. Die Policies, welche im Rahmen der IDS verwendet werden, gelten ausschließlich in den Systemen der IDS. Wird beispielsweise an einen Connector ein externes ERP-System angeschlossen und es fließen Daten in dieses System ab, so ist es nicht möglich Policies für diese Daten durchzusetzen, da sie den Einflussbereich des Connectors und der IDS verlassen.

Dies ist keine Besonderheit des IDS, sondern eine direkte Folge der Tatsache, dass in solchen Situationen Daten den Connector verlassen und in ein System übergehen, welches für den jeweiligen Connector eine Blackbox darstellt. Dieser Punkt ist von hoher Wichtigkeit und muss beachtet werden, um den Schutz von Daten gewährleisten zu können. In der Folge werden einzelne Usage Control Lösungen der IDS erklärt

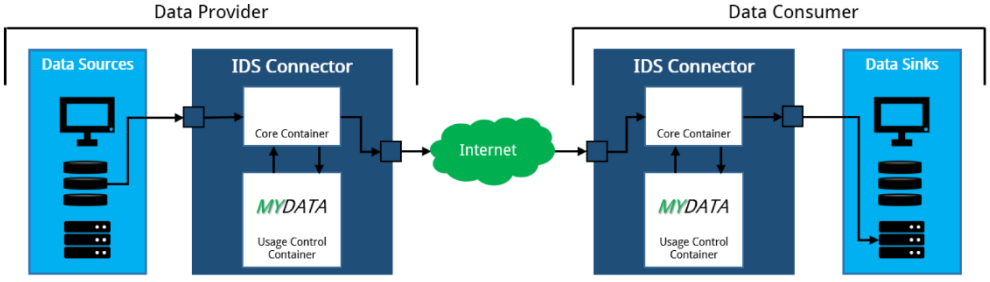
3.1.2.1 MYDATA

Bei MYDATA handelt es sich um eine Usage Control Lösung, welche auf dem Framework IND²UCE basiert, welches vom Fraunhofer IESE entwickelt wurde [14].

Bei der eXtensible Access Control Markup Language (XACML) handelt es sich um eine XML-basierte Policysprache, welche auch eine passende Architektur für die Durchsetzung der Policies enthält[3]. Dabei beschreibt XACML verschiedene zentrale Komponenten, welche in ihrer Gesamtheit dazu verwendet werden können Policies in einem System zu definieren, zu speichern und durchzusetzen. In MYDATA wird diese Architektur um zusätzliche Komponenten erweitert, welche den Funktionsumfang von XACML noch zusätzlich um erweiterte Management- und Ausführungsfunktionalitäten erweitern.

Verwendung:

Zur Funktionserfüllt stellt MYDATA einen sogenannten Usage Control Container (UCC) zur Verfügung, welcher in IDS Connectoren eingebunden werden kann und anschließend vom Core Container angesprochen wird, um Usage Control Anfragen zu beantworten. Abbildung 6 zeigt eine entsprechende schematische Darstellung.



[3] extensible access control markup language (xacml) version 3.0. 2011-09-24. http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf

Abbildung 6: Kommunikationsfluss und Integrationskonzept in MYDATA gemäß [14]

Auch wenn in der vorherigen Abbildung die Integration des UCC in die Connectoren von Konsument und Anbieter identisch dargestellt ist, werden hierbei verschiedene Verfahren verwendet. Dies ist bedingt durch die Tatsache, dass zwei Verschiedene Möglichkeiten zur Integration von Usage Control mittels MYDATA in Connectoren existieren.. Dabei wird zwischen Datenanbietern und Datenkonsumenten unterschieden. Dies kann der nachfolgenden Abbildung 7 entnommen werden und wird in den folgenden Abschnitten genauer beschrieben.

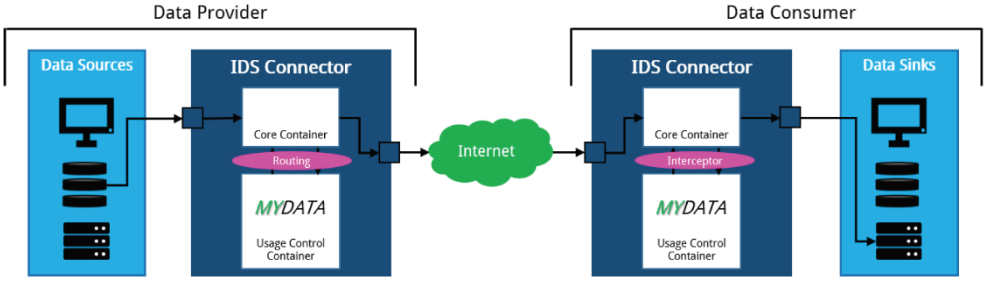


Abbildung 7: Ansatz Routing und Interceptor in MYDATA gemäß [14]

Usage Control Container Integration für Datenanbieter:

Datenanbieter haben ein eigenes Interesse daran, dass ihre Daten durch den UCC bearbeitet werden, um die Einhaltung der definierten Nutzungsbedingungen sicherstellen zu können. Aus diesem Grund liegt die Integration des UCC auf Anbieterseite in der Verantwortung des Datenanbieters selber. Er muss sicherstellen, dass Daten, welche den Connector verlassen unmittelbar vor dem Versand durch den UCC Container bearbeitet werden. Dabei kann die Integration beispielsweise durch die Verwendung von Apache Camel[4] erfolgen. Apache Camel ist eine Nachrichtenbasierte Middleware, welche häufig in Connectoren als Message Router verwendet wird.

Usage Control Container Integration für Datenkonsumenten:

Auf der Seite des Datenkonsumenten muss sichergestellt werden, dass der UCC alle eingehenden Daten erhält, um entsprechende Usage Control Entscheidungen fällen zu können. Dabei müssen die Daten direkt beim Eintreffen am Connector an den UCC weitergeleitet werden, bevor die Daten an andere Applikationen zur Verarbeitung/Abspeicherung weitergeleitet werden. Um im Connector des Konsumenten eine (initiale) Weiterleitung der eintreffenden Daten an den UCC zu erzwingen wird das Interceptor Entwurfsmuster[5] verwendet. Beim Interceptor Entwurfsmuster werden alle Verbindungen zwischen zwei Komponenten unterbunden und über eine zusätzliche Komponente geleitet. MYDATA bietet eine entsprechende Implementation für einen Apache Camel Interceptor.

Funktionsweise:

Ohne im Detail die Funktionsweise und die einzelnen Komponenten von XACML und MYDATA zu beschreiben, soll an dieser Stelle in Kürze eine grobe Darstellung der Funktionsweise von MYDATA erfolgen und auf vorhandene Besonderheiten hingewiesen werden.

Die nachfolgende Abbildung zeigt den internen Aufbau des UCC. Der Policy Execution Point (PEP) und der Policy Decision Point (PDP) sind Komponenten aus XACML, welche zur Überprüfung von Policies verwendet werden. Der Policy Management Point (PMP) ist eine Erweiterung von MYDATA und erlaubt die Ablage und Verwaltung von Policies und kann von außen angesteuert werden.

[4] https://camel.apache.org/
[5] http://software-pattern.org/Interceptor

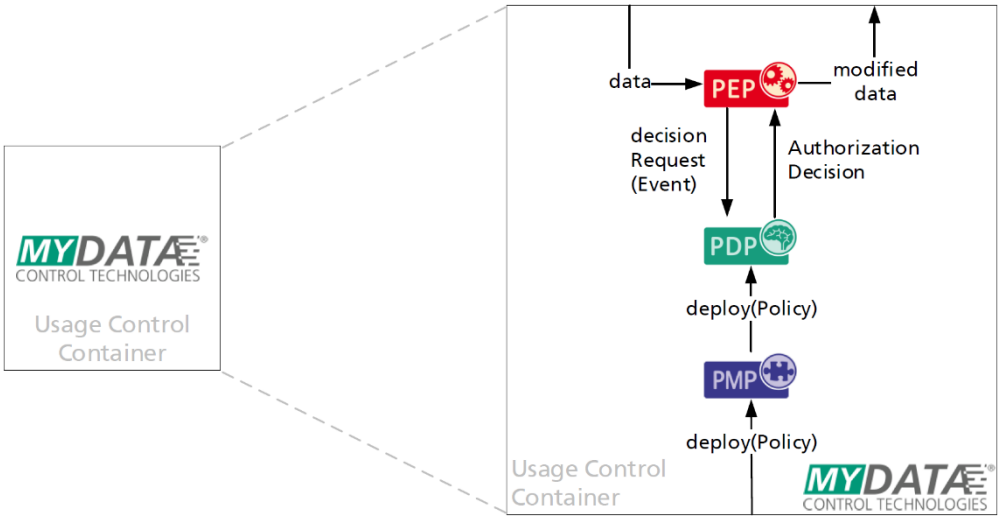


Abbildung 8: Funktionsweise MYDATA UCC gemäß [14]

Eine wichtige Eigenschaft von MYDATA kann der Abbildung entnommen werden: MYDATA erhält Daten als Eingabe und gibt (eventuell) modifizierte Daten zurück. Dies kann etwa der Fall sein, wenn eine Weiterverarbeitung der Ursprungsdaten gemäß der Nutzungsvereinbarung nicht erlaubt ist, eine Modifikation der Daten jedoch eine Weiterverarbeitung erlaubt. In solchen Situationen kann MYDATA die Daten vor der Weitergabe anpassen. Dies muss berücksichtigt werden, da es für bestimmte Szenarien erforderlich oder unerwünscht sein kann.

Beispiel:

Bei der Verwendung von MYDATA ist die zentrale Aufgabe des Anwenders, die definierten Policies für MYDATA nutzbar bereitzustellen. Wird MYDATA beispielsweise verwendet, indem eine spezielle Usage Controll App in eine Apache Camel Route integriert wird, so geschieht dies über eine spezielle HTTP-Schnittstelle. MYDATA verwendet dazu XACML.

Diese Policies können zwar händisch erstellt werden. Dies ist aber im Kontext der IDS nicht sinnvoll, da hier nur eine bestimmte Menge an Policyklassen verwendet wird, die aber im IDS Format vorliegen. Um eine automatische Transformation zwischen diesen IDS-Policies und MYDATA-Policies bereitzustellen bietet MYDATA den Policy Administration Point (PAP). Neben der Transformation von IDS-Policies zu MYDATA-Policies erlaubt der PAP auch das Erzeugen der einzelnen MYDATA-Policies, nachdem die notwendigen Daten in ein entsprechendes Format eingetragen wurden. Hierzu bietet der PAP eine Weboberfläche an, welche in der nachfolgenden Abbildung gezeigt wird.

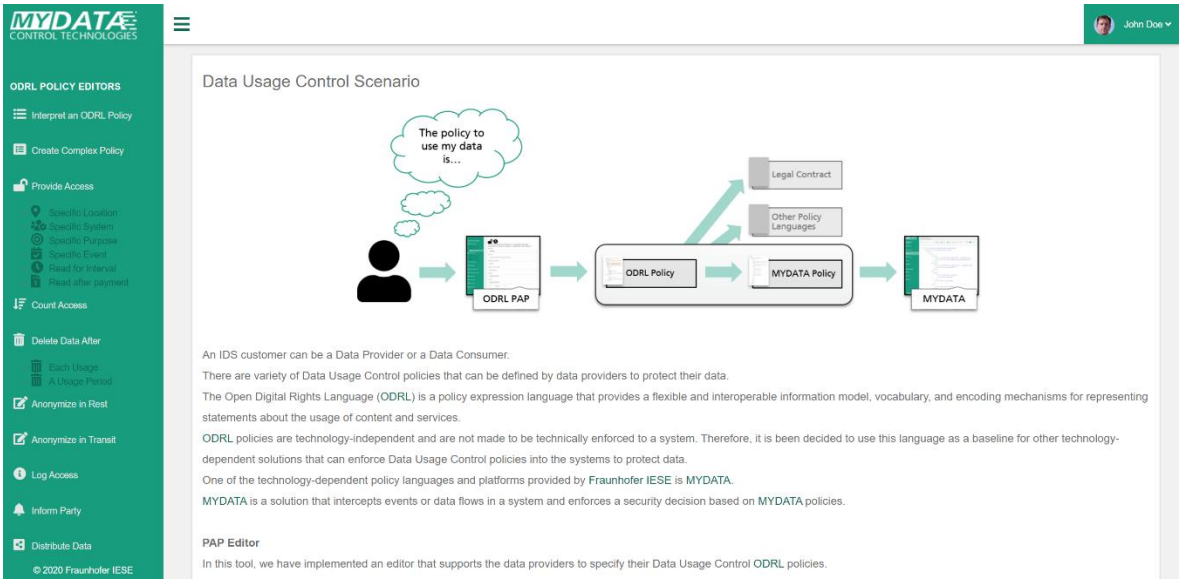


Abbildung 9: ODRL Policy Editor[6]

Im nachfolgenden Codebeispiel befindet sich ein Beispiel für eine IDS-Policy der Kategorie 4 ausgedrückt in ODRL. Policies der Kategorie 4 werden dazu verwendet, die Datennutzung nur für bestimmte Zwecke zu erlauben. Dabei verfügt die Policy neben der eigentlichen Einschränkung (in diesem Fall die Zweckgebundene Nutzung von Daten) mit target, assigner und assignee alle notwendigen Informationen, um in die MYDATA-Policysprache überführt zu werden.

```
{
 "@context": "http://www.w3.org/ns/odrl.jsonld",
 "@type": "Agreement",
 "uid": "http://example.com/policy:restrict-usage-purpose",
 "permission": [{
     "target": "http://example.com/ids-data:my-data",
     "assigner": "http://example.com/ids-party:my-party",
     "assignee": "http://example.com/ids-party:my-data-consu-
mer",
     "action": "ids:use"
     "constraint": [{
         "leftOperand": "ids:purpose",
         "operator": "eq",
         "rightOperand":  { "@value": "http://example.com/ids-
purpose:risk-management-purpose", "@type": "xsd:anyURI" }
       }]
   }]
}
```

Wird eine solche Transformation (beispielsweise mit dem MYDATA-PAP) durchgeführt, erhält man MYDATA-Policy, welche semantisch äquivalent ist. Das Ergebnis einer solchen Transformation kann im folgenden Codebeispiel gefunden werden.

[6] *https://odrl-pap.mydata-control.de/*

```
<policy id='urn:policy:my-data-consumer:restrict-usage-
      purpose'>
  <mechanism event='urn:action:my-data-consumer:use'>
    <if>
      <and>
        <equal>
          <event:string eventParameter='target' jsonPathQuery='$
            .uri'/>
          <constant:string value=http://example.com/ids-data:my-
            data'/>
        </equal>
        <equal>
          <event:string eventParameter='assignee' jsonPathQuery=
            '$.name'/>
          <constant:string value='my-data-consumer'/>
        </equal>
        <pip:boolean method='urn:info:my-data-
            consumer:purpose' default='false'>
          <parameter:string name='purpose-
                  uri' value='http://example.com/ids-pur-
                  pose:risk-management-purpose'/>
        </pip:boolean>
      </and>
      <then>
        <Allow/>
      </then>
    </if>
    <else>
      <Inhibit/>
    </else>
  </mechanism>
</policy>
```

Der MYDATA-PAP minimiert somit den Übersetzungsaufwand von IDS-Policies, welche im IDS-Format vorliegen in die MYDATA-Policysprache. Ursprünglich verfügte nur MYDATA über eine automatische Transformation von IDS-Policies in die eigene Policysprache. Da dies aber die Verwendung und Integration von IDS-Policies erheblich vereinfacht, befinden sich auch für die anderen Usage Control Lösungen vergleichbare Lösungen in der Entwicklung.

3.1.2.2 Logic based Usage CONtrol (LUCON)

Logic based Usage CONtrol (LUCON) ist ein fester Bestandteil des Trusted Connectors, welcher durch das Fraunhofer AISEC entwickelt wird [14]. Bei LUCON handelt es sich um eine Policy Sprache, die verwendet wird, um Datenflüsse zwischen Endpunkten zu kontrollieren. Dabei basiert die Arbeitsweise von LUCON auf einer Menge von Labels, welche an Daten geheftet werden. Diese Labels können im Laufe der Verarbeitung modifiziert werden, d.h. es können neue Labels hinzugefügt und bestehende Labels entfernt werden. Darüber hinaus werden diese Label verwendet um Anfragen und Überprüfungen, welche für die definierten Policies relevant sind, auszuwerten.

Verwendung:

Die Policies, welche für LUCON geschrieben werden, müssen vor ihrer Verwendung zunächst nach Prolog übersetzt werden müssen. Bei Prolog handelt es sich um eine logische Programmiersprache. Diese Übersetzung wird durch die Verwendung eines speziellen Eclipse-Plugins erreicht. Bei Eclipse handelt es sich um ein quelloffenes Werkzeug zur Entwicklung von Software. Die nach Prolog übersetzten Policies werden anschließend in den Trusted Connector geladen und über die Oberfläche aktiviert.

Da der Trusted Connector intern Apache Camel als Message Router verwendet, könnte LUCON auch in anderen Umgebungen, welche Apache Camel verwenden, eingesetzt werden.

Analog zu MYDATA verwendet LUCON das Interceptor-Muster und den entsprechenden Mechanismus in Apache Camel um eine Integration in die betrachteten Datenflüsse zu erreichen. Dies ist in Abbildung 10 dargestellt.

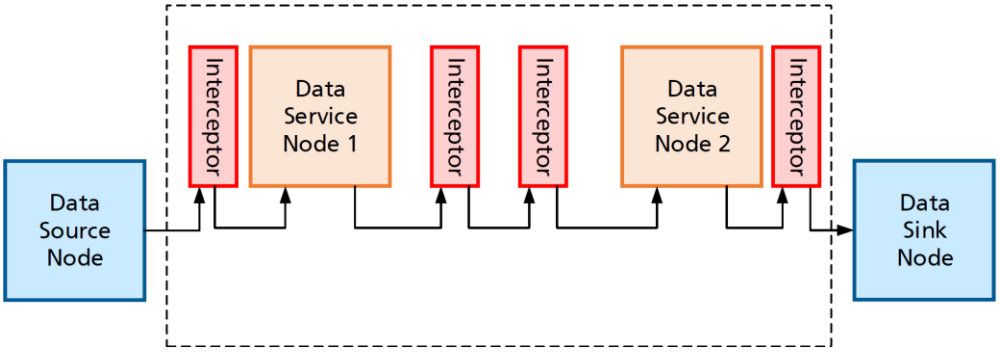


Abbildung 10: Datenfluss in LUCON mit Interceptor gemäß [14]

Funktionsweise:

Abbildung 11 zeigt ein kleines Beispiel für die Funktionsweise von LUCON.

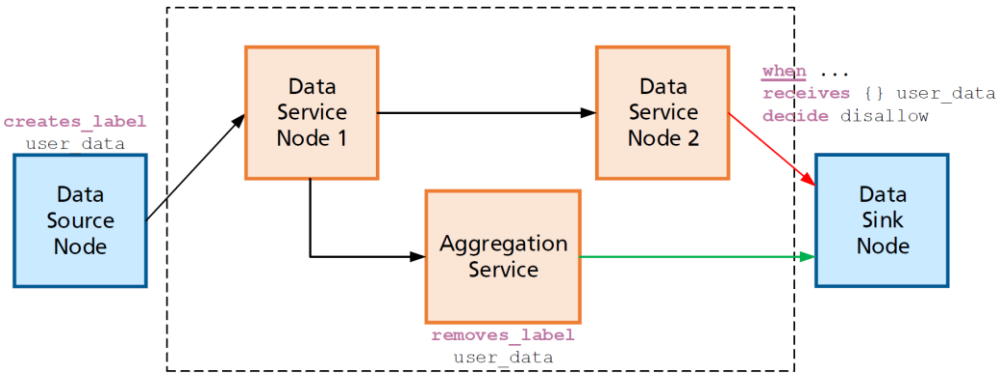


Abbildung 11: LUCON Beispiel mit Service zur Datenaggregation gemäß [14]

Daten die das System (durch die gestrichelte Box gekennzeichnet) betreten, werden mit dem Label user_data versehen. Nach dem Data Service Node 1 gibt es zwei Möglichkeiten, wo die Daten hinfließen können.

Im Aggregation Service werden die Daten durch die Aggregation hinreichend anonymisiert, dass das Label user_data wieder entfernt werden kann.

Data Service Node 2 hat keinen Einfluss auf die Label, die an den Daten hängen.

Die Ausgaben von beiden Knoten (Aggregation Service und Data Service Node 2) sollen das System verlassen und in einen Data Sink Node fließen. Da aber eine Policy in LUCON definiert ist, welche es verbietet, das Daten, welche über das Label user_data verfügen, das System verlassen, werden nur die Ausgaben des Aggregation Service an den Data Sink Node weitergeleitet.

Beispiel:

Nachfolgend werden Beispiele für die Verwendung von LUCON im Umfeld des Trusted Connectors aufgeführt. Diese sind an die offizielle Dokumentation des Trusted Connectors[7] angelehnt. Die Beispiele lassen sich auch auf andere Einsatzszenarien übertragen, in denen LUCON verwendet wird. Dabei wird in diesem Dokument nicht auf die Installation und Verwendung des Eclipse-Plugins eingegangen, sondern nur die Definition der notwendigen Elemente für die Verwendung von LUCON beschrieben.

Zunächst betrachten wir Policies, welche mit der Policysprache von LUCON definiert werden und im Kontext von LUCON Regeln genannt werden.

```
flow_rule {
  id publicData                        // Rule id
  description "Do not leak personal or internal data"
  when publicEndpoint                  // Target identifier
  receives {
    label(personal) or label(internal)
  }                                    // Received message labels
  decide drop                          // Drop message
}
```

Die gezeigte Regel mit dem Namen "anonymized" überprüft Nachrichten, welche am "publicEndpoint" eintreffen und verwirft alle Nachrichten, die über das Label "personal" oder "internal" verfügen. Für ein vollständiges Verständnis der definierten Regel und ihrer Funktionsweise ist es notwendig den Endpunkt "publicEndpoint" weiter aufzuschlüsseln. Diese ist für das gegebene Beispiel sehr kurz und definiert nur den Endpunkt in Apache Camel, auf den sich die Servicedefinition beruft. Die entsprechende Definition kann dem nachfolgenden Codebeispiel entnommen werden.

```
service {
  id publicEndpoint

  // Defines the Camel endpoints for which this service
     description applies, using a specific endpoint address.
  endpoint 'http://localhost/service'
}
```

Um sicherzustellen, dass die zuvor definierte Regel auch eingehalten wird, wird in die entsprechende Route im Apache Camel ein Service zur Anonymisierung integriert. Damit dies auch entsprechend von LUCON während der Auswertung der Regeln berücksichtigt werden kann, ist es notwendig, eine entsprechende, für LUCON nutzbare, Definition des Service anzulegen. Diese enthält neben dem Endpunkt auch eine Auflistung von Eigenschaften, die der Service bereitstellen kann. Darüber hinaus

[7] https://industrial-data-space.github.io/trusted-connector-documentation/

ist definiert, wie die Labels, die an einer Nachricht heften, modifiziert werden, wenn der Service aufgerufen wird.

```
service {
  id anonymizerService

  // Defines the Camel endpoints for which this service
     description applies, using a specific endpoint address.
  endpoint 'http://localhost/anonymizer'

  // Capabilities can be required by a flow_rule. If not
     required, nothing will happen
  capabilities
    anonymization: personal_data([surname,name])

  // Properties describe the service's behavior.
  removes_label personal
  removes_label internal
}
```

3.1.2.3 D°

Bei D° (gesprochen di'grē) handelt es sich um eine domänenspezifische Programmiersprache (DSL), welche über integrierte Mechanismen für Usage Control verfügt ([14], Kapitel 4.3). Die von D° adressierte Domäne ist die Datenverarbeitung, d.h. D° erlaubt es Entwicklern eine Vielzahl von datenverarbeitenden Applikationen (sog. Data Apps) zu entwickeln. D° ist im Rahmen der IDS entstanden und wird durch das Fraunhofer ISST entwickelt. Bei D° handelt es sich nicht um eine interpretierte Sprache. Auch wird D°-Code nicht direkt in maschinenausführbaren Code übersetzt. Stattdessen verwendet D° eine andere Programmiersprache als sog. Host Language. Dabei wird der D°-Code zunächst in Programmcode der Host Language übersetzt, welcher anschließend in eine ausführbare Applikation übersetzt wird. Die aktuelle Implementation von D° verwendet Java als Host Language.

Verwendung:

Anders als bei LUCON und MYDATA ist es nicht das Ziel von D°, existierende Software und Datenflüsse mit Usage Control nachzurüsten. Stattdessen betrachtet D° Usage Control als elementaren und untrennbaren Bestandteil von Applikationen vom Beginn der Entwicklung an. Da D° dem Programmierparadigma der policy-agnostischen Programmierung folgt, werden die Policies und der notwendige Code für deren Umsetzung nicht in den Code der Applikation eingefügt.

Das Programmierparadigma der policy-agnostischen Programmierung beschreibt eine Trennung von Applikationslogik und Policies (sowie deren Enforcement) eine Verknüpfung der beiden Konzepte findet zu einem späteren Zeitpunkt statt und bildet dann eine Einheit in der die beiden Konzepte untrennbar miteinander verwoben sind. Aus diesem Grund sind der eigentlichen Applikationsentwicklung in D° zwei zusätzliche Schritte vorgelagert.

Bevor die eigentliche Applikationslogik als D°-Programm ausgedrückt werden kann muss überprüft werden, ob alle benötigten Elemente vorhanden sind, um das Problem abzubilden. D° bietet Systeme für Datentypen, Policies und Aktivitäten. Im Kontext von D° ist eine Aktivität eine (für D°) atomare Funktion, welche eine beliebige Menge Programmcode der Host Language enthalten kann. Alle drei dieser Systeme können durch den Anwender erweitert werden, um D° für verschiedenste Szenarien

verwenden zu können. Sollten noch Elemente für die zu entwickelnde Data App fehlen, müssen diese in das System eingefügt werden. Dabei muss immer eine textuelle Spezifikation erfolgen und im Falle von Policies und Aktivitäten muss zusätzlich noch eine Implementation in der Host Sprache geschehen. Sobald alle notwendigen Elemente definiert sind, kann zum nächsten Schritt übergegangen werden.

Die Definitionen der Elemente können zunächst nicht in D°-Applikationen verwendet werden. Zuvor ist es notwendig, dass aus den Aktivitäten Instanzen erzeugt werden. Das besondere an diesen Instanzen ist, dass sie mit Policies versehen werden können und so zu einem festen Bestandteil der Aktivität werden. Für Datentypen ist ein ähnliches Verfahren in Planung, aber aktuell können die Definitionen der Datentypen direkt in D° Applikationen verwendet werden. Policies können zu keinem Zeitpunkt direkt im D°-Code verwendet werden, müssen aber ebenfalls instanziiert werden, um mit anderen Elementen verknüpft werden zu können. Bei der Instanziierung von Policies ist es möglich, eventuell vorhandene Parameter mit konstanten Werten zu belegen, sollte dies gewünscht sein. Sobald alle notwendigen Instanzen für die Applikation vorhanden sind, kann die eigentliche Entwicklung der Applikation beginnen.

Die vorgelagerten Schritte vor der Entwicklung einer Data App mit D° verursachen unter Umständen einen initialen Mehraufwand, welcher aber durch den wachsenden Funktionsumfang der Sprache und die mögliche Verwendung von Erweiterungen Dritter (vergleichbar mit Bibliotheken in anderen Programmiersprachen) minimiert werden kann.

Die eigentliche Entwicklung mit D° unterscheidet sich nicht von der Verwendung anderer Programmiersprachen. Was an dieser Stelle noch verbleibt ist die Integration der Policies, welche an den Aktivitäten hängen, in die Applikationslogik. Dieser Schritt wird im Rahmen der Übersetzung automatisch durchgeführt. Der Codegenerator, welcher Teil des D°-Compilers ist, stellt sicher, dass alle verknüpften Policies an den notwendigen Stellen überprüft werden.

Funktionsweise:

Eine Policy im Kontext von D° ist alleine nicht ausführbar und ist immer mit einem anderen Element verknüpft. Dabei besteht jede Policy aus drei unterschiedlichen Teilen:

1. Eine Precondition, welche vor dem Aufruf des verknüpften Elements überprüft wird.
2. Eine Postcondition, welche nach dem Aufruf des verknüpften Elements überprüft wird.
3. Eine Security Manager Intervention, welche bei Bedarf aufgerufen wird.

Während Pre- und Postcondition in jedem Fall ausgewertet werden (sofern die Ausführung nicht wegen Policyverstößen abgebrochen wird), findet eine Security Manager Intervention nur dann statt, wenn bestimmte Funktionen der Host Language verwendet werden. Um eine bestimmte Funktion bzw. Schnittstelle durch den D°-Security Manager schützen zu können, ist es notwendig die entsprechenden Funktionen zu instrumentieren und vor der eigentlichen Ausführung eine Anfrage an den Security Manager zu schicken. In Java sind diverse APIs (bspw. I/O und Netzwerk) bereits instrumentiert, um entsprechende Prüfungen durch den Java Security Manager zu erlauben. Die aktuelle Implementation von D° enthält eine angepasste Implementation des Java Security Managers und setzt auf dieses System auf.

Hieraus ergeben sich mehrere Vorteile:

- Bei korrekter und vollständiger Instrumentierung ist es nicht möglich eine Funktion aufzurufen, ohne dass der Security Manager darüber entscheiden kann.
- Die Anfrage an den Security Manager erfolgt unmittelbar vor der eigentlichen Ausführung der geschützten Funktion. Aus diesem Grund sind auch schon die verwendeten Parameter bekannt

und es ist beispielsweise möglich mit Policies zu limitieren, welche Datenmengen per Netzwerk versendet oder auf die Festplatte geschrieben werden dürfen.

Die nachfolgende Abbildung enthält eine schematische Darstellung der Codegenerierung für ein einzelnes Statement (bspw. einen Aktivitätsaufruf). Dabei repräsentieren grüne Knoten Statements bzw. deren Übersetzung in die Host Language und rote Knoten Code der für die Überprüfung von Policies notwendig ist. Es ist erkennbar, dass der generierte Kontrollfluss in der unteren Hälfte in immer durch Pre- und Postcondition läuft, sofern kein Fehler auftritt.

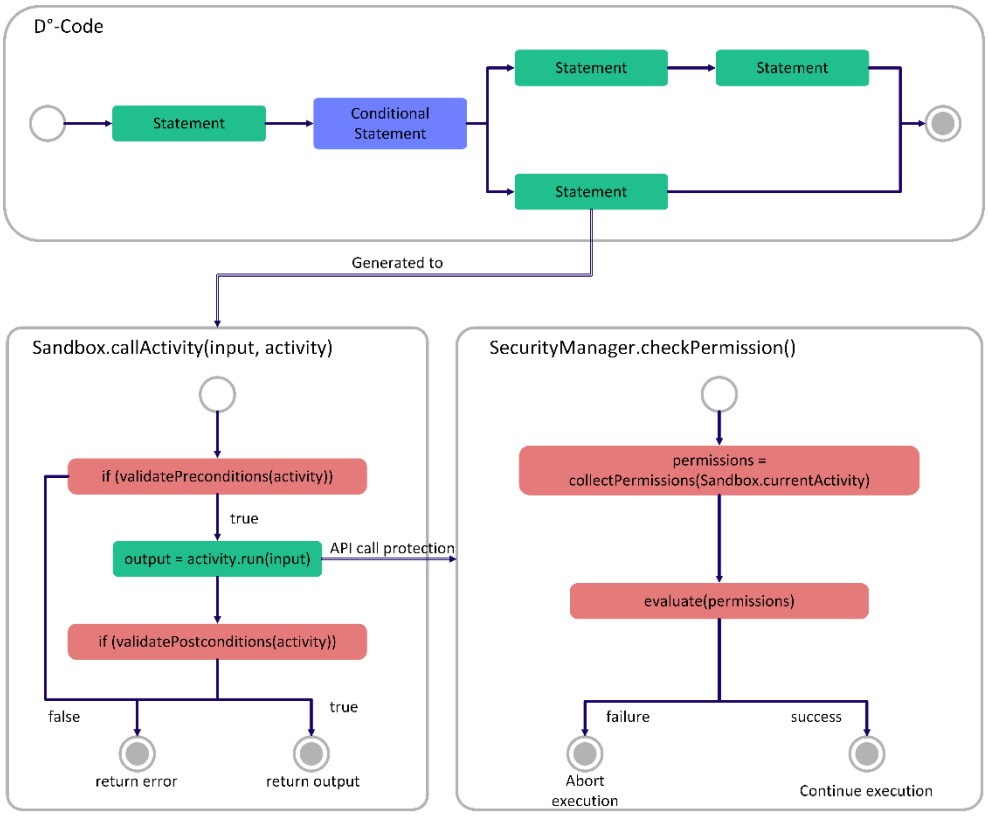


Abbildung 12: Schematische Darstellung der D°-Codegenerierung

Beispiel

Da Usage Control und entsprechende Policies ein fester Bestandteil der Programmiersprache D° sind, funktioniert die Verwendung grundlegend anders als bei LUCON und MYDATA. Bereits während der Applikationsentwicklung ist es notwendig die entsprechenden Policies in die Applikation zu integrieren. Dies wird nachfolgend an einem Beispiel erklärt, ohne dabei im Detail auf den Aufbau und die Verwendung von D° (bspw. im Bezug auf die Syntax) einzugehen.

In besagtem Beispiel soll eine simple Echo-Applikation implementiert werden, welche Nachrichten entgegennimmt und ohne Veränderungen zurückgibt. Dabei sollen die folgenden Eigenschaften erfüllt werden:

- Die Applikation bietet einen einzelnen HTTP-Endpunkt unter der URL /echo, welcher zum Aufruf der Applikation verwendet wird
- Die Applikation ist auf Port 5000 erreichbar
- Jede behandelte Nachricht wird auf der Kommandozeile wiedergegeben
- Die Applikation kann nur im Zeitraum 01.01.2010 bis 31.12.2025 verwendet werden
- Es werden nur Nachrichten geechot, die eine maximale Länge von 1024 Zeichen haben

Die letzten beiden Punkte der Auflistung sollen dabei durch Usage Control Policies umgesetzt werden. Das nachfolgende Codebeispiel enthält den Programmcode für die beschriebene Applikation, umgesetzt in D°.

```
configuration
    - namespace : "de_fhg_isst_oe270.degree"
    - name : "demoProject"
    - version : "0.0.1-1-SNAPSHOT"
    - startupPolicies : "UseNotBeforeTimeStamp2010, UseNotAfterTimeStamp2025"
    - tags: "EXAMPLE"
    - execution: "single"
    - port : "5000"
    - url: "echo"

code
    [payload = $Text] -> begin
        echoMsg = $Text(@write["Received message:"]);
        UnconstrainedPrintToConsole[echoMsg];

        EchoServicePrintToConsole[payload];

        echoMsg = $Text(@write["End of message."]);
        UnconstrainedPrintToConsole[echoMsg];

        return [payload];
    end
```

Abbildung 13: Beispielapplikation in D°

Dabei ist der Code einer in D° entwickelten Applikation (sog. Data App) in zweigeteilt. Zum einen ist die Konfiguration der Data App und zum anderen die eigentliche Applikationslogik enthalten. Im Beispiel sehen wir, dass in der Konfiguration "startupPolicies" definert sind, welche ihrem Namen nach die zuvor geforderte zeitliche Beschränkung umsetzen. In der eigentlichen Applikationslogik finden wir aber keinen Hinweis auf die Verwendung irgendwelcher Policies. Dies liegt darin begründet, dass D° das Paradigma der policy-agnostischen Programmierung umsetzt. Dieses basiert auf der Trennung von Usage Control und Applikationslogik und einer späteren Verknüpfung zu einer untrennbaren Einheit. Dabei wird diese Verknüpfung durch den Compiler während der Übersetzung von D°-Code in Java-Code geleistet. Dieser Java-Code wird dann übersetzt, um die eigentliche ausführbare Applikation zu erhalten.

Um zu verstehen, wo die Informationen über die verknüpften Policies verbleiben, ist es notwendig, das Bausteinprinzip von D° zu betrachten: Alle Sprachelemente, die der Nutzer verwendet (Datentypen, Policies und Aktivitäten [dies sind Funktionsblöcke, die innerhalb von D° atomar sind]) verfügen über eine textuelle Definition und, sofern notwendig, über eine zusätzliche Implementation. Die so definierten Elemente sind nicht direkt in Data Apps verwendbar (mit Ausnahme von Datentypen) und müssen zunächst instanziiert werden. Dabei können Eingabeparameter mit konstanten Werten belegt werden, sofern dies für den Anwendungsfall notwendig ist. Viel wichtiger ist aber, dass die Instanziierung es erlaubt, Policies mit Aktivitäten zu verknüpfen und somit die notwendige Information erzeugt, welche der Eingabeparameter notwendig sind, um die richtigen Policies an den passenden Stellen zu überprüfen. Die nachfolgende Abbildung stellt dies nochmal übersichtlich dar.

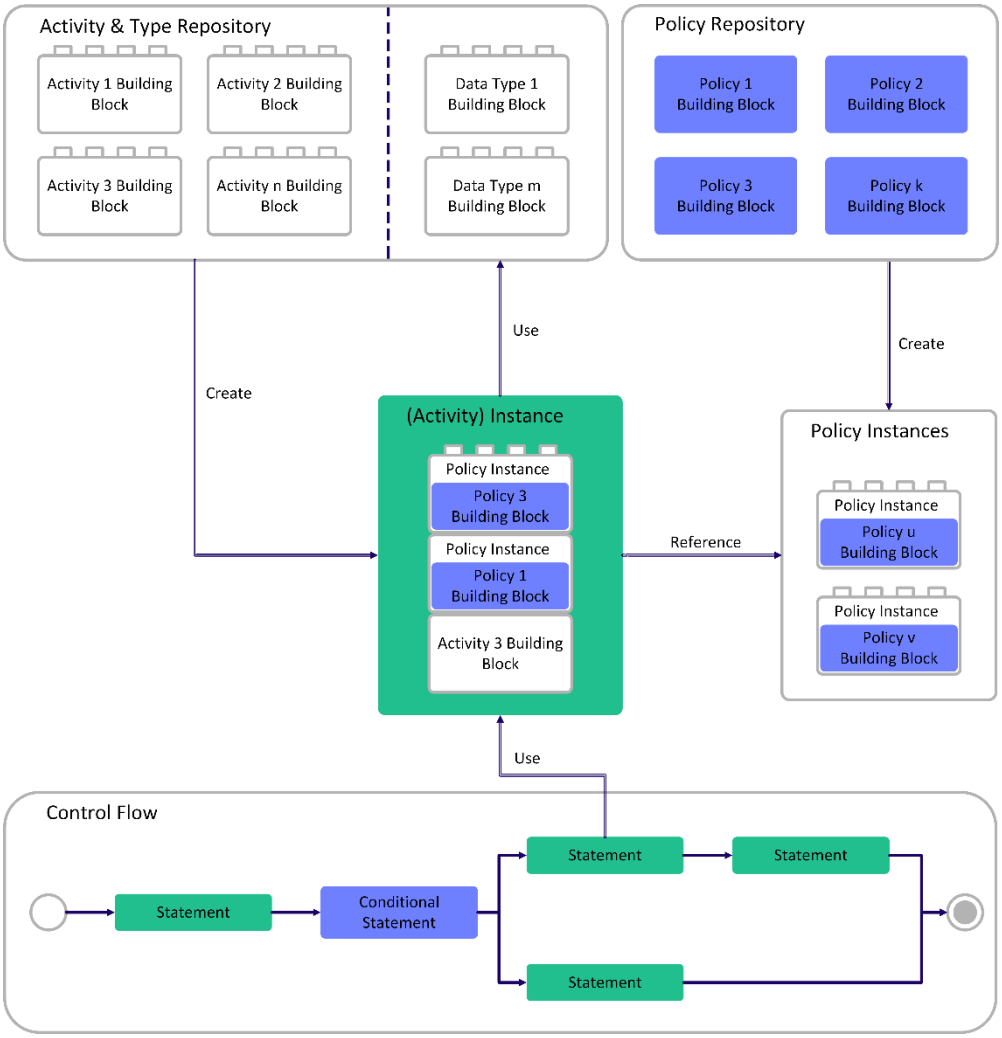


Abbildung 14: Schematische Darstellung der Komposition verschiedener Sprachelemente von D°

Betrachten wir nun exemplarisch die Umsetzung der Längenbegrenzung für geechote Nachrichten. Dabei soll die entsprechende Policy überprüft werden, sobald die eingegangene Nachricht auf die Konsole ausgegeben wird. Bevor die Policy betrachtet wird, muss zunächst die Aktivität für die Ausgabe auf die Konsole angelegt werden. Diese verfügt über einen einzelnen Eingabeparameter, nämlich den auszugebenden Text. Die entsprechende Definition findet sich in der folgenden Abbildung.

```
PrintToConsole:
  degree.Activity@PrintToConsole:
    name:
      Identifier: "PrintToConsole"
    inputParameters:
      degree.Parameter:
        - name:
            Identifier: "text"
          type:
            Type: "Text"
    executionContainer:
      degree.ExecutionContainer: "java"
```

Abbildung 15: D°-Aktivitätsdefinition

Um den Umfang dieses Whitepapers nicht zu sprengen, wird auf die Beschreibung der Implementationen verzichtet. In Abbildung 15 ist eine Aktivität, die eine gewünschte Funktionalität bereitstellt (Echo der Nachricht), definiert worden und kann instanziiert und anschließend in Data Apps verwendet werden. Da für das Beispiel aber auch eine Policy benötigt wird, muss diese ebenso in einem vorangegangenen Schritt definiert und implementiert werden. Diese verfügt über zwei Eingabeparameter: Den zu überprüfenden Text und die gültige Maximallänge. Die entsprechende Definition findet sich in der nächsten Abbildung.

```
MaxLength:
  degree.Constraint@MaxLength:
    name:
      Identifier: "MaxLength"
    attribute:
      degree.Parameter:
        - name:
            Identifier: "maxLength"
          type:
            Type: "UnsignedInteger"
        - name:
            Identifier: "content"
          type:
            Type: "Text"
```

Abbildung 16: D°-Constraintdefinition

An dieser Stelle sind alle Elemente für eine längenbeschränkte Ausgabe auf der Console verfügbar und die entsprechenden Instanzen können erzeugt werden. Zunächst wird eine Instanz für die Policy erzeugt. Da wir in diesem Beispiel immer eine maximale Nachrichtenlänge von 1024 Zeichen haben, kann dies direkt in der Instanz vermerkt werden. Somit werden spätere (versehentliche oder absichtliche) falsche Nutzungen verhindert. Im nächsten Codebeispiel findet sich die entsprechende Instanziierung.

```
MaxLength1024:
  degree.ConstraintInstance@MaxLength1024:
    name:
      Identifier: "MaxLength1024"
    definition:
      degree.ConstraintReference: "degree.Constraint@MaxLength"
    mappedElements:
      degree.InstanceMap:
        - key:
            Text: "maxLength"
          value:
            Json: "{\"core.UnsignedInteger\":\"1024\"}"
```

Abbildung 17: D°-Constraintinstanz

Der letzte Schritt ist es, die eigentliche Aktivität zu instanziieren und dabei die notwendigen Policies zu verknüpfen. Dabei ist es an der Stelle noch wichtig, dass der Eingabeparameter der Aktivität auf den verbleibenden freien Eingabeparameter der Policy abgebildet wird. Hierdurch ist sichergestellt, dass die Policy während ihrer Überprüfung Zugriff auf die notwendigen Daten hat. Gleichzeitig wird

vermieden, dass die Policy auf eventuell vorhandene Daten Zugriff hat, die für die Evaluierung nicht notwendig sind.

In der nachfolgenden Abbildung ist die entsprechende Instanziierung der Aktivität zu sehen. Neben der zuvor beschriebenen Policy zur Prüfung der maximalen Länge ist noch eine weitere Policy vorhanden. Diese erlaubt Aufrufe nur in einem gegebenen Zeitintervall. Dieses wurde für die konkrete Instanz auf die Zeiten gesetzt, welche zuvor in den Anforderungen definiert wurden.

```
EchoServicePrintToConsole:
  degree.ActivityInstance:
    name:
      Identifier: "EchoServicePrintToConsole"
    definition:
      degree.ActivityReference: "degree.Activity@PrintToConsole"
    policies:
      degree.MappedPolicyInstanceMap:
        - key:
            Text: "maxLength"
          value:
            degree.ConstraintOrPolicyInstanceReference: "degree.ConstraintInstance@MaxLength1024"
        - key:
            Text: "allowTimeInterval"
          value:
            degree.ConstraintOrPolicyInstanceReference: "degree.PolicyInstance@AllowedTimeInterval2010to2025"
    parameterMappings:
      degree.ParameterMappingsMap:
        - key:
            Text: "text"
          value:
            Text:
              - "maxLength.content"
```

Abbildung 18: : D°-Aktivitätsinstanz

Auf diese Weise werden die Usage Control Mechanismen innerhalb von D° verwendet. Der Entwickler, welcher D° verwendet, muss dabei nichts für die Überprüfungen der Policies tun, da die Prüfung automatisch durch den Code Generator geleistet wird.

Im aktuellen Entwicklungsstand von D° ist es noch nicht möglich, Policies auch mit Datentypen zu verknüpfen. Diese Funktionalität befindet sich in Arbeit und wird die Nutzung des D°-Policy Systems weiter vereinfachen und flexibilisieren.

3.1.2.4 Vergleich

Ein Vergleich der unterschiedlichen Lösungen für Usage Control in den IDS kann im Dokument "Usage Control in the International Data Spaces", Version 2.0 in den Abschnitten 4.4 und 4.5 gefunden werden. Dabei unterscheiden sich die verschiedenen Lösungen sowohl in ihren Ansätzen und verwendeten Verfahren, als auch in ihrer Mächtigkeit. Jede der Lösungen hat eine eigene Untermenge der in den IDS definierten Policy-Muster, welche umgesetzt werden können. Darüber hinaus unterscheiden sich die Lösungen auch in ihrem Reifegrad. MYDATA ist dabei in der Entwicklung am weitesten fortgeschritten und verfügt über einen Technology Readiness Level (TRL) von 7-8, wogegen LUCON einen TRL von 5 und D° von 4 aufweisen.

Zur Auswahl des Usage Controll Modells und zur Priorisierung möglicher Szenarien ist es dabei notwendig, die Präferenzen der Teilnehmer und die Umstände des Datenaustauschs in IIP-Ecosphere genauer zu untersuchen.

3.1.2.5 Sonderfall Echtzeitdaten

Als Echtzeit versteht man den Betrieb eines Rechensystems, das Programme zur Verarbeitung anfallender Daten stetig betriebsbereit hält. Man unterscheidet zwischen weicher und harter Echtzeit. Bei harten Echtzeitanforderungen müssen alle Berechnungen innerhalb eines vorgegebenen

Zeitintervalls abgeschlossen werden. Dies ist vor allem für sicherheitsrelevante Bereiche, wie etwa das Stoppen einer Maschine bei auftretenden Problemen, wichtig. Bei weichen Echtzeitanforderungen können hingegen Abweichungen von den vorgegebenen Zeitintervallen toleriert werden. Für die Datenverarbeitung bedeutet Echtzeit, dass die zeitliche Verzögerung zwischen einem Ereignis im Betrieb und der dazugehörigen Datenaufnahme, Datenanalyse und Bereitstellung der Ergebnisse in einem bestimmten Zeitrahmen erfolgen müssen, da das Ergebnis sonst an Nutzen verliert [17]. Werden mit einer Echtzeitanforderung mit Partnern geteilt oder analysiert, so müssen bei der Implementierung von Datentransformationen oder Maßnahmen wie Access und Usage-Control die dadurch bedingten Verzögerungen betrachtet werden. Es gilt zu verhindern, dass eine Implementierung dieser Maßnahmen eine zeitliche Verzögerung der Datenübertragung hervorruft, die eine Nutzenreduktion von Daten oder Service zur Folge hat. Demnach ist es wichtig, den jeweiligen Anwendungsfall zu beurteilen, um zu ermitteln, ob Usage Control Maßnahmen eine akzeptable Verzögerung hervorrufen und verwendet werden können.

3.1.3 Gestaltung des Datenmarktplatzes

Daten stellen das Fundament der digitalen Wirtschaft dar. Auf der Basis von Daten werden neue digitale Leistungsangebote entwickelt und neuartige Geschäftsmodelle entstehen [18]. Weiterhin können (industrielle) Prozesse durch die Nutzung von Daten optimiert und automatisiert und die Produktivität gesteigert werden. Gleichermaßen werden im Rahmen der Datenökonomie Daten immer mehr zu einem eigenständigen ökonomischen Gut, dass einen Wert besitzt und an andere Organisationen verkauft werden kann [8]. Obgleich des hohen Potentials bleiben Geschäftsdaten oftmals privat. Dies bevorteilt Großunternehmen, die in der Lage sind immense Datensätze aufzubauen und diese gewinnbringend zu verwerten. Um datenbasierte Innovationen auch für kleine und mittelständische Unternehmen mit geringerem Datenbestand möglich zu machen besitzen Datenmarktplätze eine entscheidende Funktion. Datenmarktplätze fungieren als intermediäre, die verschiedene Teilnehmergruppen miteinander verbinden und die Transaktion von Daten effizienter gestalten. Dazu stellen sie eine Infrastruktur zum Austausch von Daten und datenbezogenen Dienstleistungen bereit, welche die Bereitschaft der Teilnehmer zum Datenaustausch erhöht. Ein Datenmarktplatz stellt sicher, dass Daten hochwertiger Qualität angeboten werden. Auf der anderen Seite vereinfacht er die Möglichkeit die Datenbasis eines Unternehmens zu monetarisieren [6]. Dementsprechend kann ein Datenmarktplatz als Handelsplattform für Datengüter einen wichtigen Aspekt innerhalb eines Datenökosystems, das auch vergleichsweise kleinere Teilnehmer beheimatet, wie etwa IIP-Ecosphere, darstellen.

Basierend auf den zuvor erläuterten Eigenschaften wird ein Datenmarktplatz zusammenfassend folgend definiert: Bei einem Datenmarktplatz handelt es sich um eine digitale Plattform, auf welcher Datenprodukte gehandelt werden. Diese Plattform muss als neutraler Intermediär handeln und möglichst vielen Teilnehmern erlauben ihre Daten auf der Plattform bereitzustellen und zu verkaufen. Dabei kann die Plattform sowohl statische Datenpakete als auch Daten-Streams bereitstellen. Der Zugriff auf die Daten kann dabei auf verschiedene Art und Weise geschehen. Zu den möglichen Zugriffsarten können u.a. individuelle Downloads, Anwender-Programmierschnittstellen oder Webschnittstellen gehören. Weiterhin verfügen Datenmarktplätze über standardisierte Lizenzmodelle und Richtlinien hinsichtlich des Datenzugriffs und der Nutzungskontrolle [6].

Die Entwicklung neuartiger Datenplattformen im Bereich des kommerziellen Datenhandels stellt einen unbestreitbaren Trend dar [6]. Unternehmen beginnen die Notwendigkeit des Datenaustausches zu verstehen und Plattformanbieter versuchen sich frühzeitig zu positionieren. Allerdings zögern viele Unternehmen noch mit dem Angebot von Daten auf Datenmarktplätzen und dem Erwerb von Daten. Dabei spielen vielfältige Ursachen eine Rolle. So mangelt es vielen Unternehmen an Vertrauen bei Nutzung der Marktplätze [10]. Die Organisationen haben Angst, dass Konkurrenten ihr Datenangebot nutzen könnten, um ihr Konkurrenzangebot zu verbessern oder Geschäftsgeheimnisse zu

entschlüsseln [19]. Ein weiteres, in diesen Bereich fallendes, Problem stellt der Kontrollverlust der Daten nach der Übertragung dar. Selbst wenn die Daten nicht an ein Konkurrenzunternehmen übertragen werden, haben die potentiellen Datenbieter Angst, dass ihre Daten für nicht autorisierte Zwecke verwendet werden [7]. Ein weiteres Hindernis stellt für viele Teilnehmer die Sicherheit der Daten während des Datenaustausches sowie bei Nutzung durch den Käufer dar. Sie befürchten, dass nicht autorisierte Teilnehmer sich unerlaubt Zugang zu den Daten verschaffen und diese missbrauchen [1]. Auf der Seite der Datenkäufer stellt sich vor allem die geringe Zahlungsbereitschaft als hinderlich heraus. Aufgrund der Eigenschaft von Daten als Erfahrungsgut oder gar als Kredenzgut kann der Datenwert oftmals erst während der Nutzung oder, im letzteren Fall, nach Verwendung der Daten bestimmt werden [6].

Um einen Datenmarktplatz, wie er in IIP-Ecosphere entstehen soll, für möglichst viele potentielle Teilnehmer attraktiv zu gestalten müssen sowohl technische als auch organisatorische Maßnahmen getroffen werden, die den Vorbehalten gegenüber der Verwendung eines Datenmarktplatzes zum Austausch von Daten entgegenwirken. Zuvor wurden bereits die Gestaltung der Architektur und die Nutzung von Usage Control als potentielle Lösungsmaßnahmen erläutert. Diese sind neben der Plattform zur Nutzung der Services auch für den Datenmarktplatz relevant. Zusätzlich ist es zur Gewährleistung von Datenschutz und Datensicherheit wichtig, dass jeder Teilnehmer seine jeweiligen Rollen auf dem Datenmarktplatz versteht, um eventuelle Governance Maßnahmen zu treffen.

Rollen in Datenmarktplätzen:

Zur Bewertung des durch den Datenmarktplatz gelieferten Mehrwertes für die Organisation muss diese ihre strategischen Optionen bewerten. Dazu gehört die Festlegung der einzunehmenden Rollen. Diese bedingt das organisatorische Setup und die technischen Erfordernisse einer Teilnahme am Datenaustausch. Ein Unternehmen kann auf einem Datenmarktplatz eine oder mehrere Rollen einnehmen. Dies ist etwa möglich, wenn sowohl Daten bezogen, als auch angeboten werden. Die wichtigsten Rollen auf Datenmärkten werden in Anlehnung an das IDS RAM [13] folgend dargestellt.

Datenanbieter besitzen Daten und stellen diese als Produkte auf dem Datenmarktplatz zur Verfügung. Sie legen fest, unter welchen Bedingungen die von Ihnen angebotenen Datenprodukte gesehen oder abgerufen werden können. Zur Verteilung der Daten an die anderen Teilnehmer verwendet der Datenanbieter technische Lösungen, die ggf. extern vorgegeben werden können. Um sein Angebot sichtbar zu machen und den Datenaustausch zu vereinfachen, sollte der Datenanbieter sein Datenangebot mittels qualitativ hochwertiger Metadaten beschreiben. Zudem sollte der Datenanbieter die Transaktion loggen, um bei möglichen Komplikationen Fehler nachvollziehen zu können.

Datennutzer beziehen auf dem Datenmarktplatz gehandelte Datenprodukte vom Datenanbieter. Bevor der Austausch eines Datenprodukts angestoßen wird, kann der Datennutzer mittels Intermediär das Angebot des Datenmarktplatzes basierend auf Metadaten durchsuchen. Zur Etablierung eines Datenaustauschs muss der Datennutzer die Sicherheitsanforderungen des Datenanbieters erfüllen. Die Datennutzung darf nur hinsichtlich des zuvor festgelegten Kontexts erfolgen und einzelne Methoden der Datenverarbeitung können durch den Datenanbieter verboten sein. Äquivalent zum Datenanbieter sollte auch auf Seiten des Datennutzers ein Logging stattfinden, um bei etwaigen Komplikationen die Fehler nachvollziehen zu können.

Zur vereinfachten Vermittlung von Daten und Transaktionsdurchführung werden auf Datenmarktplätzen die Funktionen sogenannter Intermediäre genutzt. Funktionen eines Intermediäres nimmt zumeist der Datenmarktplatzbetreiber ein. Zur Vermittlung von Angebot und Nachfrage werden Metadaten durch den Intermediär (*Broker*) gesammelt, verwaltet und zur Verfügung gestellt. Der Datenmarktplatz stellt dazu organisatorische und technische Mittel bereit.

Dazu gehört zum Beispiel die Festlegung des Metadaten-Modells, ein Interface zur Bereitstellung der Metadaten und die Möglichkeit diese Metadaten abzufragen. Bei Einnahme der Rolle als *Clearing House* sorgt der Intermediär für die Transaktionsabwicklung. Das Clearing House zeichnet alle Aktivitäten im Rahmen des Datenaustauschs auf. Nachdem der Datenaustausch vollzogen wurde, senden Datenanbieter und Datennutzer ebenso ihr Logging, um den Abschluss des Datentransfers zu bestätigen. Basierend auf diesen Informationen wird der Datentransfer anschließend abgerechnet. Weiterhin können die Aufzeichnungen auch dazu genutzt werden eventuell auftretende Konflikte zu lösen. Oftmals werden die Rollen Broker und Clearing House von demselben Intermediär ausgefüllt, da es sich bei beidem um eine vertrauenswürdige Instanz handelt, die zwischen Datenanbieter und Datennutzer vermittelt.

3.1.4 Governance-Mechanismen

Während zuvor technische Maßnahmen und Methoden zur Gewährung von Datenschutz und Datensicherheit durch die Entscheidungsinstanz vorgestellt wurden, werden in diesem Abschnitt, basierend auf der in Abschnitt 2 vorgestellten Problemstellung, Governance-Mechanismen und organisatorische Methoden in Daten- und Plattformökosystemen betrachtet. Während sich organisationsinterne Governance mit der Maximierung des Datenwertes durch Verbesserung der Datenqualität, der Klärung von Verantwortlichkeiten für einzelne Datenkategorien und dem Aufbau von Strukturen zur Datenverarbeitung innerhalb des Unternehmens befasst, haben diese etablierten Maßnahmen nur beschränkte Relevanz in einem Datenökosystem. Governance von Datenökosystemen befasst sich hingegen mit den Mechanismen zur Etablierung eines unternehmensübergreifenden Ökosystems unter dem Zielkonflikt von Offenheit und Kontrolle mit der Intention Anreize zur Teilnahme am Datenaustausch und zur Wertschöpfung mittels der Plattform zu schaffen [20].

Da es sich bei einem Plattformökosystem um ein neuartiges Konzept handelt, wurde der Gestaltung von Governance in diesem Bereich in der Forschung bisher nur eine geringe Aufmerksamkeit zuteil [21]. Der Fokus lag dabei zumeist auf der Analyse der Governance-Struktur bereits erfolgreich etablierter Plattformen des B2C-Marktes wie etwa Youtube, eBay oder Facebook. Diese Plattformbetreiber besitzen jedoch aufgrund Marktmacht die Möglichkeit Governance-Entscheidungen ohne umfangreiche Abstimmung mit den anderen Plattformteilnehmern zu treffen. Sie können nahezu exklusiv die Dynamik und die Art des Datenaustauschs auf der Plattform festlegen. Aufgrund des großen Kontrollbereichs werden dort Ökoystem-Governance-Entscheidungen ähnlich zur organisationsinternen Governance getroffen [20]. Dementsprechend kann größtenteils nicht auf etablierte Rahmenwerke zur Governance von dezentral organisierten Datenökosystemen zurückgegriffen werden.

Die unternehmensinterne Governance soll sich an den Zielen der Organisation ausrichten. Äquivalent dazu ist es notwendig, dass sich die Plattform-Governance an den Zielen der Plattform ausrichtet und diese durch ihre Mechanismen unterstützt. In der Wissenschaftscommunity besteht Konsens, dass die Festlegung von Rollen, die Aufteilung der Einnahmen, Plattformkontrolle und das Erzeugen von Vertrauen Kernkonzepte für die Data Governance von Plattform-Ökosystemen darstellen [22]. Zum Design von guter Governance in Plattform-Ökosystemen werden sechs fundamentale Prinzipien vorgeschlagen [23]:

- Transparenz: Governance sollte jedem Stakeholder ein klares Bild vermitteln.
- Fairness: Regeln sollen gleich für alle Teilnehmergruppen gelten. Zudem sollte jeder Teilnehmer gleiche Teilnahmemöglichkeiten besitzen. Eine hohe Fairness führt zu einer höheren Teilnahmebereitschaft und mehr Innovationen.
- Einfachheit: Governance sollte so einfach und effizient wie möglich eingesetzt werden.

- Realitätsbezug: Governance-Methoden sollten für alle Teilnehmer anwendbar und auf den jeweiligen Kontext bezogen sein.
- Shared value: Die erstellten Regeln sollten einen Mehrwert für alle Teilnehmer bieten.
- Teilhabe: Alle Teilnehmergruppen sollten die Möglichkeit zur Einflussnahme besitzen.

So wird ebenso deutlich, dass Plattform-Governance flexibel hinsichtlich möglicher Auslegungen agieren und sich an dem jeweiligen Kontext der Plattform ausrichten muss.

Governance-Entscheidungen mit Bezug zu Datenschutz und Datensicherheit in Plattform-Ökosystemen können in vier übergeordnete Entscheidungsfelder kategorisiert werden. In den Bereich der regulatorischen Umgebung fallen die Entscheidungen, welche Regularien, Standards, Regeln und Richtlinien berücksichtigt werden sollen und welche externen Regularien einen Einfluss auf das Ökosystem besitzen. Dazu werden die externen Regularien identifiziert und für die Plattform umgesetzt sowie weitere interne Richtlinien erstellt. Die Einhaltung der Richtlinien und Regularien sollte von einer unabhängigen Stelle geprüft werden, um Voreingenommenheit oder Interessenskonflikte zu vermeiden. Einen weiteren Bereich stellen Data Ownership und Zugangsberechtigungen dar. Dort wird definiert, wer die im Ökosystem verwendeten Daten besitzt, und wie festgelegt wird, wer auf die Daten zugreifen kann. Dazu müssen Rollen sowie entsprechende Rechte und Pflichten festgelegt werden. Im Rahmen des Bereichs Data Use Cases wird festgelegt, wie Daten genutzt werden können, ohne dass der Data Owner die Kontrolle über diese verliert. Dabei können Maßnahmen wie das Data Monitoring und Data Provenance als Mechanismen implementiert werden. Weiterhin können Use Cases dabei helfen, Rechte und Pflichten darzulegen und unerwünschte Nutzung von Daten zu vermeiden. Das vierte Entscheidungsfeld stellt die Datenbewertung dar. Dort wird definiert, wie der Wert der Daten festgelegt wird und welche Vergütung Teilnehmer erhalten, wenn sie ihre Daten im Ökosystem anzubieten. Zudem fällt unter diesen Bereich auch die Identifikation weiterer Belohnungen für Datennutzer, wie etwa ein Bewertungssystem [24].

Betrachtung von Regeln und Richtlinien

Beispielhaft für Maßnahmen der Governance wird im Folgenden der Bereich der regulatorischen Umgebung, insbesondere von Regeln und Richtlinien (engl.: Policies und Rules), erläutert. Im Rahmen dieses Dokuments werden unter Regeln und Richtlinien grundlegende Teilnahmebedingungen verstanden, die jeder Interessent zur Teilnahme am digitalen Ökosystem, respektive zur Nutzung der dort angebotenen Dienstleistungen erfüllen muss. Diese Teilnahmebedingungen beeinflussen die Offenheit der Plattform und somit die Zugänglichkeit für ihre Teilnehmer. Bei Auslegung der Richtlinien ist der Trade-off zwischen der Attraktivität für möglichst viele Teilnehmer und der Kontrolle über ebendiese und ihre Inhalte zu beachten. Eine sehr offene Plattform erlaubt einer Menge an Unternehmen Zugang und erleichtert die Entwicklung innovativer Lösungen. Andererseits kann durch eine vollkommen offene Plattform auch Wert verloren gehen, indem "Rauschen" durch unerwünschte Produkte, Teilnehmer oder Verhaltensweisen induziert wird. Die Regeln einer Plattform sind nicht statisch, sondern können sich mit der Zeit ändern. So starten Plattformen zumeist in einer relativ geschlossenen Form und öffnen sich mit der Zeit für weitere Teilnehmer und Produkte [25]. Plattformrelevante Richtlinien ergeben sich einerseits aus dem extern vorgegebenen regulatorischen Rahmenwerk. In der Europäischen Union sind dabei Gesetze wie die EU Datenschutz-Grundverordnung oder der EU Cybersecurity Act relevant. Zum anderen können Richtlinien auch von der Plattform selbst definiert werden. Mögliche Entscheidungsbereiche stellen dabei die einzuhaltenden Standards und Zertifizierungen von Teilnehmern und Services dar. Weiterhin sind auch Richtlinien für das Verhalten zwischen den Teilnehmern zu definieren.

Die Umsetzung der Zugangsbeschränkungen kann mittels standardisierter technischer Mechanismen erfolgen. Ein Beispiel dazu stellen Anwendungsprogrammierschnittstellen (API) dar. Diese erlauben neben dem standardisierten Austausch von Daten basierend auf Datenmodellen ebenso die Möglichkeit Datenzugriffe und -nutzungen zu verwalten. Zusätzlich sorgen APIs für eine gesteigerte Transparenz der Datenzugriffe, da diese aufgezeichnet und analysiert werden können [1].

3.1.5 Identitäts- und Zugriffsmanagement

Unternehmen müssen bei Nutzung und Bereitstellung von Daten und Services sicherstellen können, dass hinter der virtuellen Identität wirklich das zugehörige reale Unternehmen steht, um valide Entscheidungen hinsichtlich Zugriffs- und Nutzungsrechten zu treffen. Ein akzeptiertes Modell zur Sicherstellung der Identität fördert Vertrauen in das Ökosystem. Eine Identität besteht aus einem eindeutigen Bezeichner und weiteren die Entität beschreibenden Eigenschaften, wie etwa der Standort oder erhaltene Zertifizierungen. Zur Sicherstellung einer Identität erfolgt ein Authentifizierungsprozess. Dieser besteht aus den Schritten Identifikation und Authentifizierung, Zur Identifikation muss ein Teilnehmer angeben, wer er ist. Dies kann mittels eines Schlüssels, Benutzernamen oder einer E-Mailadresse geschehen. Authentifizierung beschreibt den Prozess, indem das System sicherstellt, dass es sich bei der Entität wirklich um die angegebene Person handelt.

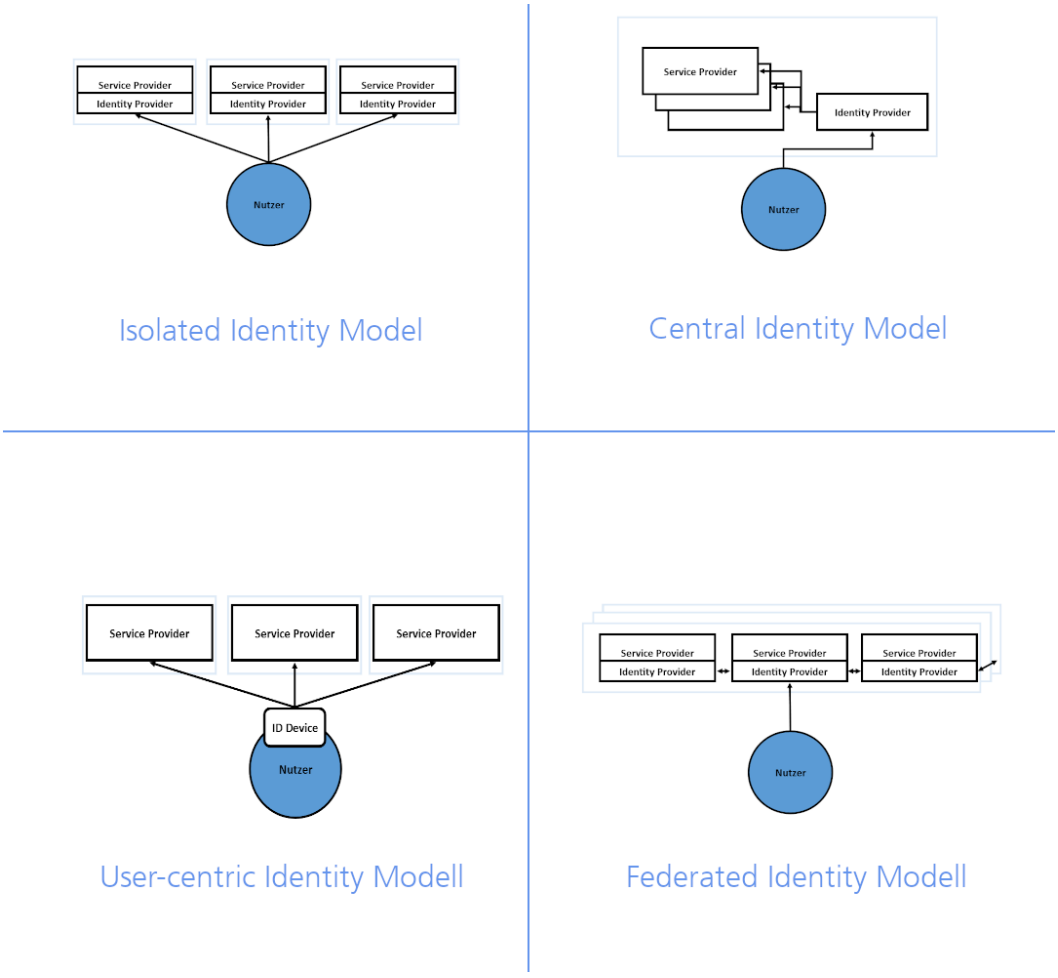


Abbildung 19: Grafische Gegenüberstellung verschiedener Identitätsmanagementmodelle

Hinsichtlich der Umsetzung von Identitätsmanagement in IIP-Ecosphere sind die in Abbildung 19 dargestellten und folgend beschriebenen Varianten denkbar. Im *isolated identity model* identifiziert und authentifiziert jeder Service Provider selbst seine Nutzer. Der Nutzer muss folglich für jeden Service Provider eine separate Registrierung durchführen. Bei simplen Registrierungsverfahren stellt dieser Prozess für den Service Provider die einfachste Form des Identitätsmanagements dar. Werden

viele Charakteristika des Nutzers abgefragt, handelt es sich auch für den Service Provider um einen aufwändiges Verfahren. Für den Nutzer ergibt sich generell ein hoher Aufwand. Einerseits muss dieser bei jeder Nutzung eines neuen Services ein Identifizierungsverfahren durchlaufen. Andererseits muss er die vielen Identitäten des Unternehmens verwalten.

Das *central identity model* wird charakterisiert durch einen zentralen Identity Provider, der die Identitäten identifiziert, authentifiziert und entlang des Lebenszyklus verwaltet. Service Provider können die Identitäten der Nutzer vom Identity Provider abrufen. Für den Nutzer erfolgt ein verringerter Aufwand durch eine einmalige Registrierung. Ein Beispiel für das zentrale Identitätsmodell stellt Kerberos dar [26].

Im *user-centric identity model* werden alle Identitätsdaten direkt durch den User, etwa mittels eines Tokens oder einer Smart Card, gespeichert. Die Identität wird dementsprechend nur an den Service Provider übertragen, falls der User sein Einverständnis dazu gibt oder eine aktive Anmeldung betreibt. Dadurch wird die Privatsphäre des Users geschützt [27].

Das *federated identity model* speichert die Identitätsdaten verteilt über verschiedene Identity Provider und Service Provider. Sie bilden dazu ein vertrauenswürdiges Netzwerk auf organisatorischer Ebene. Die Durchsetzung der Identifizierung wird durch die Plattform getätigt. Durch dieses Modell, ergibt sich ein geringer Aufwand für den Nutzer. Service Provider können die Identität eines Teilnehmers untereinander verifizieren. Ein Beispiel für ein solches Identitätsmanagement stellt Shibboleth dar [28]. Tabelle 2 fasst die Vor-und Nachteile der Modelle zum Identitätsmanagement zusammen.

Zur Umsetzung des Identitätsmanagements wird, im Kontext von Plattformökosystemen, das central identity model als angemessenste Lösung beurteilt. Es kombiniert die Provider-seitigen Vorteile einer effektiven Verwaltung von Nutzerkonten und einem geringen Aufwand für die Service Provider mit den anwenderseitigen Vorteilen im Bereich einfacher Nutzbarkeit. Bei geschlossenen Plattformen mit klaren Regeln und Richtlinien können Nutzerkonten durch eine zentrale vertrauenswürdige Instanz verwaltet werden. Zudem wird bei der Implementierung eines zentralen Identitätsmanagements von einem geringeren Aufwand im Vergleich mit den userzentrierten oder föderierten Ansätzen ausgegangen.

In enger Verbindung mit dem Identitätsmanagement steht das Zugriffsmanagement für den Datenmarktplatz und die damit verbundene Sichtbarkeit der Metadaten dar. Da die Metadaten des Datenmarktplatzes durch eine zentrale Instanz verwaltet werden, können einzelne Teilnehmer nicht mehr direkt entscheiden, wer die Metadaten ihres Datenangebots sehen kann. Unternehmen haben jedoch auch in diesem Zusammenhang den Wert der Daten erkannt und wollen Einblicke durch Mitbewerber vermeiden [29]. So wollen Unternehmen die Offenlegung der Metadaten bezüglich der Datenquellen und Aufnahmemethoden verhindern um relevante Informationen strategisch zu schützen. Zudem können Methoden der Datenbehandlung selbst wertvolle Geschäftsgeheimnisse darstellen, die den Mitbewerbern verborgen bleiben sollen [7]. Um die Sichtbarkeit der Metadaten für gewisse Teilnehmer zu begrenzen, kann die Identität des Teilnehmers genutzt werden. Dazu kann der Datenbesitzer definieren, welche Eigenschaften der potentielle Kunde besitzen soll, um von diesen Vorgaben abweichende Unternehmen eine Sichtbarkeit des Angebots einzuschränken. Die Umsetzung erfolgt anschließend durch die zentrale Instanz mittels eines Abgleichs der definierten Eigenschaften mit der Identität.

Tabelle 2: Gegenüberstellung von Modellen zum Identitätsmanagement

Identity Management Model	Vorteile	Nachteile
Isolated Identity Model	*Einfache Umsetzung für Service Provider [27]* *Informationen werden nur an Service Provider gesendet [26]*	*Nutzer müssen hohe Menge an Identitäten verwalten [26]* *Werden Passwörter zu hoch sensiblen Anwendungen vergessen, besteht ein hoher Aufwand zu deren Wiederherstellung [26]*
Central Identity Model	*Effektiv bei der Verwaltung vieler Nutzer [26]* *Nutzer benötigen nur einen Zugang (Single-Sign-On) [26]* *Gute Nutzbarkeit in geschlossenen Netzwerken [27]*	*Datendiebstahl ermöglicht Zugang zu vielen Services [26]* *In Umgebungen ohne gemeinsame Rahmenbedingungen nicht nutzbar [27]*
User-centric Identity Model	*Single-Sign-On und hohe Datensicherheit möglich [27]*	*Aufwändige Implementierung [27]*
Federated Identity Model	*Möglichkeit des Single-Sign-On in offenen Netzwerken [27]* *Hohe Kontrolle über Nutzerkonten [26]*	*Hoher Implementierungs-aufwand [27]* *Vertrauen zwischen Identity Provider benötigt [27]* *Datenkonsistenz als große Herausforderung [27]*

3.2 Unternehmensinterne Ansätze

In diesem Abschnitt werden Lösungsansätze vorgestellt, die ein jedes Unternehmen intern und ohne weitgreifende Kommunikation mit anderen Teilnehmern des Ökosystems implementieren kann, um für den Schutz und die Sicherheit der im Ökosystem verwendeten Daten zu sorgen. Dazu werden u.a. die Einschätzung des Risikos für die Freigabe von Daten für das Ökosystem, die technischen Maßnahmen zur Behandlung personenbezogener Daten oder weitere Maßnahmen der unternehmensinternen Data Governance beschrieben.

3.2.1 Behandlung von personenbezogenen Daten

Als personenbezogene Daten werden gemäß Datenschutz-Grundverordnung (DSGVO), Art. 4 Daten verstanden, die eine natürliche Person identifizieren oder identifizierbar machen. Beispiele für unmittelbar personenbezogene Daten stellen etwa der Name, die Mitarbeiternummer, die Anschrift oder die Telefonnummer dar. Mit steigender Verbreitung von Sensoren und Analysemöglichkeiten steigt auch die Identifizierbarkeit einer vermeintlich anonymisierten natürlichen Person. Diese ist gegeben, wenn mittels Aufbringung weiterer Mittel oder der Mittel Dritter eine Identifizierung möglich ist. Dies kann beispielsweise durch die Zusammenführung von Informationen geschehen. Sollten etwa

Dienstpläne der Arbeitnehmer in der Produktion mit den Einsatzorten und den Rückmeldedaten der Maschinen kombiniert werden, ist es möglich, die individuelle Leistung eines Mitarbeiters zu bestimmen und zu bewerten [30]. Insbesondere sind zudem Daten mit besonderem Schutzbedürfnis (Art. 9 Nr. 1 DSGVO) zu betrachten. Dazu gehören Daten aus den Bereichen von ethnischer Herkunft, aber auch Gesundheitsdaten und biometrische Daten. Im Bereich der individualisierten Produktion nehmen insbesondere letztere etwa im Bereich der Produktion von Implantaten oder anderen personalisierten Hilfsmitteln eine große Rolle ein. Für die Verarbeitung und Weitergabe von personenbezogenen Daten müssen die Regeln der DSGVO beachtet werden. Für Daten ohne Personenbezug gelten diese Regeln nicht. Nicht-personenbezogene Daten können ohne Befristung und Zweckgenehmigung verarbeitet werden. Zudem ist es erlaubt, Daten ohne Personenbezug durch Dritte verarbeiten zu lassen, ohne dass durch das Datenschutzrecht eine Rechtsgrundlage oder ein Vertrag zur Auftragsdatenverarbeitung gefordert wird. Insgesamt wird deutlich, dass verhindert werden muss, dass durch einen in Datenökosystemen durchgeführten Datenaustausch eine Identifizierung natürlicher Personen und ein Eingriff in deren Persönlichkeitsrechte stattfinden kann.

Um die Identifizierung natürlicher Personen zu verhindern sind Maßnahmen der Verschlüsselung, Pseudonymisierung oder Anonymisierung denkbar. *Anonymisieren* ist gemäß Bundesdatenschutzgesetz (BDSG) "das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können“. Anonymisierte Daten lassen somit keine Ableitung von Einzelangaben zu natürlichen Personen zu. Neben den Daten, die einer Anonymisierung unterzogen wurden, sind auch von vornherein erfasste Daten ohne Personenbezug anonym.

Unter *Pseudonymisierung* versteht das BDSG das "Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren." Pseudonyme ersetzen somit die Identität einer Person oder die Person identifizierende Informationen. Dies basiert auf Grundlage einer Zuordnungsregel. Bei Kenntnis dieser Zuordnungsregeln kann der Personenbezug über die Daten wiederhergestellt werden. Dies stellt den Unterschied zur Anonymisierung dar, bei der ein Personenbezug der Daten dauerhaft unterbunden wird. Verfügt eine Stelle über die Zuordnungsregel, besitzen die pseudonymisierten Daten für diese Stelle einen Personenbezug. Dementsprechend ist es wichtig, dass die Zuordnungsregeln von Pseudonymen und Ursprungsregeln vor dem Zugriff Unberechtigter geschützt werden. Im Laufe der Zeit können Pseudonyme in etwa durch eine Anreicherung des Datenbestands abgeschwächt werden [31].

Mittels Verfahren der *Verschlüsselung* werden in Klartext angegebene Zeichenfolgen, in etwa in einer Datenbank, mithilfe eines Schlüssels in eine für Dritte unverständliche Zeichenfolge gebracht. Ist ein Datensatz ausreichend verschlüsselt, so kann ein unberechtigter zwar die Chifrate lesen, ist allerdings nicht in der Lage den Inhalt zu erschließen. Für die Inhaber der Schlüssel handelt es sich weiterhin um personenbezogene Daten. Werden einzelne Merkmale verschlüsselt, können Unberechtigte etwa durch statistische Analysen Einzelangaben und identifizierende Merkmale rekonstruieren. Wird der gesamte Datensatz verschlüsselt, so sind keine Schlussfolgerungen für Unberechtigte mehr möglich. Dabei sind alle Werte geschützt und – im Gegensatz zur Pseudonymisierung – ist auch durch eine zukünftige Anreicherung des Datenbestands keine Identifikation möglich. Bei der Datenweitergabe liegen für den Dritten keine personenbezogenen Daten vor [31]. Bei der Anwendung von Verschlüsselung wird grundsätzlich zwischen Methoden der symmetrischen Verschlüsselung und Methoden der asymmetrischen Verschlüsselung unterschieden. Bei der symmetrischen Verschlüsselung, auch Geheimer-Schlüssel-Kryptographie genannt, wird nur ein Schlüssel, sowohl für die Ver- als auch die Entschlüsselung, verwendet oder es ist möglich, den Schlüssel zur Entschlüsselung

leicht aus dem Verschlüsselungsschlüssel zu berechnen. Beispiele für symmetrische Verschlüsselungsverfahren sind etwa der Data Encryption Standard (DES) oder der Advanced Encryption Standard (AES). Im Gegensatz dazu werden bei der asymmetrischen Verschlüsselung, auch Öffentlicher-Schlüssel-Kryptographie genannt, zwei zueinander passende Schlüssel erzeugt, wobei ein Schlüssel zur Verschlüsselung und ein anderer zur Entschlüsselung der Daten genutzt wird. Der Schlüssel zur Chiffrierung der Daten ist dabei öffentlich verfügbar. Einzig die Besitzer des privaten Schlüssels zur Dechiffrierung der Zeichenfolgen können allerdings die Informationen auslesen. Beispiele für asymmetrische Verschlüsselungsverfahen sind RSA, benannt nach seinen Entwicklern und Elliptische-Kurven-Verschlüsselung (ECC) [32].

Zur weiteren Erläuterung von Methoden des Datenschutzes werden im Folgenden formale Datenschutzmodelle, sowie weitere Methoden zur Erreichung von Datenschutz dargestellt.

K-anonymity stellt ein formales Datenschutzmodell zur Verhinderung der Identifizierung von Personen durch die Zusammenführung von Informationen beim Austausch von Daten mit vormaligem Personenbezug dar. Daten besitzen dabei einen k-anonymity-Schutz, wenn es innerhalb eines Datensatzes mindestens einen weiteren Datensatz mit gleichen personenbezogenen Merkmalen gibt. So entstehen Äquivalenzklassen mit k Tupeln. Zur Anwendung von k-anonymity müssen zunächst alle quasi-Identifier ermittelt werden. Dabei handelt es sich um Attribute, die bei einer Zusammenführung des Datensatzes mit weiteren Informationen dafür sorgen könnten, dass eine natürliche Person identifiziert wird. Im weiteren Verlauf wird der Datensatz so abstrahiert, dass mindestens k Datentupel hinsichtlich der quasi-Identifier dieselben Werte aufweisen [33]. K-anonymity ist aufgrund der Einfachheit des Konzepts sowie der einfachen Überführung in Algorithmen ein weit verbreitetes Konzept zum Schutz von Personendaten. Es kann effektiv vor der Enthüllung einer Identität schützen. Allerdings verbleiben bei Nutzung des Konzepts die Gefahr, dass Angreifer mittels gewisser Methoden die Eigenschaften einzelner Personen aufdecken können. So wird bei einer Homogenitätsattacke ausgenutzt, dass alle Tupel in einer Äquivalenzklasse dasselbe sensible Attribut aufweisen. So kann eine Person bei Bekanntheit gewisser Eigenschaften einer Äquivalenzklasse zugeordnet und die sensible Eigenschaft identifiziert werden. Weiterhin sind Attacken basierend auf dem Hintergrundwissen über eine Person möglich. Kann die Person einer Äquivalenzklasse zugeordnet werden, so ist es möglich, dass durch Kenntnis weiterer Eigenschaften der Person die sensible Eigenschaft identifiziert werden kann [34].

L-diversity adressiert die zuvor genannten Probleme des Konzepts von k-anonymity. L-diversity erweitert k-anonymity dahin, dass innerhalb einer Äquivalenzklasse mit k Tupeln nicht nur ein sensibles (personenbezogenes) Attribut, sondern mindestens l andere "gut repräsentierte" sensible Werte für ein Attribut existieren, damit das sensible Attribut einer spezifischen Person nicht mehr identifiziert werden kann. Ein Angreifer müsste demnach l-1 andere Werte ausschließen können, um die Eigenschaft der Person zu identifizieren. Dabei gilt: besitzen alle Äquivalenzklassen eines Datensatzes l-diversity, so besitzt auch der Datensatz l-diversity. Zur Definition eines "gut repräsentierten" Attributs stellen die Autoren fünf verschiedene Instanzen von l-diversity vor [34], die in Auswahl folgend umrissen werden. Die einfachste Instanz geht davon aus, dass in einer Äquivalenzklasse eine Anzahl von l unterschiedlichen Werten für ein Attribut vorhanden sind. Ist dabei allerdings ein Wert deutlich stärker repräsentiert als die anderen, so kann mit hoher Wahrscheinlichkeit diese Ausprägung für die Person angenommen werden. Es sind also folglich Attacken auf Basis von Wahrscheinlichkeitsverteilungen möglich. Als strengere Instanz wird daher die rekursive (c,l)-diversity definiert. So besitzt eine Äquivalenzklasse (c,l)-diversity falls der häufigste Wert weniger als c-mal der Summe der anderen Werte in einer Äquivalenzklasse auftritt. Insgesamt werden wie Verfahren der l-diversity als Weiterentwicklung der k-anonymity akzeptiert. Allerdings wird auch Kritik an ihnen geäußert. So ist es vor allem bei binären Attributen schwierig, l-diversity zu erreichen.

Zudem können durch den Vergleich einer Äquivalenzklasse mit der Gesamtpopulation Eigenschaften von Personen aufgedeckt werden [35].

Auf Basis der Limitationen von k-anonymity und l-diversity wurde die Methode *t-closeness* entwickelt. Insbesondere ist das Ziel der Methode zu verhindern, dass aus der Gesamtheit der Daten Aussagen über eine einzelne Person möglich sind. T-closeness fordert daher, dass alle Äquivalenzklassen hinsichtlich der sensiblen Attribute eine ähnliche Verteilung wie die Grundgesamtheit aufweisen. Schlussfolgerungen auf die Eigenschaften einzelner Personen sind dementsprechend schwieriger durchzuführen. Als komplexeste Methode kann ein Schutz einzelner Personen vor der Identifizierung ihrer Eigenschaften durch t-closeness gewährleistet werden. Allerdings ist t-closeness im Vergleich mit den zuvor genannten Methoden am komplexesten. Eine ähnliche Verteilung der sensiblen Attribute innerhalb der Referenzklassen im Vergleich zur Grundgesamtheit ist schwierig zu erreichen und erfordert einen hohen Aufwand [36].

Eine weitere Methode zur Vermeidung von Aussagen über einzelne Datensubjekte stellt *Differential Privacy* dar. Ansätze von Differential Privacy können sowohl Datenbanken als auch Datenstreams vor der Enthüllung von Eigenschaften natürlicher Personen schützen. Differential Privacy Verfahren können dabei sowohl bei der Aufzeichnung von Informationen, als auch bei der Veröffentlichung statistischer Datensätze angewandt werden. Die Daten werden hierbei durch die Verwendung verschiedener Algorithmen (u.a. Gauss, Laplace) mit einem zufälligen Rauschen versehen. Einzelne Datentupel werden dabei in einem Maße verändert, dass auch bei Verwendung von Datenanalysemethoden keine Rückschlüsse auf eine Person mehr möglich sind. Die Aussage des Datensatzes bleibt dabei allerdings gleich, da die statistische Verteilung bei Anwendung der Methode unverändert bleibt [37]. Im Gegensatz zu den zuvor genannten Verfahren kann bei Differential Privacy sichergestellt werden, dass auch in Zukunft keine Identifikation der Eigenschaften natürlicher Personen möglich ist. Je nach Sicherheitserfordernis können verschiedene Algorithmen ausgewählt werden. Weiterhin führt diese Methode nur zu einem geringen zusätzlichen Rechenaufwand. Allerdings ist die Anwendung von Differential Privacy nur für ausreichend große Datensätze möglich, da sonst Informationen verfälscht oder verloren gehen können. Weiterhin besteht ein hoher Aufwand zur Auswahl und zur Implementierung eines geeigneten Algorithmus unter der Berücksichtigung des Trade-offs von Datenschutz und Informationsgehalt des Datensatzes [38].

Homomorphe Verschlüsselung verfolgt explizit das Ziel die Möglichkeit der Bearbeitung von Daten auf Cloud Systemen und Datensicherheit zu vereinen. Bei der Anwendung von homomorpher Verschlüsselung werden die Daten so verschlüsselt, dass weiterhin Rechnungen mit diesen durchgeführt werden können. Im Kontext von Cloud-Services ist dazu folgendes Szenario denkbar: Der Cloud-Nutzer verschlüsselt seine Daten mittels einer homomorphen Methode und überträgt diese in die Cloud. Dort werden Datenanalysen auf den verschlüsselten Daten ausgeführt und das verschlüsselte Ergebnis zurück an den Nutzer übertragen. Dieser ist nun in der Lage das Ergebnis zu entschlüsseln und somit in Klartext zu sichten. Vollständig homomorphe Verfahren erlauben dabei sämtliche Rechenoperationen mit den Daten durchzuführen. Neben Problemen der deskriptiven Statistik können etwa auch Bild- oder Audiodaten auf diesem Weg verschlüsselt und analysiert werden. Allerdings handelt es sich hierbei noch nicht um einen ausgereiften Stand der Technik. Vielmehr existiert erweiterter Forschungsbedarf die Limitationen zu reduzieren. Auch wenn eine Reihe an Untersuchungen zur Verbesserung der Performance stattgefunden hat, existieren weiterhin Probleme bei der Performance der Verfahren, insbesondere bedingt durch einen hohen Rechenaufwand. Zudem können bei mehrfacher Multiplikation von Werten Ungenauigkeiten entstehen [39].

Eine Voraussetzung zur Anwendung der zuvor präsentierten Maßnahmen zur Verhinderung der Identifizierung natürlicher Personen ist die Feststellung, welche Datenfelder eine natürliche Person identifizierbar machen. Die manuelle Identifizierung personenbezogener Merkmale in Daten kann

dabei, je nach Art der vorliegenden Daten, mit einem hohen Aufwand verbunden sein. Die *automatisierte Detektion von personenbezogenen Daten* kann dementsprechend dabei helfen diesen Prozess schneller zu gestalten und das notwendige Fachwissen zu verringern. Die Identifikation von Merkmalen mit Personenbezug kann dabei basierend auf regulären Ausdrücken oder mittels Machine Learning erfolgen. Machine Learning bietet dabei mehr Flexibilität, eine höhere Robustheit (z.B. bei Rechtschreibfehlern) und eine einfachere Anpassung an den vorliegenden Anwendungsfall. Für wohlgeformte Datensätze sind jedoch deterministische Ansätze von Vorteil. Diese bieten zudem die Möglichkeit, verwandte Wörter mithilfe von semantischen Graphen aufzudecken [40].

3.2.2 Risikobewertung von Daten

Firmen haben Vorbehalte, durch den Austausch von Daten kritisches Wissen und Geschäftsgeheimnisse zu verraten und ihre Wettbewerbsfähigkeit zu verlieren. Allerdings wird selten eine objektive Beurteilung des Risikos eines Datenaustausches durchgeführt. Stattdessen agieren Unternehmen oftmals nach dem Motto „better safe than sorry“. Unternehmen verweigern sich somit von vornherein einer Datenweitergabe. Dadurch verbleiben die Chancen eines unternehmensübergreifenden Datenaustausches ungenutzt [41]. Dies liegt vor allem an den Schwierigkeiten bei der Anwendung von Risikomanagement-Ansätzen. So fehlt es beispielsweise an einem gemeinsamen Verständnis hinsichtlich der Kategorisierung und Bewertung von möglichen Bedrohungen. Gleichermaßen werden Risikomanagement-Methoden zur grundsätzlichen Vermeidung von Risiko anstatt als Mittel zum Management von Unsicherheiten verwendet [1]. Insbesondere vor dem Hintergrund des unvermeidbaren Wandels zum vermehrten Datenaustausch mit anderen Unternehmen ist es jedoch wichtig, Risikomanagement-Methoden ordnungsgemäß zu nutzen, um ein optimales Verhältnis zwischen Ertrag und Bedrohung für das eigene Unternehmen zu finden. Risikobewertung betrifft dabei nicht nur die Speicherung der Daten und ihren Schutz, sondern beeinflusst auch Maßnahmen der Governance, Compliance und Datensouveränität eines Unternehmens. Da derzeit vor allem Methoden des innerbetrieblichen Risikomanagements in der Praxis verwendet werden, ist es nötig, diese auf den Datenaustausch zwischen Unternehmen zu übertragen.

Das zentrale Element des Risikomanagements von Daten stellt die Risikoanalyse dar. Sie identifiziert und analysiert das Risiko des Einsatzes von Informationstechnologie. Dieses Risiko kann daraufhin akzeptiert, kontrolliert oder minimiert werden. Ein Risiko beschreibt dabei stets die Wahrscheinlichkeit des Auftretens eines Problems, das eine unerwünschte Auswirkung auf das System besitzt. Risiken können dabei zumeist nicht vollständig ausgeschlossen werden, da es sich bei der Risikoanalyse stets um eine Abschätzung handelt. Abbildung 20 zeigt das Schema einer Risikoanalyse. Eine Risikoanalyse umfasst demnach folgende Teilbereiche:

- Feststellung des Wertes einer Ressource entlang ihres Lebenszyklus
- Identifikation von Bedrohungen
- Feststellung der Wahrscheinlichkeit des Auftretens einer Bedrohung
- Feststellung der Konsequenzen eines möglichen Datenverlustes
- Analyse möglicher Schutzmaßnahmen hinsichtlich deren Effektivität

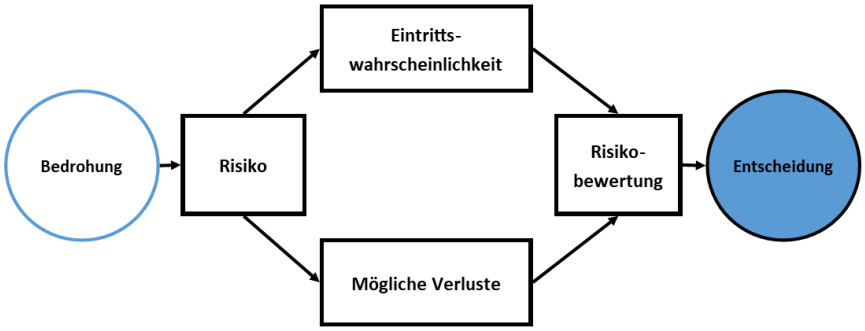


Abbildung 20: Ablauf einer Risikoanalyse

Folgend werden traditionelle Ansätze des Risikomanagements für Daten vorgestellt, um mögliche Implikationen für das Risikomanagement für den Datenaustausch abzuleiten. Die verfügbaren Methoden der Risikoanalyse werden in quantitative und qualitative Methoden unterteilt. Bei der Verwendung von quantitativen Methoden wird das Risiko anhand numerischer Werte bemessen. Diese können etwa Mengen, Frequenzen oder Verlustwerte darstellen. Die Ergebnisse dieser Werte können anhand numerischer Indikatoren dargestellt werden. Qualitative Methoden stellen ihre Ergebnisse in Form von Beschreibungen oder Empfehlungen dar. Einzelne Dimensionen können etwa anhand von Ordinalskalen gekennzeichnet werden oder Angriffsszenarien erstellt werden [42].

Tabelle 3 zeigt die Gegenüberstellung der Typen von Risikoanalysen sowie ausgewählte Beispielmethoden. Diese traditionellen Methoden wie NIST SP800-30 gehen davon aus, dass sich die Datenbestände einer Organisation innerhalb des eigenen Datencenters befinden und die Organisation diese selbst vollständig verwalten sowie die Sicherheitsprozesse selbst bestimmen kann. Werden Daten innerhalb eines Ökosystems gehandelt oder genutzt, sind diese Annahmen – äquivalent zur Nutzung von Cloud-Computing Diensten – nicht mehr gültig. Das Gegenteilige ist der Fall. Oftmals mangelt es an Informationen hinsichtlich der sicherheitsrelevanten Technologien des Geschäftspartners. Dementsprechend gilt es, die existierenden Modelle zur Risikoanalyse auf den Kontext von Datenökosystemen anzupassen oder neuartige Methoden zur Risikoanalyse zu entwickeln.

Im Rahmen des Datensharings und der Datennutzung in Datenökosystemen ergeben sich drei Arten von Risiken, die im Folgenden getrennt betrachtet werden. Auf der Seite des Datenowner sind dies einerseits das Risiko eines Unternehmens durch die Weitergabe von Daten an Dritte sensible Informationen preiszugeben oder zu verlieren. Andererseits existieren für den Datenowner mit der Nutzung von Services auf den etablierten Plattformen verbundene Risiken. Auf der Seite des Datenempfängers können bei der Übernahme von Informationen Dritter Verfügbarkeitsausfälle entstehen, die negative Effekte auf die Geschäftsprozesse eines Unternehmens besitzen.

Tabelle 3: Zusammenfassung der Risikoanalysetypen nach [42]

Typ der Risikoanalyse	Qualitative Methoden	Quantitative Methoden
Vorteile	*Vermittlung eines genaueren Bildes über das Risiko* *Durch numerische Werte wird die Entscheidungsfindung erleichtert*	*Einfache Durchführung* *Bereiche erhöhten Risikos werden schnell identifiziert* *Risiken können priorisiert werden*
Nachteile	*Teilweise ungenaue Bewertung* *Numerische Methoden müssen um qualitative Methoden ergänzt werden um Handlungsempfehlungen zu erhalten* *Hoher Aufwand und mehr Erfahrung benötigt*	*Keine Einordnung des Risikos mithilfe von Zahlen* *Schwierige Kosten-Nutzen-Analyse* *Nur ungefähre Aussagen möglich*
Beispiele	*ALE* *Courtney Method* *ISRAM Modell*	*FMEA* *FMECA* *NIST SP800-30* *CRAMM*

Zur Bewertung des Risikos bei Nutzung von Services in können bereits existierende Methoden zur Analyse der Risiken von Cloud-Services genutzt werden. Im Gegensatz zum Risikomanagement von Datenweitergabe und Datenhandel ist, aufgrund der durch Sicherheitsrisiken bedingten großen Wiederstände zur Adoption von Cloud Dienstleistungen, bereits eine Reihe an Vorgehensmodellen zur Risikoanalyse bei Cloud-Services vorhanden. Deren Resultate werden von Organisationen als Orientierungshilfe zur Etablierung von Schutzmaßnahmen für die das Unternehmen verlassenden Daten genutzt. Ein Vergleich von Methoden zur Risikoanalyse von Cloud Anwendungen wird bereits an anderer Stelle durchgeführt [43]. Eine Lösung zur Verhinderung jeglichen Risikos stellt die Bereitstellung der Services zum Einsatz in der Edge-Cloud dar.

Während die dort beschriebenen Methoden wie etwa OCTAVE oder CCRAMM das Risiko aus der Perspektive des Cloud-Nutzers beschreiben, sind für die Analyse des Risikos bei der Bereitstellung von Services, wie etwa der Weitergabe von Daten an Dritte nur geringe Referenzen vorhanden. Ein Zusammenhang, bei dem das Risiko der Freigabe von Informationen an Dritte untersucht wird, stellt die sogenannte risikobasierte Zugangskontrolle dar. Dort wird, basierend auf einer vorherigen Analyse des Risikos der Daten- oder Dienstfreigabe an einen Dritten, der Zugriff für diesen Dritten auf die Ressource genehmigt oder verweigert. In ihrer strukturierten Literaturanalyse stellen Atlam et al. [44] relevante Faktoren zur Abschätzung des Risikos bei der Freigabe von Ressourcen vor. Die wichtigsten Faktoren sind dabei die Risikohistorie des potentiellen Nutzers, die Sensitivität der Ressource sowie der Nutzungskontext. Im Rahmen ihrer Untersuchung stellen sie ebenso verschiedene genutzte

Techniken zur Abschätzung des Risikos dar. Dabei wird deutlich, dass die Abschätzung des keinesfalls eine triviale Aufgabe darstellt. Beispielsweise führen ungenaue oder unvollständige Informationen zu Problemen bei der Ermittlung ihres Wertes. Dementsprechend gilt es im Rahmen von IIP-Ecosphere, die notwendigen Informationen zur Abschätzung des Risikos bereitzustellen. Da die risikobasierte Zugangskontrolle automatisch erfolgen soll, werden zur Risikobewertung zumeist quantitative Methoden verwendet. Diese basieren etwa auf mathematischen Gleichungen oder Fuzzylogik. Aber auch traditionelle Verfahren der Risikoanalyse finden Verwendung. So können etwa einzelne Methoden zur Risikoabschätzung im Rahmenwerk von NIST SP800-30 eingesetzt werden [45].

Risiken beim Datenbezug können auf Basis der Nutzung eines Datenkatalogs mitsamt entsprechendem Metadatenmodell analysiert werden. Dazu wird zunächst die Etablierung eines Metadatenkatalogs benötigt, der den unternehmensübergreifenden Austausch von Metadaten im Datenökosystem ermöglicht. Dieser Metadatenkatalog muss zudem ein geeignetes Metadatenmodell zur Beschreibung der für eine Risikoabschätzung relevanten Informationen bieten. Ein solches Datenmodell stellt das Metadata Model for Data Goods (M4DG) dar. Dieses bietet die Möglichkeit im Unternehmen vorhandene Daten im Sinne eines Wirtschaftsgutes und Datenressourcen in Datennetzwerken zu beschreiben. So können etwa Relationen zwischen Datenressourcen und Nutzern dargestellt werden. Ein auf Basis von M4DG implementierter Datenkatalog ermöglicht es dem Datenowner Metadaten mit Relevanz für eine Risikobewertung zur Verfügung zu stellen. Aufgrund der unterschiedlichen Anwendungsfälle der Datennutzung und der Präferenzen eines jeden Unternehmens, werden die Unternehmen selbst befähigt, eigene Metriken zur Erfassung des Risikos zu implementieren. Es wird die Implementierung von Metriken mit textuellen und numerischen Ergebnissen ermöglicht. Hinsichtlich der Architektur des Datenkatalogs sind sowohl zentrale als auch dezentrale Strukturen denkbar. Durch eine mögliche Skalierung und Teilautomatisierung der Metriken ist es zudem möglich, das Datenangebot verschiedener Teilnehmer hinsichtlich des Risikos zu bewerten und darauf basierend eine Auswahl der genutzten Daten zu treffen [29].

3.2.3 Governance-Mechanismen

Basierend auf den dargestellten Entscheidungsdimensionen der Plattform Governance ist es möglich, ein initiales Framework für die Data Governance für einzelne Teilnehmer in Plattform-Ökosystemen abzuleiten [46]. Abbildung 21 fasst die aus dem etablierten Kontext von unternehmensinterner Data Governance abgeleiteten Handlungsfelder für Data Governance in Datenökosystemen zusammen.

Der Bereich der originalen Datenqualität beschreibt das Handlungsfeld, Daten von hoher Qualität in das Datenökosystem einzubringen. Dies ist zunächst identisch mit dem Handlungsfeld von hoher Datenqualität bei unternehmensinterner Data Governance. Die Bereitstellung qualitativ hochwertiger Metadaten ist für den Datenaustausch besonders relevant, da diese die Eigenschaften des Datensatzes für den Tauschpartner beschreiben. Verantwortlich für die Gewährleistung einer ausreichenden Qualität der Daten sind dabei die Unternehmen selbst, allerdings können Datenmodelle zum Austausch von Metadaten von der zentralen Instanz vorgegeben werden. Im Rahmen des Data Ownership wird definiert, wer welche Daten besitzt und weiterhin, wer die Berechtigung besitzt, auf diese Daten zuzugreifen. Im Bereich Data Stewardship soll gewährleistet werden, dass die Daten und Datenmodelle dem Verwendungszweck entsprechen. Bei Untersuchung des Datenlebenszyklus erfolgt die Festlegung, wer welche Daten zu welchem Zeitpunkt teilen oder abrufen kann und wann das Teilen der Daten beendet wird. Sollte kein Datenmodell für Metadaten vorgegeben sein, gilt es außerdem, ein Format für die Metadaten festzulegen, dass Interoperabilität zwischen den Teilnehmern fördert und eine Vergleichbarkeit der Daten gewährleistet. Im Bereich Datenqualität der Plattform wird adressiert, wie das Unternehmen die Integrität, Verfügbarkeit, Authentizität und Sicherheit der Daten auf der digitalen Plattform unterstützen kann. Weiterhin müssen - gemäß des Bereichs Datennutzung und Bewertung adäquate Geschäftsmodelle innerhalb der Organisation etabliert und die Wertschöpfung unter den Teilnehmern aufgeteilt werden. Im Gegensatz zur unternehmensinternen

Data Governance zeigen die Bereiche Data Ownership und Zugangsberechtigungen, Datennutzung und Bewertung und Data Stewardship die größten Unterschiede, da dort die externen und teilweise gegenläufigen Interessen mehrerer Organisationen berücksichtigt werden müssen [46].

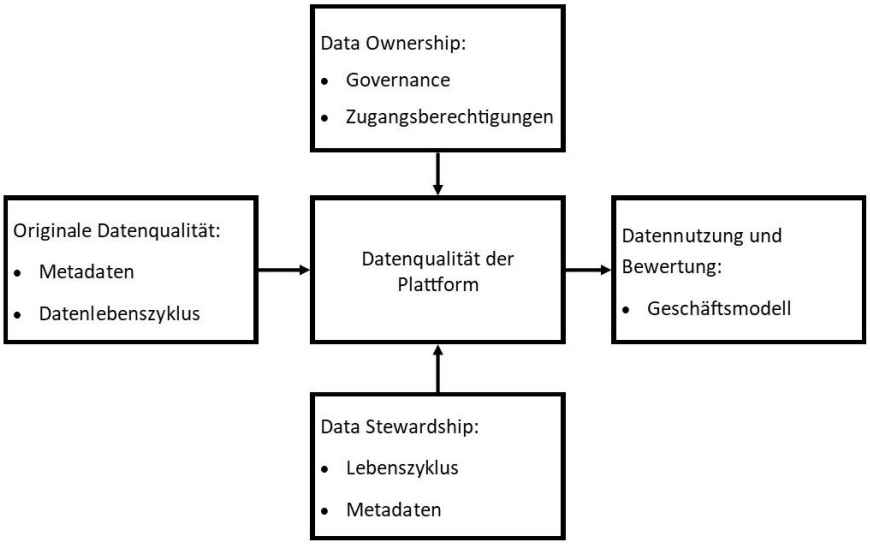


Abbildung 21: Unternehmensperspektive von Data Governance in Datenökosystemen in Anlehnung an [46]

3.2.4 Festlegung von Zugangsbedingungen zu Ressourcen

Bei der Teilnahme in Datenökosystemen ist für die Teilnehmer wichtig, zu bestimmen, welchen anderen Unternehmen Zugriff zu den von ihnen angebotenen Ressourcen gewährt werden kann. Dazu werden im Folgenden unter dem Oberbegriff Access Control die Festlegung von Entscheidungsmodellen diskutiert, die Anforderungen an potentielle Kooperationspartner, sogenannte Policies, überprüfen. Die Entwicklung von Access Control Systemen wird in drei Schritte unterteilt. Zunächst werden in der Definitionsphase die Regeln auf einer nicht-technischen Ebene festgelegt. In der Formalisierungsphase werden die Regeln in ein formales Modell überführt, das in der Implementierungsphase technisch realisiert wird. Während der Ablauf der Formalisierung und Implementierung der Zugangsbedingungen durch die übergreifend definierte technische Lösung vorgegeben wird, ist es die Aufgabe eines jeden Unternehmens die durch Dritte zu erfüllenden Bedingungen zum Zugriff auf ihre Ressource festzulegen. Dabei handelt es sich um einen schwierigen Prozess. So müssen die komplexen und oft zweideutigen Regeln der realen Welt in eindeutig definierte und von Computersystemen realisierbare Policies umgewandelt werden. Für die Definitionsphase von Access Control Policies stehen derzeit keine allgemeinen Rahmenwerke zur Verfügung. Verfügbare Standards konzentrieren sich eher auf die Implementierung der Policies anstatt Hilfestellung zu deren Entwicklung zu geben. Nützliche Entwicklungshinweise können allerdings aus dem übergeordneten Bereich der Informationssicherheitsrichtlinien abgeleitet werden. Zwar existiert auch hier weiterhin Forschungsbedarf, allerdings sind dort umfassendere Forschungsarbeiten vorhanden.

Eine Gegenüberstellung verschiedener Modelle zur Entwicklung von Policies zeigt, dass sich die notwendigen Tätigkeiten in die Schritte Input, Development und Output aufteilen lassen [47]. Eine Ausführung der Schritte kann konsekutiv, iterativ oder simultan erfolgen. Eine Übersicht des Entwicklungsprozesses ist in Abbildung 22 dargestellt.

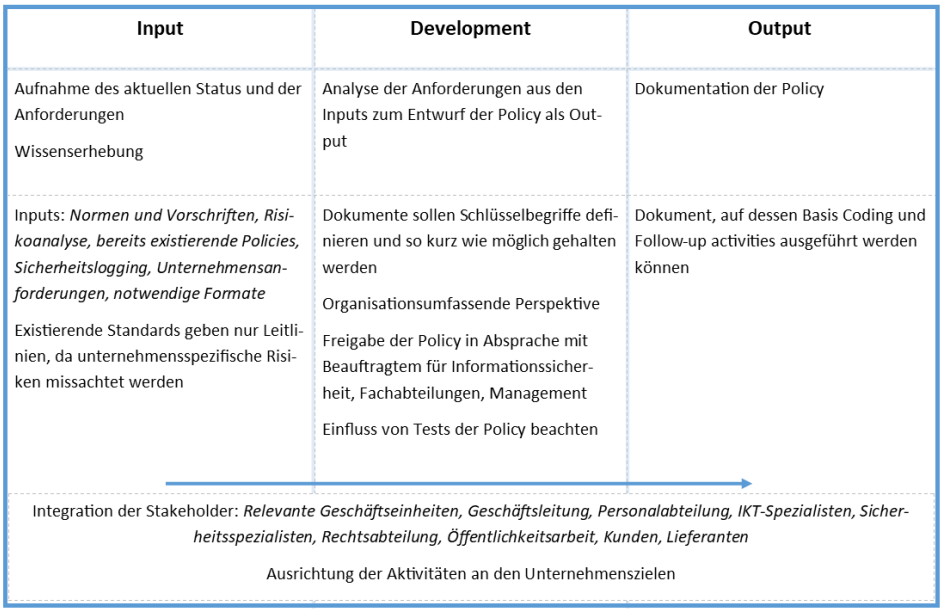


Abbildung 22: Entwicklung von Sicherheitspolicies in Unternehmen in Anlehnung an [47]

In der Inputphase werden sämtliche notwendigen Informationen zur Erstellung der Policies gesammelt. Dabei kann es sich unter anderem um externe Regularien, wie etwa die DSGVO, sowie intern vorhandene Risikoanalysen, bereits existierende Policies, Loggings von Sicherheitsapplikationen und Standarddokumente zur Beschreibung des Formats der Policies handeln. Viele Wissenschaftler warnen dabei vor der ausschließlichen Nutzung vordefinierter Standards oder Richtlinien, da diese keine unternehmensspezifischen Risiken beschreiben können und somit die Gefahr von deren Vernachlässigung besteht.

In der Entwicklungsphase werden die aus den Inputs abgeleiteten Anforderungen genutzt, um Policies zu formulieren. Dabei müssen Entscheidungen hinsichtlich des Designs, wie etwa die Architektur, das Abstraktionslevel und das Format der Policy festgelegt werden. Das Entwicklungsteam sollte dabei aus Personen bestehen, die einen Gesamtüberblick über die Vorgänge im jeweiligen Unternehmen besitzen. Hinsichtlich des eigentlichen Dokuments wird empfohlen, Schlüsselbegriffe zu definieren, da diese oftmals unterschiedlich aufgefasst werden können. Weiterhin sollten die Regeln kurz und präzise formuliert werden. Um zu prüfen, ob eine Policy ihren Anforderungen genügt, sollte diese vor ihrem produktiven Einsatz getestet werden. Dabei ist vor allem relevant, ob die Policy die Anforderungen erfüllen kann, den Vorgaben hinsichtlich des Design entspricht und ob diese technisch umgesetzt werden kann. Die Policy kann zudem weiteren Tests in den anderen Stufen der Entwicklung unterzogen werden. Die Ergebnisse der zusätzlichen Tests, etwa nach der Implementierung, müssen entsprechend iterativ in der Implementierungsphase berücksichtigt werden. Nach der Fertigstellung der Policy sollte diese zur Freigabe mit Personen aus dem Bereich der Informationssicherheit, den Fachabteilungen und dem Management diskutiert werden.

Als Output des Entwicklungsprozesses steht ein strukturiertes Dokument zur Verfügung, das die Policy beschreibt und auf dessen Basis die Umsetzung durch Programmierung oder weitere Folgeaktivitäten ausgeführt werden können. Während des gesamten Entwicklungsprozess ist es wichtig, die relevanten Stakeholder einzubinden. Diese können etwa aus den Bereichen der Geschäftsleitung, der Informationssicherheit, den Fachabteilungen sowie aus dem Kreis von Kunden oder Lieferanten entstammen. Einerseits kann durch deren Einbeziehung weitere technische Expertise gewonnen werden. Andererseits die Integration außenstehender in den Entwicklungsprozess bei der Erzeugung von Akzeptanz. Weiterhin gilt es, die Policies an den Unternehmenszielen auszurichten, um eine einheitliche Strategie zu verfolgen [47].

4 Anforderungen des Konsortiums

In Abschnitt 2 wurden mögliche Herausforderungen im Bereich Datenschutz und Datensicherheit für den Datenaustausch in Datenökosystemen mittels einer Literaturrecherche ermittelt. In Abhängigkeit davon wurden auf selben Wege in Abschnitt 3 denkbare technische und organisatorische Maßnahmen zur Behandlung der Herausforderungen vorgestellt und diskutiert. Um die Probleme und Lösungskonzepte im Bereich Datenschutz und Datensicherheit aus Sicht der Teilnehmer in IIP-Ecosphere adressieren zu können, war es notwendig die Anforderungen des Konsortiums in diesem Bereich aufzunehmen. Dazu wurde, durch den Think Tank Daten, eine Umfrage im Onlineformat durchgeführt. Um Synergieeffekte nutzen zu können, wurde der Fragebogen zusammen mit dem Bereich der rechtlichen Begleitforschung gestaltet. Insbesondere wurde dabei auf die zuvor ermittelten Problemstellungen und Lösungen Bezug genommen. Die Entwicklung des Fragebogens erfolgte in mehreren Iterationsschritten. Eine interne Validierung des Fragebogens fand mittels Methoden des lauten Denkens statt. Die finale Umfrage umfasste insgesamt etwa 60 Fragen, von denen etwa die Hälfte eine Relevanz für den Inhalt diese Whitepapers besaß, und wurde mittels des Online-Umfragetools Limesurvey implementiert. Einen großen Anteil der Befragung machten dabei Fragen aus, in denen die Teilnehmer Maßnahmen und Umstände auf einer ordinalen Skala mit sechs Abstufungen bewerten sollten.

Die Umfrage wurde den Partnern elektronisch zur Verfügung gestellt. Da die Fragestellung nicht für alle Projektpartner Relevanz besitzt, wurde innerhalb des Anschreibens an den Verteiler auf die Zielgruppe innerhalb der Projektpartner hingewiesen. Die relevanten Ansprechpartner der Assoziierten des Projekts wurden direkt adressiert.

Durch die Methodik der Befragung geben individuelle Personen Auskunft über die Eigenschaften ihrer Organisation. Da diese jedoch als Ansprechpartner ihres Unternehmens in IIP-Ecosphere zur Verfügung standen, kann von einer hohen Erfahrung der Befragten zu den besprochenen Themen und zu der durch ihn vertretenen Organisation ausgegangen werden. Zur Beantwortung des Fragebogens wurde den Partnern ein Zeitraum von 25 Tagen gewährt. Die Assoziierten hatten zwei Wochen Zeit zur Beantwortung der Fragen. Für den Fragebogen wurden insgesamt 30 relevante Unternehmen unter den Teilnehmern ausgemacht. Nach Beendigung des Fragezeitraums waren 12 Fragebögen vollständig und 19 Fragebögen teilweise beantwortet. Damit ergibt sich eine Rücklaufquote von 40 % (63 %). Die demografische Verteilung der teilgenommenen Unternehmen ist Tabelle 4 zu entnehmen.

Zur Auswertung der Daten wurden diese zunächst aus Limesurvey extrahiert. Eine initiale Datenaufbereitung und Sichtung fand anschließend in Microsoft Excel statt. Die Erstellung von Plots und weitergehenden statistischen Analysen wurden in R getätigt. Folgend werden zunächst die Ergebnisse der Untersuchung dargelegt. Im Anschluss werden aus diesen in Kombination mit dem erarbeiteten Hintergrundwissen Handlungs- und Gestaltungsempfehlungen für das Design des Ökosystems in IIP-Ecosphere getätigt.

Tabelle 4: Demografische Verteilung der Umfrageteilnehmer

Unternehmensart	Anzahl	Mitarbeiterzahl	Anzahl	Umsatz	Anzahl
Maschinenbau & Produzierende Unternehmen	*5*	*10-49*	*3*	*< 1 Mio. Euro*	3
Softwarehersteller	*5*	*50-249*	*4*	*1 Mio. Euro - 5 Mio. Euro*	2
Sonstige	*2*	*2500-5000*	*2*	*10 Mio Euro - 25 Mio. Euro*	2
		>5000	*3*	*> 500 Mio. Euro*	4

4.1 Darstellung der Umfrageergebnisse

Im Rahmen dieses Abschnitts werden die Ergebnisse der Umfrage zu Datenschutz und Datensicherheit in den Bereichen der Nutzung der zentralen Plattform, des Datensharings, der Datenhaltung und der Datennutzung dargestellt, um im folgenden Abschnitt Handlungs- und Gestaltungsempfehlungen an die Maßnahmen zu Datenschutz und Datensicherheit in IIP-Ecosphere abzuleiten. Die ebenso im Rahmen der Befragung erhobenen Anforderungen im Bereich rechtliche Begleitforschung werden an dieser Stelle nicht adressiert.

Ein Ziel der Befragung war es, die Vorstellungen der Teilnehmer hinsichtlich der zukünftigen Nutzung der KI-Services zu erfassen. Die meisten Teilnehmer prognostizieren, dass ihr Unternehmen die KI-Services der entstehenden Plattform nutzen werden. Lediglich zwei Unternehmen sehen es als eher unwahrscheinlich, dass die entstehende Plattform von ihnen genutzt wird. Die Unternehmen sind sich dabei relativ sicher, dass sie dazu die von ihnen selbst erhobenen Daten nutzen werden. Hinsichtlich der Nutzung von KI-Services mittels Daten Dritter zeigt sich ein gemischtes Bild. Während eine Hälfte der Teilnehmer ein Fremdbezug von Daten zumindest eher wahrscheinlich hält, sieht die andere Hälfte der Befragten dies für ihr Unternehmen als eher unwahrscheinlich. Auch im Hinblick auf eine mögliche Datenübertragung an eine zentrale Plattform zur Nutzung der KI-Services kann keine gemeinsame Haltung ausgemacht werden. Einerseits existieren Unternehmen innerhalb von IIP-Ecosphere mit einer sehr hohen Bereitschaft Daten an eine Plattform zur Nutzung ihrer Services zu übertragen. Andererseits steht wiederum ein Teil (5) der Unternehmen der Übertragung ihrer Daten auf eine Plattform skeptisch gegenüber.

Zur Etablierung von Maßnahmen aus dem Bereich der Datenschutz und Datensicherheit ist Kenntnis über die Eigenschaften der verwendeten Daten unerlässlich. In IIP-Ecosphere sind zehn der zwölf befragten Unternehmen grundsätzlich in der Lage, die Schützenswürdigkeit ihrer in IIP-Ecosphere genutzten Daten zu bewerten. Von Diesen erachten acht ihre Daten als schützenswert und nur zwei Unternehmen gehen nicht von einer Schützenswürdigkeit ihrer Daten aus. Wird eine Datenverarbeitung oder -analyse in Echtzeit durchgeführt, ergeben sich besondere Anfordernisse an mögliche Schutzmaßnahmen, da diese nur geringe Verzögerungen des Gesamtablaufs bedingen dürfen. Im Rahmen dieser Umfrage sagen insgesamt acht Unternehmen aus, dass sie Anwendungen mit Echtzeitbedarf in IIP-Ecosphere nutzen möchten. Die anderen vier Unternehmen nutzen entweder keine Daten in Echtzeit oder konnten an dieser Stelle keine Aussage darüber tätigen. Auch die Nutzung von Personendaten sorgt für einen erhöhten Schutzbedarf. In IIP-Ecosphere planen nur zwei der befragten Unternehmen eine Behandlung von Daten mit Personenbezug. Ein weiteres Unternehmen gibt an, derzeit kein Wissen über die Nutzung personenbezogener Daten zu besitzen.

Um die Präferenzen der Teilnehmer hinsichtlich der Anwendung von Methoden zur Behandlung von personenbezogene Daten aufzudecken, wurden diese gefragt, inwiefern die Methoden Verschlüsselung, Pseudonymisierung und Anonymisierung im Rahmen von IIP-Ecosphere denkbar wären. Abbildung 23 zeigt eine Zusammenfassung der Bewertung in Form von Boxplots. Dabei erscheint die Verschlüsselung aufgrund des geringeren Medians zunächst als bevorzugte Methode der Teilnehmer. Um diese Hypothese zu testen, wurde der Friedman-Test für die drei untersuchten Mechanismen angewandt. Dieser untersucht, ob ein systematischer Unterschied zwischen der Beliebtheit der Verfahren existiert, oder die Abweichungen der Stichprobe nur zufällig auftreten. Aufgrund des hohen p-Werts von 0,738 kann kein systematischer Unterschied festgestellt werden. Daher kann von einer ähnlichen Beliebtheit der Methoden innerhalb des IIP-Ecosphere-Konsortiums ausgegangen werden.

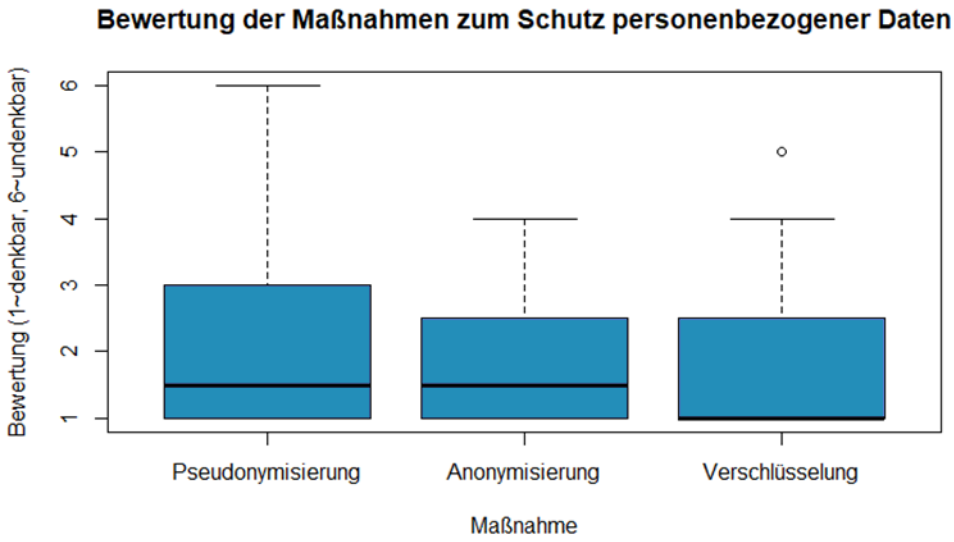


Abbildung 23: Bewertung der Maßnahmen zum Schutz personenbezogener Daten

Neben den zuvor präsentierten Maßnahmen zum Schutz der personenbezogenen Daten sind weitere, allgemeine Maßnahmen für die Etablierung von Datenschutz und Datensicherheit auf der entstehenden Plattform denkbar. Um das Meinungsbild der Teilnehmer hinsichtlich der erforderlichen Maßnahmen zum Datenschutz zu erfassen, wurden diese gefragt, für wie notwendig sie folgende Datenschutzmaßnahmen bei einer Teilnahme an IIP-Ecosphere erachten:

1. Zugriff durch Dritte verhindern
2. Garantiertes Löschen der Daten nach Verwendung
3. Ausschließliche Verwendung der Daten für den Untersuchungsgegenstand
4. Verhinderung der Speicherung Ihrer Resultate durch den Serviceanbieter
5. Verhinderung des Verlusts von Daten auf der Plattform

Abbildung 24 zeigt die Notwendigkeit der weiteren Datenschutzmaßnahmen für die Konsortialteilnehmer. Die auf der X-Achse aufgeführten Nummern entsprechen der Nummerierung in vorangegangener Aufzählung. Es wird deutlich, dass ein hoher Konsens in IIP-Ecosphere darüber herrscht, dass ein Datenzugriff durch Dritte während der Nutzung der Services zu verhindern ist. Hinsichtlich der anderen Faktoren ergibt sich ein gemischtes Meinungsbild. Zur Prüfung ob sich der Median der Bewertungen tatsächlich voneinander unterscheiden, wurde erneut ein Friedman-Test durchgeführt. Ein p-Wert von 0,002 signalisiert dabei, dass sich mindestens ein Median signifikant von einem anderen unterscheidet und somit als notwendiger angesehen wird. Durchgeführte Post-hoc-Untersuchungen mittels des Nemenyi-Test konnten allerdings nur eine höhere Notwendigkeit hinsichtlich der Verhinderung des Zugriffs durch Dritte gegenüber der Verhinderung der Speicherung der Resultate durch den Serviceanbieter ermitteln (p-Wert 0,014). Bei einem Signifikanzniveau von 0,05 konnten keine weiteren für das gesamte Konsortium representativen Präferenzen herausgestellt werden. Grundsätzlich ergibt sich jedoch für alle Maßnahmen außer der Verhinderung der Speicherung von Resultaten durch den Serviceanbieter ein hohes Erfordernis.

In diesem Kontext wurde den Teilnehmern ebenso die Möglichkeit gegeben, weitere aus ihrer Sicht notwendige Maßnahmen frei darzustellen. Dies wurde von drei Teilnehmern genutzt. Ein Teilnehmer betonte dabei die Notwendigkeit von rechtlichen Rahmenbedingungen zum Datenzugriff, die zusätzlich zu technischen Mechanismen geschaffen werden müssen. Zwei weitere Teilnehmer betonten den Wunsch nach Transparenz, insbesondere hinsichtlich des Verwendungszwecks ihrer Daten und der zur Datenbearbeitung genutzten Technologien und Dienstleister. Einer der Teilnehmer wünscht sich zudem in diesem Zusammenhang ein Vetorecht bei der Verwendung seiner Daten.

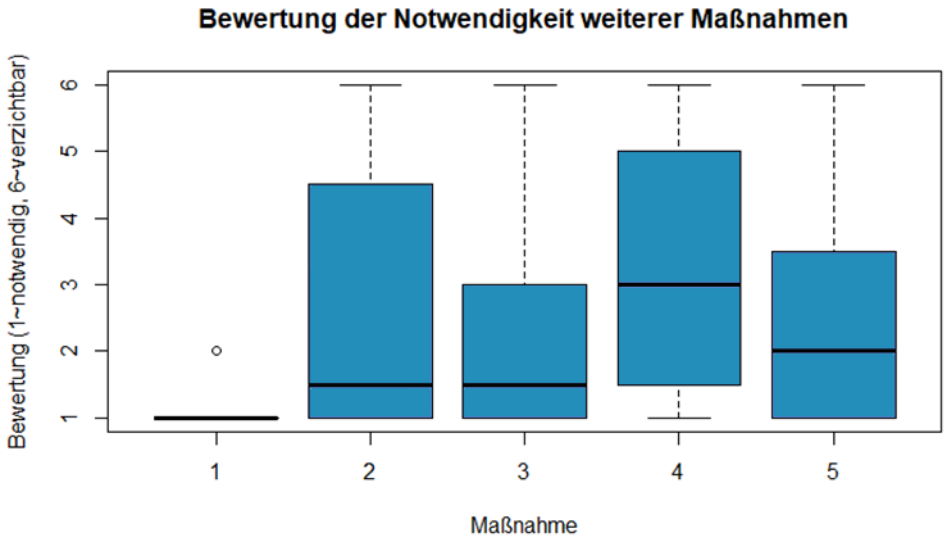


Abbildung 24: Bewertung der Notwendigkeit weiterer Datenschutzmaßnahmen

Im zweiten Fragekomplex wurden bisherige Erfahrungen der Unternehmen im Bereich des Datensharings begutachtet und die geplanten Aktionen in diesem Bereich in IIP-Ecosphere ermittelt. Aktuell bietet keines der befragten Unternehmen seine Daten Dritten (zum Kauf) an. Auch im Rahmen von IIP-Ecosphere sehen es alle Befragungsteilnehmer als eher unwahrscheinlich bis sehr unwahrscheinlich, dass Daten von ihnen anderen Unternehmen (zum Kauf) angeboten werden. Auf Seiten des Fremdbezugs von Daten zeigen sich die Befragten offener eingestellt. So Nutzen bereits etwa die Hälfte der Unternehmen Daten von Dritten. Zwei der Unternehmen geben an, diese käuflich zu erwerben. Fünf weitere geben an, frei verfügbare Daten zu nutzen. Im Rahmen von IIP-Ecosphere werden allerdings voraussichtlich nur ein Drittel der Unternehmen die Daten anderer nutzen und somit vorrangig ihre eigenen Daten zur Leistung der Services verwenden. Abbildung 25 zeigt die Relation der Unternehmen hinsichtlich des Fremddatenbezugs von Daten innerhalb und außerhalb von IIP-Ecosphere.

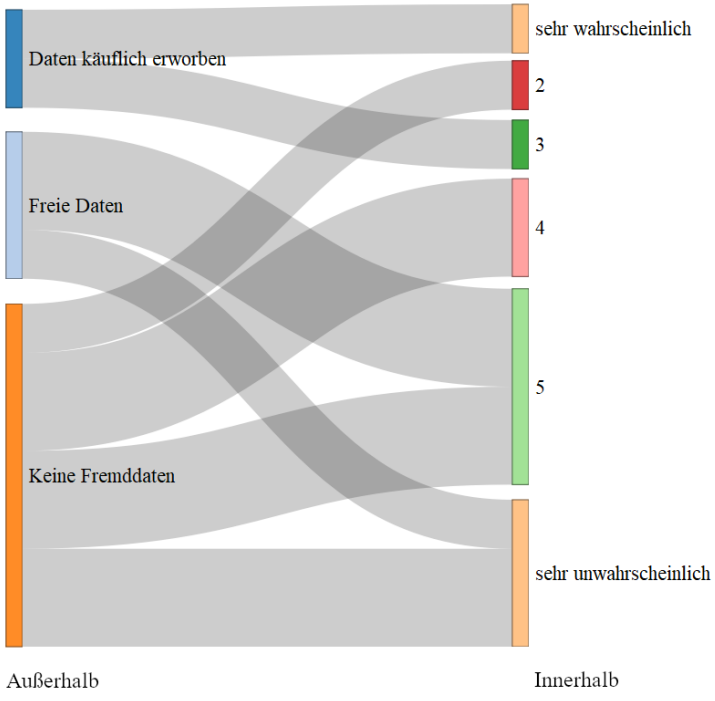


Abbildung 25: Fremdbezug von Daten außerhalb und innerhalb von IIP-Ecosphere

Ein weiterer Interessensaspekt stellt das Format der ausgetauschten Daten dar. Je nach Datenformat müssen unterschiedliche Technologien zum Schutz und zur Sicherheit der übertragenen Informationen implementiert werden. Für Bilddateien existieren andere Anforderungen als für strukturierte Daten, wie zum Beispiel Daten in tabellarischer Form. Letztere werden von den meisten befragten Unternehmen (10) in Zukunft ausgetauscht werden, während semistrukturierte (5) und unstrukturierte Daten (6) von etwa der Hälfte der Teilnehmer für einen zukünftigen Datenaustausch vorgesehen sind. Dementsprechend sollten bei der Umsetzung von Sicherheitsmaßnahmen alle Datenarten berücksichtigt werden. Weiterhin wurde, wie auch hinsichtlich der Nutzung von KI-Services, die Frage nach Nutzung von personenbezogenen und Echtzeitdaten gestellt. Hierbei wurde festgestellt, dass ein Viertel der Unternehmen planen, ihre Daten in Echtzeit an Partner zur Verfügung zu stellen. Im Gegensatz dazu geht derzeit nur eines der befragten Unternehmen von einer Übertragung von Daten mit Personenbezug an andere Stellen aus.

Um Ursachen für die geringe Bereitschaft zur Weitergabe von Daten an andere Unternehmen zu klären, wurden die Unternehmen gebeten, mögliche Gründe für eine skeptische Haltung im Bereich der Datensicherheit zu bewerten. Dabei wurden folgende Probleme berücksichtigt:

1. Preisgabe des Geschäftsgeheimnisses mittels Daten
2. Unsichere Datenübertragung
3. Sicherheit der Daten beim Nutzer
4. Missbrauch der Daten für nicht autorisierte Zwecke

Wie Abbildung 26 zeigt, bewerten die Teilnehmer insbesondere das Risiko des Missbrauchs von Daten für nicht genehmigte Zwecke und die Preisgabe von Geschäftsgeheimnissen als hoch. Auch in Bezug auf die die Gesamtheit aller Teilnehmer kann durch statistische Tests gezeigt werden, dass die beiden genannten Probleme als Hauptursachen für eine Zurückhaltung beim Austausch von Daten innerhalb der Befragten gelten. Zusätzlich zu den zur Bewertung vorgegebenen Maßnahmen konnten weitere subjektive Probleme durch die Umfrageteilnehmer genannt werden. Unter anderem sieht es ein Teilnehmer als kritisch an, dass von seinem Unternehmen weitergegebene Daten durch den Dritten falsch interpretiert werden könnten und sich somit Nachteile für sein Unternehmen ergeben könnten.

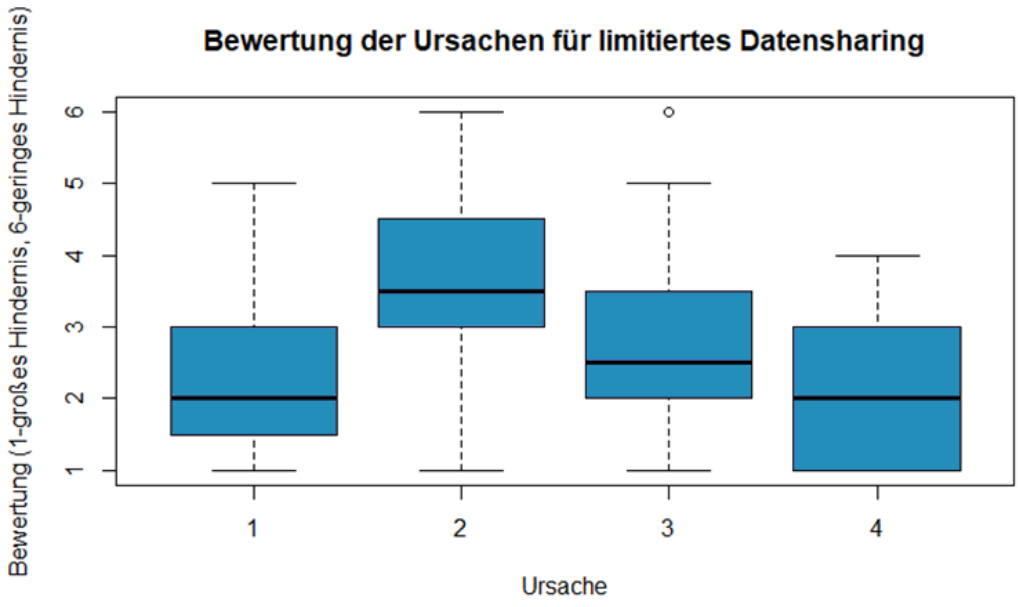


Abbildung 26: Einschätzung der Ursachen für Zurückhaltung beim Datensharing

Im Folgenden werden die Anforderungen der Teilnehmer an die Ausgestaltung des zu implementierenden Datenmarktplatzes und des Datensharings beschrieben. Allen Teilnehmern ist es wichtig, zu bestimmen, wer die von ihnen angebotenen Daten verwenden kann. Bis auf einen Befragten möchten die Unternehmen zudem selbst darüber entscheiden können, für wen das durch Metadaten beschriebene Datenangebot sichtbar ist. Werden Daten an Geschäftspartner weitergegeben ist zudem die Definition der Nutzungsbedingungen ein wichtiger Aspekt für die Teilnehmer. Daher wurde ermittelt, über welche Aspekte der Datennutzung die Teilnehmer bestimmen möchten. Abbildung 27 zeigt die Bewertung einzelner Maßnahmen zur Beschränkung der Datennutzung bei der Weitergabe von Daten durch die Umfrageteilnehmer. Es wird deutlich, dass eine Verhinderung der Weitergabe von Daten an unberechtigte Dritte von den Teilnehmern als wichtigste Maßnahme eingeschätzt wird. Alle Teilnehmer halten diese Maßnahme für notwendig. Weiterhin wünschen sich drei Viertel der Teilnehmer eine Möglichkeit, den Anwendungszweck der weitergegebenen Daten zu beschränken. Als drittnotwendigste Maßnahme sehen die Personen eine Verhinderung der Aggregation ihrer Daten mit Daten der Konkurrenz. Die weiteren möglichen Maßnahmen wie etwa die Beschränkung der Nutzungsdauer, die Verhinderung der Aggregation personenbezogener Daten und insbesondere die Beschränkung der Nutzungsanzahl werden von den Teilnehmern weniger relevant eingeschätzt. Zusätzlich wurde ebenso die Möglichkeit eingeräumt, weitere Dinge zu nennen, über die während der Datennutzung entschieden werden soll. Ein Teilnehmer nutzte diese Möglichkeit um zu äußern, dass er gerne über Veröffentlichungen basierend auf seinen Daten entscheiden möchte.

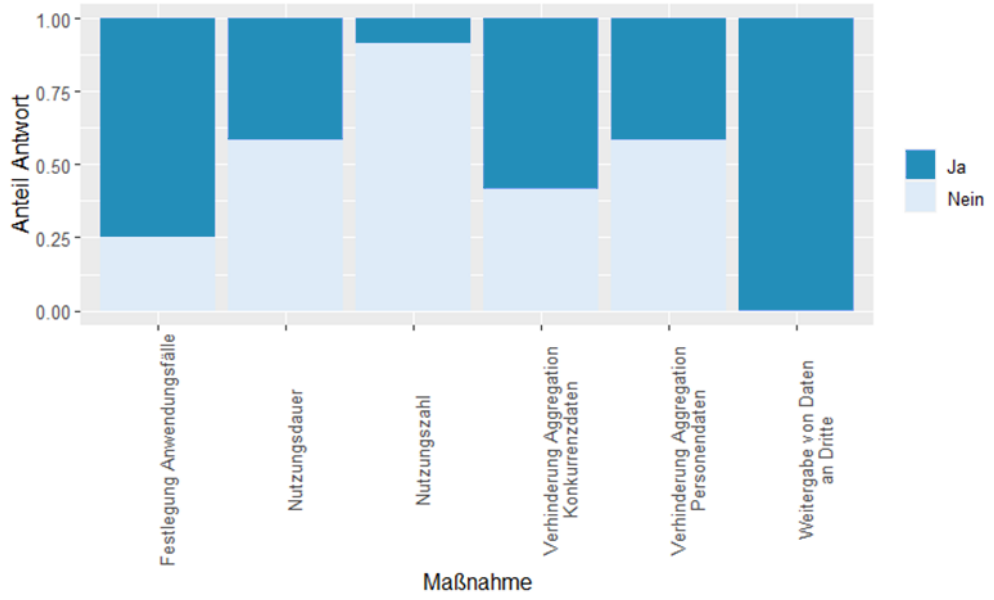


Abbildung 27: Bewertung der Notwendigkeit von Maßnahmen zur Beschränkung der Datennutzung

Technische Maßnahmen auf Basis von Usage Control ermöglichen es, die zuvor als notwendig formulierten Nutzungsbedingungen durchzusetzen. Zu ihrer Implementierung ist es notwendig, dass die partizipierenden Unternehmen einer technischen Lösung ihr Vertrauen schenken. Sollten Vorbehalte gegenüber einer technischen Lösung bestehen gilt es, Maßnahmen zu treffen, die für ein erhöhtes Vertrauen sorgen. Acht der befragten Teilnehmer bringen Usage Control ein sehr hohes oder hohes Vertrauen entgegen. Demgegenüber stehen drei Unternehmen mit größeren Vorbehalten hinsichtlich der technischen Maßnahme Usage Control. Gründe für diese Vorbehalte bestehen in a) einer hohen Komplexität einer lückenlosen Umsetzung b) einer geringen Bekanntheit und Kenntnis über die Lösung c) einer generellen Skepsis gegenüber neuer Technologien innerhalb des Unternehmens.

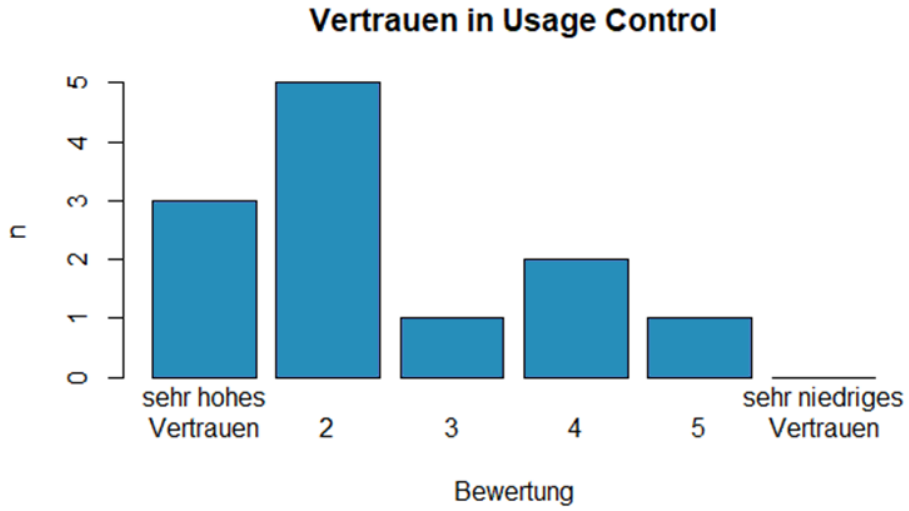


Abbildung 28: Bewertung des Vertrauens in Usage Control

Werden technische Lösungen zur Umsetzung der Datennutzungskontrolle verwendet, so entsteht ein nicht vermeidbarer Overhead. Dieser führt zu Verzögerungen beim Datenzugriff, welche idealerweise im Millisekundenbereich liegen, jedoch grundsätzlich von den Charakteristika der übertragenen Daten und der konkreten Ausprägung der verwendeten Usage Control Mechanismen abhängen. Daher gilt es zu prüfen, ob die Implementierung von Usage Control Mechanismen vor diesem Hintergrund den in IIP-Ecosphere entwickelten Services schaden würde. In diesem Zusammenhang gibt nur ein Teilnehmer an, dass er absolut nicht in der Lage ist, eine solche Verzögerung zu akzeptieren. Die große Mehrheit der Teilnehmer (10) ist jedoch bereit, eine Verzögerung bei der Datenübertragung bei der technischen Umsetzung von Usage Control in Kauf zu nehmen. Hinsichtlich der Auswahl einer Usage Control Lösung für IIP-Ecosphere ist es entscheidend, ob bereits existierende Applikationen oder Services um die Fähigkeit von Usage Control erweitert werden, oder ob neuartige Applikationen entstehen, die bereits im Entwicklungsprozess mit Usage-Control Prinzipien versehen werden können. Weitere Umfrageteilnehmer[8] geben dazu an, innerhalb von IIP-Ecosphere neuartige Services entwickeln zu wollen, die auf Daten der Geschäftspartner als essentielle Eingangsgröße basieren.

Um zu etablierende technische Maßnahmen auf den Rahmen von IIP-Ecosphere anzupassen, sind weitere Informtionen zum geplanten Datenaustausch nötig. So ist unter anderem relevant, ob es sich um einen bilateralen oder multilateralen Datenaustausch handelt und ob es sich bei den Partnern ausschließlich um Teilnehmer des IIP-Konsortiums oder ebenso um externe Partner handelt. Etwa die Hälfte der Befragten (7) plant im Rahmen von IIP-Ecosphere Projekte, bei denen Daten von mehr als einem Anbieter bezogen werden. Ist dies der Fall, so sind nur selten alle Datenlieferanten auch Teilnehmer in IIP-Ecosphere (2). Auf der Seite des Teilens der Daten ist selbiges der Fall. Auch hier existieren bei vielen Teilnehmern Projekte (8), bei denen Daten mit mehr als einem Teilnehmer geteilt werden und ebenso sind dabei oft nicht alle Empfänger der Daten Teilnehmer im IIP-Ecosphere Konsortium (6).

4.2 Ableitung von Handlungs- und Gestaltungsempfehlungen

Auf Basis der zuvor dargestellten Ergebnisse der Umfrage zu Datenschutz und Datensicherheit werden im Folgenden, unter der Berücksichtigung der in Abschnitt 3 vorgestellten Lösungskonzepte, Gestaltungsempfehlungen für die entstehende Plattform und den Datenmarktplatz sowie für die Mechanismen zur Herstellung von Datenschutz und Datensicherheit abgeleitet.

[8] Die genaue Anzahl wird in diesem Whitepaper aus Datenschutzgründen nicht genannt

4.2.1 Empfehlungen für die entstehende zentrale Plattform

Der erste Betrachtungsbereich für die Ableitung von Anforderungen stellt die im Rahmen des Projekts entstehende Plattform sowie die dort zur Nutzung bereitgestellten Services dar. In der Umfrage gibt die Hälfte der Teilnehmer an, einer Übertragung ihrer Daten auf eine zentrale Plattform skeptisch gegenüber zu stehen. Es kann somit als allgemein wichtig angesehen werden, Maßnahmen zu Datenschutz und Datensicherheit zu etablieren, um einerseits die Werte des Unternehmens zu schützen und andererseits Vertrauen in die Nutzung der zentralen Plattform zu generieren. Weiterhin wird durch eine Ermöglichung der Nutzung von Services on-premises eine Reihe an Problemen des Datenschutzes und der Datensicherheit vermieden. Ersteres ist insbesondere vor dem Hintergrund, dass drei Viertel der Unternehmen von einer Schützenswürdigkeit der verwendeten Daten ausgehen, wichtig. Hinsichtlich einer Auslegung der Mechanismen wurde in Erfahrung gebracht, dass viele Unternehmen die Services mit Echtzeitdaten nutzen möchten. Darüber hinaus gilt es, Schutzmaßnahmen so zu etablieren, dass eine Datenverarbeitung in Echtzeit weiterhin gewährleistet werden kann. Daten mit Personenbezug werden nur von einem geringen Teil der Teilnehmer analysiert. Im Vergleich der einzelnen Mechanismen zum Schutz personenbezogener Daten konnte keine präferierte Methode herausgestellt werden. Gleichwohl stoßen die Methoden Anonymisierung, Pseudonymisierung und Verschlüsselung allesamt bei den Teilnehmern auf Zustimmung. Mit KIProtect steht dem Projektkonsortium ein Partner zu Verfügung, dessen Kernkompetenz im Schutz von Personendaten bei der Verwendung von Datenanalysen und Methoden von künstlicher Intelligenz liegt. Aufgrund der gleichrangigen Wertung der Verfahren sollte zur Reduktion des Aufwands uneingeschränkt auf die Methoden von KIProtect zurückgegriffen werden, um personenbezogene Daten zu schützen.

Unabhängig von der Datenart wurden die Teilnehmer zur Bewertung verschiedener Datenschutzmaßnahmen bei Nutzung von Analyseservices aufgefordert. Zwar war die statistische Signifikanz der Unterschiede für das gesamte Konsortium beschränkt, jedoch stellten sich die Notwendigkeit eines ausreichenden Schutzes der Daten vor dem Zugriff durch Dritte, die Limitierung der Daten auf den Verwendungszweck, die Verhinderung des Datenverlustes auf der Plattform und die Möglichkeit zum Löschen der Daten nach Verwendung als wichtige Anforderungen für den Datenschutz im Rahmen der zentralen Plattform heraus. Zusätzlich ergibt sich die Anforderung an eine Transparenz hinsichtlich Technologien und Verwendungszweck der Daten, welche von mehreren Teilnehmern als zusätzliches Kriterium genannt wurde und damit als weitere wichtige Anforderung gesehen werden kann.

4.2.2 Empfehlungen für das Datensharing

Im Abschnitt Datensharing werden Anforderungen erhoben, um den Schutz und die Sicherheit von Daten während des Austausches mit anderen Unternehmen zu gewährleisten. Zunächst zeigt sich, dass keines der befragten Unternehmen Daten zum Kauf anbietet. Dies betont die Notwendigkeit weiterer Forschung zur Monetarisierung von Daten, wie sie im Think Tank Daten stattfindet, um den Teilnehmern den Einstieg in die Datenökonomie zu erleichtern. Denn auf der Seite des Datenempfängers profitieren einige Unternehmen bereits durch die Nutzung von käuflich erworbenen oder frei verfügbaren Fremddaten. Hinsichtlich der Charakteristik der ausgetauschten Daten geben die meisten Teilnehmer an, Daten in Form von Tabellendokumenten austauschen zu wollen. Aber auch hinsichtlich semistrukturierter und unstrukturierter Daten prognostizieren die Teilnehmer einen zukünftigen Datenaustausch. Zusätzlich werden Daten mit Echtzeitcharakteristik und Personenbezug genutzt. Bei der Etablierung von Schutzmaßnahmen bei weitergegebenen Daten sind daher die vorangehend beschriebenen Charakteristika zu beachten und die Maßnahmen entsprechend zu spezifizieren. Weiterhin gilt es zur Erhöhung des Daten Sharings Maßnahmen zu treffen, welche die signifikantesten Probleme beim Datenaustausch in Angriff nehmen. Als diese wurden die Preisgabe von Geschäftsgeheimnissen, der Missbrauch der Daten für nicht autorisierte Zwecke und die Sicherheit

der Daten beim Nutzer identifiziert. Besonders die Sorge vor der Preisgabe von Geschäftsgeheimnissen zeigt, dass die ausschließliche Nutzung technischer Maßnahmen nicht zur Erzeugung von Vertrauen ausreicht. Stattdessen sind für den Schutz von Geschäftsgeheimnissen auch Organisatorische Maßnahmen der Data Governance, wie der Risikobewertung notwendig. Für die weitere Erzeugung von Vertrauen sind Maßnahmen gegen den Missbrauch der Daten für nicht autorisierte Zwecke sowie Maßnahmen für die Sicherheit der Daten beim Nutzer zu treffen.

4.2.3 Empfehlungen Datenmarktplatz

In IIP-Ecosphere wird mit dem Datenmarktplatz eine Komponente zum kontrollierten Daten Sharing entwickelt. Die Befragung der Teilnehmer ergab verschiedene Anforderungen an einen Datenmarktplatz. Hinsichtlich der zu gestaltenden Architektur des Datenmarkplatzes zeigen sich die Teilnehmer unentschlossen. Eine leichte Tendenz geht dabei jedoch in Richtung einer hybriden Auslegung. Weiterhin wurden die Teilnehmer befragt, welche Bestimmungsrechte sie bei der Weitergabe ihrer Daten erhalten möchten. So fordern die Teilnehmer die Möglichkeit zu entscheiden, wer ihr Datenangebot sehen kann, wer die Daten des Unternehmens beziehen kann und welche Bedingungen nach der Nutzung der Daten eingehalten werden müssen.

4.2.4 Empfehlungen zur Umsetzung von Usage Control

Um eine Einhaltung der Nutzungsbedingungen durchzusetzen, bieten Methoden von Usage Control eine technische Realisierungsmethode. Um eine Unterstützung der Implementierung durch das Konsortium sicherzustellen, wurde zunächst geprüft, ob etwaige Vorbehalte existieren. Es zeigt sich, dass ein Großteil der Unternehmen einer technischen Umsetzung generell Vertrauen schenken würden. Die skeptischen Unternehmen besitzen Vorbehalte eher allgemeiner Natur. Diese ergeben sich etwa aus der hohen technischen Komplexität und der Tatsache, dass ohne Einzelfallprüfung generell keiner Lösung vertraut werden kann. Auch die Befragung hinsichtlich des durch Usage Control induzierten Overheads, welcher insbesondere bei der Verwertung von Echtzeitdaten ein Hindernis darstellen kann, wurde von nahezu allen Teilnehmern als akzeptabel eingestuft. Einzig ein Teilnehmer gibt an, diese zusätzliche Verzögerung nicht akzeptieren zu können. Dementsprechend kann Usage Control als eine Methode zum Schutz der Daten verwendet werden.

Mithilfe von Usage Control können verschiedene Nutzungsbedingungen, die zuvor in Form von Policies definiert wurden, umgesetzt werden. Da es sich jedoch derzeit noch um ein Forschungsfeld handelt, können die vorhandenen Lösungen mitunter nicht jegliche Nutzungsbedingung umsetzen. Daher sollte bei Auswahl der Usage Control Lösung auf eine Umsetzung der von den Teilnehmern priorisierten Nutzungsbedingungen geachtet werden. Im Rahmen dieses Whitepapers wurden Maßnahmen mit einer Zustimmungsrate von mehr als 50 % in der Umfrage als für die Teilnehmer wichtig aufgefasst. Daraus ergeben sich folgende Nutzungsbedingungen, die durch die ausgewählte Usage Control Lösung unterstützt werden sollten:

- Verhinderung der Weitergabe von Daten
- Beschränkung der Datennutzung auf bestimmte Anwendungsfälle
- Verhinderung der Zusammenführung von Daten mit Daten der Konkurrenz

Bei der Umsetzung von Usage Control sind zudem weitere Rahmenbedingungen zu beachten. So werden in einigen Projekten Daten mit mehr als einem weiteren Unternehmen geteilt oder Daten von mehr als einem Unternehmen empfangen. Darunter fallen auch Kollaborationen, bei denen nicht alle am Datenaustausch beteiligten Unternehmen auch selbst Teilnehmer in IIP-Ecosphere sind und dementsprechend für diese andere Regeln und Grundlagen gelten. Weiterhin werden innerhalb des Projekts IIP-Ecosphere neue Services entworfen. Diese Services können bereits Usage Control in den Designprozess mit einbeziehen und dadurch eine höhere Flexibilität hinsichtlich der Policies gewährleisten.

5 Konsolidierung der Erkenntnisse

Im Rahmen dieses Abschnittes werden die Erkenntnisse aus den vorangegangenen Abschnitten der Literaturanalyse und des Fragebogens miteinander verknüpft und konsolidiert, um in der Folge eine einheitliche Handlungsgrundlage für die folgenden Aktivitäten zur Gewährleistung von Datenschutz und Datensicherheit in Datenökosystemen zu schaffen. Zusätzlich werden weitere wichtige Erkenntnisse im Hinblick auf die Leistung der entstehenden Plattform und der Teilnehmer in IIP-Ecosphere konsolidiert.

5.1 Datenmarktplatz

Mit dem Datenmarktplatz entsteht im Think Tank Daten eine prototypische Komponente zum sicheren Sharing von Daten. Dabei ist es das Ziel, innerhalb von IIP-Ecosphere eine Marktplatz-Lösung zu entwerfen, die durch Datenschutz und -sicherheitskonzepte ein erhöhtes Maß an Sicherheit für die Anwender bietet. Eine Voraussetzung des Erfolgs des Datenmarktplatzes stellt dabei die Adaption der Nutzer dar, weshalb neben einer fundierten Analyse des Stands der Technik ebenso eine Berücksichtigung der Präferenzen potentieller Anwender von Nöten ist. Mit der Befragung des Konsortiums wurde dabei diesem Umstand entsprochen.

Im Abschnitt Stand der Technik wurden drei gangbare Alternativen zur Architektur der Datenhaltung und somit auch des Datenmarktplatzes vorgestellt. Bei einer zentralen Architektur werden die Daten und Metadaten auf eine zentrale Plattform übertragen. Die Daten verlassen dabei die Grenze des Unternehmens und die Unternehmen verlieren dementsprechend die Kontrolle über ihre Daten. Damit ist es ihnen nur noch bedingt möglich über die Art der Nutzung ihrer Daten zu entscheiden. Andersherum ergeben sich für den Datenempfänger eine einfache Auffindbarkeit der benötigten Datensätze und einheitliche Schnittstellen zum Empfang der Daten. Demgegenüber steht das Konzept der dezentralen Datenhaltung, bei dem sämtliche Daten beim Datenanbieter verbleiben. Dieser kann dementsprechend autonom entscheiden, welche Daten welchem Empfänger zur Verfügung gestellt werden. Allerdings ergibt sich hierbei eine Vielzahl von Schnittstellen und eine schwierige Allokation von Angebot und Nachfrage. Solche Architekturen wurden beispielhaft auf Basis von Blockchain realisiert, was dementsprechende technologiebedingte Vor-und Nachteile impliziert. Um die Vorteile beider Architekturen zu vereinen, wurden sogenannte hybride Architekturen entwickelt, bei denen die Metadaten durch eine zentrale Instanz bereitgestellt werden, die weiteren Daten jedoch im System des Datenanbieters verbleiben. Somit ergibt sich die Möglichkeit für die Datenanbieter, ihre Souveränität über die eigenen Daten zu bewahren. Die Datennachfrager erhalten andererseits die Möglichkeit das Datenangebot an einer zentralen Stelle zu durchsuchen. Einzig der Mehraufwand zur Implementierung steht dieser Architektur negativ gegenüber. Aufgrund des Betriebs der entstehenden Plattform on-premises sollte sich auch der Datenmarktplatz an diesem Konzept ausrichten. Entsprechend wird empfohlen, die Daten für den entstehenden Datenmarktplatz dezentral bei den Datenanbietern zu halten und ausschließlich die Metadaten über das verfügbare Datenangebot – ähnlich zum einem Anwendungsstore an zentraler Stelle verfügbar zu machen und somit dem Prinzip einer hybriden Architektur zu folgen.

Auch das Konsortium wurde nach den Präferenzen im Bereich der Architektur eines Marktplatzes befragt. Dabei konnte jedoch keine statistisch signifikante Präferenz der Teilnehmer herausgestellt werden. Tendenziell bewerten die Umfrageteilnehmer eine hybride Lösung als angemessener, was die Architekturentscheidung des Think Tank Daten unterstützt. Ebenso wurden weitere Anforderungen und Gestaltungswünsche bei der Auslegung des Datenmarktplatzes in der Befragung erfasst. So möchten die Teilnehmer sowohl beschränken, wer die von ihnen zur Verfügung gestellten Daten nutzen kann, als auch, wer das Datenangebot des einzelnen Teilnehmers sehen kann. Bei Nutzung der zentralen Instanz zur Speicherung der Metadaten könnte eine Einschränkung der Sichtbarkeit des Datenangebots über die Kategorisierung der Unternehmen in verschiedene Klassen und eine

dementsprechende Limitierung der Sichtbarkeit für einzelne Klassen erfolgen. Weiterhin wurden mögliche Hürden bei der Teilnahme am Datenmarktplatz untersucht. Dazu wurden die in der Literatur identifizierten Hindernisse zur Teilnahme am Datenaustausch auf einem Datenmarktplatz durch die Teilnehmer bewertet. Dabei stellte sich heraus, dass die Gefahr der Preisgabe von Geschäftsgeheimnissen und der Missbrauch von Daten für nicht genehmigte Zwecke die größten Hürden zum Sharing von Daten auf Datenmarktplätzen darstellen. Während erstes vor allem ein Problem im Bereich des unternehmensinternen Risikomanagements darstellt, liefert Usage Control Lösungen für Probleme des zweiten Bereichs. Eine weitere Hürde zur Teilnahme stellt die Besorgnis der Teilnehmer hinsichtlich der Datensicherheit beim Datennutzer dar, während die Sicherheit bei der Datenübertragung weniger problematisch gesehen wird.

Im Rahmen dieses Whitepapers wurden die unterschiedlichen Rollen auf einem Datenmarkplatz erläutert, um darzustellen, dass es für jedes Unternehmen wichtig ist, seine eigene Rolle auf dem Datenmarktplatz und die damit verbundenen Tätigkeiten zu kennen. Von besonderer Bedeutung ist dabei die Rolle des Intermediärs, der für eine Vermittlung des Angebots basierend auf den zur Verfügung gestellten Metadaten sorgt und in Form eines Clearing House für die Abwicklung der Transaktionen sorgt oder Konflikte löst. Oftmals wird diese Rolle von einem führenden Unternehmen eingenommen. Da es sich bei IIP-Ecosphere um ein Projekt ohne ein solch führendes Unternehmen handelt, muss die Implementierung Automatismen zur Lösung solcher Probleme bereitstellen.

Weiterhin sind Beiträge zur Governance und dem Identitäts- und Authentifizierungsmanagement (IAM) wichtig für den Datenmarktplatz. Die Ausgestaltung ebendessen sollte sich an den vorgestellten Prinzipien Transparenz, Fairness, Einfachheit, Realitätsbezug, Shared value und Teilhabe orientieren. Das IAM-Modell sollte dabei mit dem Plattform IAM-Modell vereinheitlicht werden.

Im Rahmen des Datensharings sehen es die Teilnehmer in der Befragung zudem als wichtig an, über die Nutzungsbedingungen entscheiden zu können und diese durchzusetzen. Usage Control stellt eine solche Technologie dar, die in diesem Kapitel in einem separaten Abschnitt betrachtet wird. Generell ist eine Integration von geeigneten Usage Control Mechanismen in den Marktplatz eine sinnvolle Ergänzung, aber die Nutzung dieser Mechanismen sollte nicht verpflichtend sein. Ansonsten würde die Akzeptanz bei Unternehmen, welche kein Vertrauen bzw. kein Interesse an technischen Umsezungen von Nutzungsbedingungen haben, gefährdet. Darüber hinaus ist zu beachten, dass Usage Control Mechanismen nur in einem System funktionieren, in dem alle beteiligten Komponenten über entsprechende Mechanismen verfügen bzw. notwendige Informationen bereitstellen. Diese Bereitstellung von Informationen kann beispielsweise mithilfe der Asset Administration Shell erfolgen.

5.2 Usage Control

Wie durch die Ergebnisse der Konsortialbefragung bestätigt, stellt Usage Control ein wichtiges Instrument zur Durchsetzung von Nutzungsbedingungen und –restriktionen beim Austausch von Daten in Datenökosystemen dar. Dabei kann Usage Control sowohl bei der Nutzung von Fremdservices als auch beim Sharing von Daten verwendet werden. Die Nutzung von Usage Control erzeugt dabei Transparenz und Vertrauen zwischen den Partnern und garantiert die digitale Souveränität der einzelnen Teilnehmer. Bei der Verwendung von Usage Control Technologien aus den IDS bietet sich auch der Einsatz von IDS Connectoren an. Diese stellen eine vertrauenswürdige Umgebung dar, in welcher Nutzungsbedingungen durchgesetzt werden können und erlauben eine einfachere Identifizierung von Systemgrenzen. Generell können die Technologien aber auch unabhängig von den IDS verwendet werden. In diesem Fall müssen aber diejenigen Systemgrenzen definiert werden, welche durch die Usage Control Technologien kontrolliert werden.

Dennoch sollte die Technologie Usage Control keineswegs als verpflichtend für die Teilnehmer während des Datenaustausches oder der Nutzung von Services festgelegt werden. Denn, wie die

Befragung des Konsortiums zeigt, besitzen nicht alle Teilnehmer ein vollständiges Vertrauen in diese Lösung. Gründe dafür bestehen u.a. in der hohen Komplexität, einer geringen Bekanntheit der Lösung oder einer generellen Skepsis gegenüber neuen Technologien. Gleichermaßen ist mit der Implementierung von Usage Control ein Aufwand verbunden, der, sollten sich Teilnehmer bereits vertrauen, lediglich zu einer zusätzlichen Belastung wird. Schlussendlich zeigte sich durch die Befragung auch, dass nicht alle Teilnehmer die durch die Verwendung von Usage Control entstehende zeitliche Verzögerung für die von ihnen verwendete Echtzeitanwendung akzeptieren können. Daher wird Usage Control für die Teilnehmer als optionale Leistung verfügbar werden.

Im Rahmen der Literaturrecherche wurden drei in den IDS entwickelte Methoden von Usage Control erläutert und gegenübergestellt. Eine aus dem IND²UCE-Framework hervorgegangene Usage Control Lösung ist MYDATA. Bei dieser Lösung werden Usage Control Container auf Seiten des Datensenders und des Datenempfängers im IDS-Connector initialisiert. Bei Sendung der Daten werden diese entsprechend der Policies durch den Usage Control Container bearbeitet. Beim Empfänger werden die Daten wiederum initial durch den dortigen Usage Control Container geleitet, bevor dieser die Daten der Vorgaben entsprechend an andere Applikationen weiterleitet. Die Erstellung von Policies kann dabei in der IDS Policy Language erfolgen. Die Policies werden von MYDATA übersetzt und implementiert.

Die Usage Control Lösung LUCON ist ein fester Bestandteil des IDS Trusted Connectors. Bei Lucon werden Labels an Daten geheftet und diese, analog zu MYDATA, mithilfe des Interference-Muster in Apache Camel. Dabei werden die an die Daten gehefteten Labels geprüft und eine Weiterverarbeitung nur erlaubt, sofern entsprechende Label vorhanden sind. Usage Policies müssen vor der Nutzung in der logischen Programmiersprache Prolog formuliert werden.

D° ist eine domänenspezifische Programmiersprache, die es erlaubt, kann Usage Control von Beginn an in die Entwicklung von Applikationen einzubinden. Die aktuelle Version von D° verwendet dazu JAVA als Host Language und kann daher bei der Entwicklung von Anwendungen in entsprechender Programmiersprache genutzt werden. Dabei ergibt sich ein initialer Mehraufwand, der allerdings, der allerdings mit wachstender Reife in Zukunft sinken wird.

Die Lösungen besitzen unterschiedliche Ansätze und technologische Reifegrade und sind in der Lage, unterschiedliche Usage Policies umzusetzen. Daher wurde im Rahmen der Befragung ermittelt, welche Rahmenbedingungen des Datenaustausches vorherrschen und welche Art von Nutzungsbedingungen die Teilnehmer durchsetzen möchten. Hinsichtlich der für den Einsatz von Usage Control relevanten Rahmenbedingungen des Datenaustausches ergab sich, dass nahezu alle Teilnehmer davon ausgehen, in Zukunft Daten im strukturierten Format auszutauschen. Zusätzlich gibt jeweils die Hälfte der Teilnehmer an, Daten in semistrukturierten oder unstrukturierten Format austauschen zu wollen. Weitere für Usage Control relevante Charakteristika von Daten stellen die Verwendung von Daten mit Personenbezug und von Echtzeitdaten dar. Aufgrund des Kontexts der Use Cases, gehen viele Unternehmen von einer Übertragung von Echtzeitdaten an die Partner aus. Usage Control muss daher gewährleisten, dass die entstehenden Verzögerungen durch den Mehraufwand so gering wie möglich gehalten werden. Der Anteil der ausgetauschten Daten mit Personenbezug wird allerdings im Rahmen von IIP-Ecosphere eher gering ausfallen.

Hinsichtlich der durchzusetzenden Nutzungsbedingungen existieren drei vorrangig gewünschte Mechanismen. Zum einen sollte die Weitergabe der Daten an unberechtigte Dritte unterbunden werden können. Weiterhin sollten die Daten nur für vorher definierte Anwendungsfälle genutzt werden. Einen weiteren wichtigen Aspekt für die Teilnehmer stellt die Verhinderung der Aggregation mit Konkurrenzdaten dar. Weniger geforderte, aber dennoch relevante Aspekte sind zudem die Festlegung der Nutzungsdauer der weitergegebenen Daten und die Verhinderung der Aggregation von Daten mit Personenbezug.

Basierend auf diesem Gerüst von Anforderungen kann keine allgemeingültige Empfehlung zur Nutzung von Usage Control Technologien in IIP-Ecosphere gegeben werden. Es muss auf Basis des jeweiligen Einzelfalls entschieden werden, ob ein bestimmter/s Service/System mit Usage Control Mechanismen ausgestattet werden soll. Dies hängt von verschiedenen Faktoren ab. Zum einen muss betrachtet werden, ob die Daten, welche von dem Service/System verarbeitet werden sollen, überhaupt ein Schutzbedürfnis haben, welches den zusätzlichen Overhead durch die Verwendung von Usage Control Mechanismen rechtfertigt. Ebenfalls muss abgewogen werden, ob die Verzögerung, welche durch die Verwendung von Usage Control Technologien entsteht, im Rahmen der (weichen) Echtzeitanforderungen des jeweiligen Anwendungsfalls genügen. Da im Rahmen von IIP-Ecosphere an vielen Stellen bestehende (Open Source) Software verwendet werden soll, ist es nicht möglich Usage Control Technologien direkt in diese Software zu integrieren. Dies kommt höchstens für neue Entwicklungen in Frage. Hierdurch wird die Auswahl an zur Verfügung stehenden Usage Control Lösungen eingeschränkt. Ebenfalls muss bei der Entscheidung und Auswahl berücksichtigt werden, dass für manche Usage Control Lösungen Lizenzgebühren anfallen würden, was im Rahmen von IIP-Ecosphere nicht wünschenswert ist. Es kann nur die allgemeine Empfehlung ausgesprochen werden, die Verwendung von Usage Control Technologien als optional zu gestalten. Dies ist ein Ergebnis der durchgeführten Befragung des Konsortiums.

5.3 Relevante Erkenntnisse für die entstehende Plattform

Neben den primär im Think Tank Daten behandelten Themen Datenmarktplatz und Usage Control, konnten mittels der Literaturanalyse und der anschließenden Konsortialbefragung weitere relevante Aspekte für die in IIP-Ecpshere entstehende Plattform identifiziert werden. Diese gehen dabei oftmals Hand in Hand mit den zuvor konsolidierten Resultaten der zentralen Plattform.

Ein erster Betrachtungsaspekt stellt dabei die Datenhaltung dar. Wie bereits im Teilbereich Datenmarktplatz beschrieben, existieren auch für die Implementierung der Plattform die drei möglichen Grundprinzipien zentral, hybrid oder vollkommen dezentral. Dabei wurde im Rahmen der Konsortialbefragung ermittelt, dass ein Großteil der Teilnehmer die für die Nutzung der bereitgestellten Services verwendeten Daten als Schützenswert hält. Weiterhin steht die Hälfte der Teilnehmer der Übertragung von Daten auf eine zentrale Plattform zur Nutzung von Services skeptisch gegenüber. Unabhängig von den vorhandenen technischen Restriktionen, etwa die Nutzung von Echtzeitdaten, die sich durch eine solche zentrale Plattform ergeben, sprechen somit weitere Faktoren für eine dezentrale oder hybride Ausführung.

Weiterhin wurden die Teilnehmer befragt, welche Maßnahmen im Bereich von Datenschutz und Datensicherheit im Hinblick auf die Nutzung von Services ein erhöhtes Vertrauen in die IIP-Ecosphere Plattform erzeugen würden. Dabei ist für die Teilnehmer ein ausreichender Schutz vor dem Datenzugriff durch Dritte wichtig. Ebenso wird eine Limitierung des Verwendungszwecks, eine Verhinderung des Datenverlustes auf der Plattform und die Möglichkeit zum Löschen der Daten nach Verwendung als wichtig angesehen. Zusätzlich fordern die Teilnehmer eine Transparenz hinsichtlich der Verwendeten Technologien eines Service. Letzteres könnte in etwa durch eine ausführliche Servicebeschreibung mittels Metadaten oder Use Case Beschreibungen erreicht werden.

Äquivalent zum Datenmarktplatz sollen auch bei der zentralen Plattform Governance-Mechanismen für ein hohes Vertrauen unter den Nutzern sorgen und ebenso attraktive Rahmenbedingungen zur Teilnahme schaffen. Diese Governance Regeln sollen dabei den sechs in Abschnitt 3.1.4 präsentierten Prinzipien folgen. Dabei konnten vier besonders relevante Entscheidungsfelder identifiziert werden:

- Regulatorische Umgebung: Identifikation und Umsetzung relevanter rechtlicher Regularien; Erstellung und kontinuierliches Monitoring interner Richtlinien

- Ownership und Zugangsberechtigungen: Wer besitzt die Services, Daten und die Ergebnisse und wer kann auf diese zugreifen?
- Data Use Cases: Wie werden Daten genutzt? Welche Rahmenbedingungen, Rechte und Pflichten existieren?
- Bewertung: Wie wird der erzielte Mehrwert unter den Teilnehmern aufgeteilt; Bewertungssystem

Weiterhin wurde die Einführung von Regeln und Richtlinien als ein Steuerungsmechanismus aus dem Bereich der Governance beschrieben. Aufgrund der Tatsache, dass im Projekt IIP-Ecosphere bereits während der Konzeptionierungsphase Rahmenbedingungen und Regeln für die Arbeit im Projekt definiert wurden, kann davon ausgegangen werden, dass bereits eine vertrauenswürdige Umgebung für die gemeinsame Zusammenarbeit geschaffen wurde, die auch für die Nutzung der zentralen Plattform gilt. Dementsprechend müssen initial keine weiteren Richtlinien geschaffen werden. Sollten sich in einem reiferen Stadium der Plattform zusätzlich notwendige Rahmenbedingungen, Standards oder Zertifizierungen ergeben, so können diese nachträglich gefordert werden.

Zur weiteren Erzeugung von Vertrauen unter den Teilnehmern ist die Implementierung eines geeigneten IAM-Modells notwendig. Dabei wurden im Rahmen des Whitepapers vier mögliche Modelle (Isolated Identity Model, Central Identity Model, User-centric Identity Model, Federated Identity Model) gegenübergestellt. Für IIP-Ecosphere wurde das Central Identity Model als angemessenste Methode aufgefasst. In diesem werden die Identitäten durch eine zentrale Instanz identifiziert, authentifiziert und entlang des Lebenszyklus verwaltet. Für die Service- oder Datenprovider können die erforderlichen Daten einfach abgerufen werden. Nutzer der Plattform haben den Vorteil mittels eines Benutzerkontos sämtliche Services nutzen zu können.

5.4 Teilnehmerrelevante Ergebnisse

Im Hinblick auf die Teilname an der Data Economy und zur weiteren Etablierung eines sicheren und souveränen Datenaustausches konnten im Rahmen der Erarbeitung dieses Whitepapers weitere Maßnahmen identifiziert werden, die innerhalb der Unternehmen getroffen werden können, um Datenschutz und Datensicherheit zu gewährleisten.

Ausgangslage ist dabei die Unwissenheit über einige Charakteristika der durch die Unternehmen erzeugten und verwendeten Daten. So sind einige Unternehmen (auch in der Unternehmensbefragung) nicht vollständig in der Lage, die Schützenswürdigkeit ihrer Daten zu beziffern. Weiterhin haben viele Unternehmen Angst, dass eine Weitergabe ihrer Daten an Partnerunternehmen zu einer Offenbarung der Geschäftsgeheimnisse und damit zu einem Verlust des kompetitiven Vorteils oder einer Verhandlungsposition führen könnte. Diese nicht ausreichende Analyse des Datenpotentials und der damit verbundenen Risiken führen vielerorts zu einer Zurückhaltung hinsichtlich des unternehmensübergreifenden Austauschs von Daten oder der Nutzung neuer Technologien gemäß dem Ansatz „better safe than sorry“.

Zur Behandlung dieser Situation und der damit verbundenen Nutzung sich neu ergebener Geschäftsmodelle und technologischer Potentiale können nicht ausschließlich technische Maßnahmen genutzt werden. Stattdessen sind weitere organisatorische Maßnahmen aus dem Feld der Data Governance denkbar. Insbesondere wurde dabei im Rahmen dieses Whitepapers das Daten-Risikomanagement betrachtet. Im Gegensatz zu den unternehmensinternen Ansätzen existiert dabei oftmals ein Informationsdefizit hinsichtlich der sicherheitsrelevanten Technologien der Geschäftspartner. Diese besitzt dementsprechend eine höhere Komplexität im Vergleich zu unternehmensinterner Data Governance. Insgesamt wurden drei Risikotypen identifizert und dazugehörige Lösungsansätze ermittelt. Durch die geringe Verbreitung des Daten Sharings handelt es sich hierbei jedoch nicht um bereits ausgereifte Konzepte.

Das Risiko durch die Weitergabe von Daten an Dritte sensible Informationen zu verlieren wurde analog zur risikobasierten Zugangskontrolle betrachtet, da es sich bei der Weitergabe von Daten – stark vereinfacht – ebenso um den Zugang zu einem Service handelt. Relevante Faktoren zur Abschätzung eines Risikos stellen dabei die Risikohistorie, die Ressourcensensitivität und der Nutzungskontext dar. Um das Risiko in den jeweiligen Bereichen abschätzen zu können, ist es in der Umgebung eines digitalen Ökosystems wie IIP-Ecosphere notwendig, die relevanten Daten zur Analyse transparent bereitzustellen. Das Risiko bei der Nutzung von Fremdservices können bereits existierende Methoden des Risikomanagements im Rahmen von Cloud-Computing verwendet werden. Dort sind bereits eine Reihe von Vorgehensmodellen vorhanden, die zusätzlich notwendige Schutzmaßnahmen vorgeben. Schlussendlich existiert zudem ein Risiko bei der Verwendung von Daten Dritter, etwa durch einen Ausfall der Datenquelle. Die Analyse dieser Risiken während des Datenbezugs kann auf Basis eines Metadatenkatalogs erfolgen, der ein geeignetes Metadatenmodell zur Beschreibung der relevanten Informationen besitzt. Ein geeignetes Metadatenmodell wird durch das Metadata Model for Data Goods (M4DG) bereitgestellt. Unternehmen werden unter Verwendung der bereitgestellten Daten dabei selbst befähigt, eigene Metriken zur Risikoanalyse zu applizieren. Dabei kann dieses Datenmodell auch für die in IIP-Ecosphere vorgesehenen dezentralen Strukturen verwendet werden.

Ein eng mit der Risikoanalyse verbundenes Thema stellt die Aufstellung von Nutzungsbedingungen für zur Verfügung gestellte Ressourcen, wie etwa bei der Weitergabe von Daten, dar. Dabei wurden in Abschnitt 3.2.4 bisherige Erkenntnisse im Rahmen dieses Whitepapers zusammengefasst und Empfehlungen für die drei Phasen Input, Development und Output ausgesprochen.

Nicht zuletzt stellt die Behandlung von personenbezogenen Daten einen wichtigen Teilaspekt im Bereich von Datenschutz und Datensicherheit in IIP-Ecosphere dar. Die Verhinderung der Identifizierung natürlicher Personen ist eine obligatorische Aufgabe für die datenbesitzenden Unternehmen, die sowohl einerseits während der Nutzung von angebotenen Services als auch im Rahmen der Weitergabe von Daten an andere Unternehmen gewährleistet werden muss. Dazu liegt es an den Unternehmen, adäquate Maßnahmen zu treffen. Eine Reihe dieser Maßnahmen und Konzepte wurde in Abschnitt 3.2.1 vorgestellt. Die Befragung des Konsortiums ergab allerdings, keine bevorzugte Maßnahme hinsichtlich der Konzepte Anonymisierung, Pseudonymisierung und Verschlüsselung, wobei allerdings alle drei möglichen Ansätze als angemessen eingeschätzt wurden. Mit KIProtect existiert ein Konsortialpartner, dessen primäre Kompetenz der Schutz von personenbezogenen Daten während des Daten Sharings darstellt und dessen Entwicklungen frei zur Verfügung stehen[9]. Im Rahmen von IIP-Ecosphere sollten dementsprechend diese Kompetenz abgerufen und die vorhandenen Konzepte für das Datensharing und die Servicenutzung integriert werden. Insgesamt wird jedoch festgestellt, dass derzeit noch ein hoher Forschungsbedarf hinsichtlich adäquater Maßnahmen zur Vorbereitung von Unternehmen zur Bereitstellung von Daten in Datenökosystemen existiert.

[9] https://github.com/kiprotect/kodex

6 Zusammenfassung

Daten- bzw. Plattformökosysteme stellen eine neuartige Organisationsform dar, in welcher Innovation und wirtschaftlicher Nutzen durch die Beziehungen von Unternehmen untereinander erzeugt werden. In Datenökosystemen werden Datenquellen verschiedener Organisationen und über die Grenzen einzelner Industrien angereichert. Zur Gewinnung von teilnehmenden Unternehmen ist es erforderlich, ein Mindestmaß von Datenschutz und Datensicherheit zu gewährleisten. Damit verbunden sind insbesondere die Erzeugung von Vertrauen zwischen den Teilnehmern, die Gewährleistung von Souveränität über die eingebrachten Daten und die Governance von Daten und des Ökosystems.

Im Forschungsprojekt IIP-Ecosphere wird anvisiert ein Datenökosystem um die entstehende IoT-Plattform mitsamt einem prototypischen Datenmarktplatz zum Austausch von Daten zu entwickeln. Dementsprechend sind etwaige Maßnahmen zu treffen, um die Sicherheit und den Schutz der Daten sicherzustellen und die Souveränität der Teilnehmer zu gewährleisten. Dieses Whitepaper befasst sich mit möglichen Ausprägungen von Datenschutz und Datensicherheit in Datenökosystem unter besonderer Berücksichtigung der in IIP-Ecosphere entstehenden Komponenten. Im Rahmen einer Literaturrecherche wurden mögliche Probleme in ebendiesem Bereich erfasst und mögliche Lösungsmaßnahmen und –komponenten identifiziert. Diese wurden im Anschluss auf deren Eignung in IIP-Ecosphere bewertet. Insbesondere wurde dies durch eine Befragung des Projektkonsortiums unterstützt.

Im Bereich des Datenmarkplatzes wurde ermittelt, dass dieser einer hybriden Architektur folgen sollte. Dies bedeutet, dass einzig die Metadaten in einer zentralen Komponente zur Auffindung des Datenangebots bereitgestellt werden. Die eigentlichen Daten verbleiben bis zum Abschluss bei den Anbietern. Weiterhin wurden die größten Hemmnisse einer Teilnahme durch potentielle Teilnehmer ermittelt. Insbesondere stellten sich dabei die Preisgabe von Geschäftsgeheimnissen und der Missbrauch von Daten für nicht genehmigte Zwecke als „Pain-Points“ heraus.

Als eine technische Maßnahme zur Gewährleistung von Datensouveränität und der Verhinderung von Datenmissbrauch wurde Usage Control im Rahmen dieses Whitepapers eine hohe Beachtung geschenkt. Usage Control ermöglicht die Durchsetzung von Nutzungsbedingungen in vertrauenswürdigen Umgebungen. Bei der Implementierung einer solchen Lösung in IIP-Ecosphere stellten sich drei Maßnahmen (Verhinderung der Datenweitergabe, Verwendung der Daten für definierten Anwendungsfall, Aggregation der Daten) als zentral heraus. Insgesamt soll ein Angebot zur Verwendung von Usage Control geschaffen werden, dass optional durch die Teilnehmer verwendet werden kann, da je einzelnem Anwendungsfall untersucht werden muss, welche Vorteile eine Usage Control Lösung bietet.

Abschließend konnten weitere Empfehlungen hinsichtlich der entstehenden Plattform und einzelner Nutzer getätigt werden. Hinsichtlich erstem Teilaspekt wurden unter anderem eine (teilweise) dezentrale Ausführung der Plattform als Anforderung der Teilnehmer ermittelt. Verbunden dazu sollte die Plattform eine ausreichende Transparenz hinsichtlich der angebotenen Services bieten. Weiterhin wird die Unterstützung eines Central Identity Model empfohlen. In Bezug auf einzelne Teilnehmer wurde deutlich, dass insgesamt noch ein hoher Forschungsbedarf im Bereich Risikomanagement und Freigabe von Unternehmensdaten in Datenökosystemen notwendig ist. Auch die Behandlung personenbezogener Daten vor der Datenweitergabe stellt weiterhin ein offenes Forschungsfeld dar, das in Zukunft angegangen werden sollte, um insbesondere kleinen und mittelständischen Unternehmen die Teilnahme an der Data Economy zu erleichtern.

7 Referenzen

[1] OECD, *Enhancing Access to and Sharing of Data*. OECD, 2019.

[2] R. a. M. Roundtable on Environmental Health Sciences, B. o. P. H. a. P. H. Practice, H. a. M. Division, E. a. M. National Academies of Sciences und E. Rusch, *Principles and Obstacles for Sharing Data from Environmental Health Research: Workshop Summary*. Washington, D.C: National Academies Press, 2016. [Online]. Verfügbar unter: http://gbv.eblib.com/patron/FullRecord.aspx?p=4528551

[3] A. S. Figueiredo, „Data Sharing: Convert Challenges into Opportunities“ (eng), *Frontiers in public health*, Jg. 5, S. 327, 2017, doi: 10.3389/fpubh.2017.00327.

[4] M. D'Agostino, P. Pellaton und A. Brown, „Mobility Data Sharing: Challenges and Policy Recommendations“, UC Davis, 2019.

[5] J. Kaewkungwal, P. Adams, J. Sattabongkot, R. K. Lie und D. Wendler, „Issues and Challenges Associated with Data-Sharing in LMICs: Perspectives of Researchers in Thailand“ (eng), *The American journal of tropical medicine and hygiene*, Jg. 103, Nr. 1, S. 528–536, 2020, doi: 10.4269/ajtmh.19-0651.

[6] M. Spiekermann, „Data Marketplaces: Trends and Monetisation of Data Goods“, *Intereconomics*, Jg. 54, Nr. 4, S. 208–216, 2019, doi: 10.1007/s10272-019-0826-z.

[7] P. Koutroumpis, A. Leiponen und L. D. W. Thomas, „The (unfullfilled) potential of data marketplaces“, *ETLA Working Papers*, 2017.

[8] M. Spiekermann, D. Tebernum, S. Wenzel und B. Otto, „A metadata model for data goods“, *Multikonferenz Wirtschaftsinformatik*, S. 326–337, 2018.

[9] B. Otto, M. ten Hompel und S. Wrobel, „International Data Spaces“ in *Digital Transformation*, R. Neugebauer, Hg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2019, S. 109–128, doi: 10.1007/978-3-662-58134-6_8.

[10] Yuri Demchenko, Wouter Los, Cees de Laat, „Data as economic goods: Definitions, properties, challenges, enabling technologies for future data markets“, 2018.

[11] B. Otto und M. Jarke, „Designing a multi-sided data platform: findings from the International Data Spaces case“, *Electron Markets*, Jg. 29, Nr. 4, S. 561–580, 2019, doi: 10.1007/s12525-019-00362-x.

[12] D. Lis, „Ökosysteme für Daten und Künstliche Intelligenz“, 2019.

[13] International Data Spaces Association, Hg., „International Data Spaces: Reference Architecture Model Version 3“, 2019. Zugriff am: 10. September 2020.

[14] Andreas Eitel, Christian Jung, Christian Kühnle, Fabian Bruckner, Gerd Brost, Pascal Birnstill, Ralf Nagel, Sebastian Bader, „Usage Control in International Data Spaces“.

[15] A. Munoz-Arcentales, S. López-Pernas, A. Pozo, Á. Alonso, J. Salvachúa und G. Huecas, „Data Usage and Access Control in Industrial Data Spaces: Implementation Using FIWARE“, *Sustainability*, Jg. 12, Nr. 9, S. 3885, 2020, doi: 10.3390/su12093885.

[16] M. Jarke, B. Otto und S. Ram, „Data Sovereignty and Data Space Ecosystems“, *Business & Information Systems Engineering*, Jg. 61, Nr. 5, S. 549–550, 2019, doi: 10.1007/s12599-019-00614-2.

[17] G. Sejdic, *Produktionscontrolling im Kontext von Industrie 4.0*. Nomos Verlagsgesellschaft mbH & Co. KG, 2019.

[18] B. Otto und H. Österle, *Corporate Data Quality*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016.

[19] P. Koutroumpis, A. Leiponen und L. D. W. Thomas, „Markets for data“, *Industrial and Corporate Change*, Jg. 29, Nr. 3, S. 645–660, 2020, doi: 10.1093/icc/dtaa002.

[20] D. Lis und B. Otto, „Data Governance in Data Ecosystems - Insights from Organizations“, *Americas Conference on Information Systems*, 2020.

[21] M. Schreieck, M. Wiesche und H. Krcmar, „Design and Governance of Platform Ecosystems – Key Concepts and Issues for Future Research“ in *Twenty-Fourth European Conference on Information Systems*, 2016.

[22] S. U. Lee, L. Zhu und J. Ross, „Data Governance for Platform Ecosystems: Critical Factors and the State of Practice“, *Twenty First Pacific Asia Conference on Information Systems*, 2017.

[23] Lee, S. U., Zhu, L., & Jeffery, R., „Designing Data Governance in platform ecosystems“, *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[24] S. U. Lee, L. Zhu und R. Jeffery, „Data Governance Decisions for Platform Ecosystems“, 2019, doi: 10.24251/HICSS.2019.766.

[25] M. W. van Alstyne, G. G. Parker und S. P. Choudary, „Pipelines, platforms, and the new rules of strategy“, *Harvard business review*, Jg. 94, Nr. 4, S. 54–62, 2016.

[26] Y. Cao und L. Yang, „A survey of Identity Management technology“ in *2010 IEEE International Conference on Information Theory and Information Security (ICITIS)*, Beijing, China, 2010, S. 287–293, doi: 10.1109/ICITIS.2010.5689468.

[27] A. Josang, M. AlZomai und S. Suriadi, „Usability and privacy in identity management architectures“ in *ACSW Frontiers 2007: Proceedings of 5th Australasian Symposium on Grid Computing and e-Research, 5th Australasian Information Security Workshop (Privacy Enhancing Technologies), and Australasian Workshop on Health Knowledge Management and Discovery*, S. 143–152.

[28] N. Selvanathan, D. Jayakody und V. Damjanovic-Behrendt, „Federated Identity Management and Interoperability for Heterogeneous Cloud Platform Ecosystems“ in *the 14th International Conference*, Canterbury, CA, United Kingdom, 2019, S. 1–7, doi: 10.1145/3339252.3341492.

[29] D. Tebernum, M. Spiekermann, S. Wenzel und B. Otto, „Risikobewertungen in Datennetzwerken“, 2018.

[30] J. K. M. Müller, „Dateneigentum in der vierten industriellen Revolution?“, *Datenschutz und Datensicherheit-DuD*, Jg. 43, Nr. 3, S. 159–166, 2019.

[31] V. Hammer und M. Knopp, „Datenschutzinstrumente Anonymisierung, Pseudonyme und Verschlüsselung“, *Datenschutz und Datensicherheit-DuD*, Nr. 8, S. 503–509, 2015.

[32] N. Benamar, „Bildbasierte Authentifizierung und Verschlüsselung“. Dissertation, Universität Kassel, 2008.

[33] L. Sweeney, „k-anonymity: A model for protecting privacy“, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Jg. 10, Nr. 05, S. 557–570, 2002.

[34] A. Machanavajjhala, D. Kifer, J. Gehrke und M. Venkitasubramaniam, „L -diversity“, *ACM Trans. Knowl. Discov. Data*, Jg. 1, Nr. 1, S. 3, 2007, doi: 10.1145/1217299.1217302.

[35] N. Li, T. Li und S. Venkatasubramanian, „t-closeness: Privacy beyond k-anonymity and l-diversity“ in *2007 IEEE 23rd International Conference on Data Engineering*, S. 106–115.

[36] K. Rajendran, M. Jayabalan und M. E. Rana, „A Study on k-anonymity, l-diversity, and t-closeness Techniques“, *IJCSNS*, Jg. 17, Nr. 12, S. 172, 2017.

[37] C. Dwork, „Differential privacy: A survey of results“, *International conference on theory and applications of models of computation*, S. 1–19, 2008.

[38] M. U. Hassan, M. H. Rehmani und J. Chen, „Differential Privacy Techniques for Cyber Physical Systems: A Survey“, *IEEE Commun. Surv. Tutorials*, Jg. 22, Nr. 1, S. 746–789, 2020, doi: 10.1109/COMST.2019.2944748.

[39] L. Wiese, D. Homann, T. Waage und M. Brenner, „Homomorphe Verschlüsselung für Cloud-Datenbanken: Übersicht und Anforderungsanalyse“, 2018. Zugriff am: 30. Juli 2020.

[40] F. Di Cerbo und S. Trabelsi, „Towards Personal Data Identification and Anonymization Using Machine Learning Techniques“ in *Communications in Computer and Information Science*, Bd. 909, *New trends in databases and information systems: ADBIS 2018 short papers and workshops, AIQA, BIGPMED, CSACDB, M2U, BigDataMAPS, ISTREND, DC, Budapest, Hungary, September, 2-5, 2018 : proceedings*, A. Benczur, Hg., Cham: Springer, 2018, S. 118–126, doi: 10.1007/978-3-030-00063-9_13.

[41] R. Lorenz, T. H. Netland, P. Roh, V. Holzwarth, A. Kunz und K. Wegener, „Data-driven productivity improvement in machinery supply chains“, *IJMMS*, Jg. 12, 3/4, S. 255, 2019, Art. no. 103483, doi: 10.1504/IJMMS.2019.103483.

[42] A. Rot, *IT Risk Assessment: Quantitative and Qualitative Approach* (Zugriff am: 23. Juli 2020).

[43] A. Amini und N. Jamil, „A Comprehensive Review of Existing Risk Assessment Models in Cloud Computing“, *J. Phys.: Conf. Ser.*, Jg. 1018, S. 12004, 2018, doi: 10.1088/1742-6596/1018/1/012004.

[44] H. F. Atlam, M. A. Azad, M. O. Alassafi, A. A. Alshdadi und A. Alenezi, „Risk-Based Access Control Model: A Systematic Literature Review“, *Future Internet*, Jg. 12, Nr. 6, S. 103, 2020, doi: 10.3390/fi12060103.

[45] H. Khambhammettua, S. Boularesb, K. Adib und L. Logrippob, „A Framework for Risk Assessment in Access Control Systems“, *Computers & Security*, 2013.

[46] T. Nokkala, H. Salmela und J. Toivonen, „Data Governance in Digital Platforms“, *Twenty-fifth Americas Conference on Information Systems*, 2019.

[47] H. Paananen, M. Lapke und M. Siponen, „State of the art in information security policy development“, *Computers & Security*, Jg. 88, S. 101608, 2020, doi: 10.1016/j.cose.2019.101608.