

Risikobewertungen in Datennetzwerken

Daniel Tebernum¹ · Markus Spiekermann¹
Sven Wenzel¹ · Boris Otto^{1,2}

¹Fraunhofer ISST – Digitization in Service Industries
{daniel.tebernum | markus.spiekermann | sven.wenzel
boris.otto}@isst.fraunhofer.de

²TU Dortmund – Lehrstuhl für Supply Net Order Management
boris.otto@tu-dortmund.de

Zusammenfassung

Die in den Unternehmen verfügbaren Daten haben sich durch die Digitalisierung und der damit verbundenen digitalen Transformation von Geschäftsmodellen und Geschäftsprozessen zu einer strategischen Ressource entwickelt. Durch die Bildung von Datenökosystemen, in denen Unternehmen untereinander über Plattformen und Marktplätze ihre Daten austauschen und dem damit verbundenen stetigen Zuwachs an Daten aus internen und externen Quellen, wird das Management von Risikobewertungen einzelner Datenressourcen immer schwieriger. Die Komplexität der verteilten Datenarchitektur innerhalb dieser Datenökosysteme erschwert es den verantwortlichen Personen den Überblick zu behalten. Diese Arbeit stellt ein Konzept zur qualitativen Bewertung von möglichen Risiken einzelner Datenressourcen in verteilten Datennetzwerken vor. Hierbei wird der Ansatz einer innerbetrieblichen Verwaltung von Datenressourcen durch Datenkataloge und der Nutzung zugeschnittener Metadatenmodelle auf einen unternehmensübergreifenden Kontext übertragen, der eine automatisierte Analyse und Bewertung ermöglicht. In einer prototypischen Anwendung wurde das Konzept implementiert und anhand einer beispielhaften Risikobewertung evaluiert.

1 Einleitung

Für Unternehmen stellen Daten, im Rahmen der digitalen Transformation, eine strategische Ressource mit wachsender Bedeutung dar [Ott15]. Nur durch eine individuell passende und fachgerechte Bewirtschaftung der Daten können Unternehmen in Zukunft innovative Geschäftsmodelle aufbauen [OÖ16]. Neben den klassischen Informationssystemen, welche laufend enorme Datenmengen verarbeiten und produzieren, sorgen vor allem Trends der Digitalisierung wie Big Data, Machine Learning und das Internet der Dinge (IoT) für eine immer weiter steigende Flut an Daten [GR12, AYH⁺17]. Zusätzlich sorgen Plattformen und Datenmarktplätze mit ihren Datenangeboten, die weltweit über das Internet erreichbar sind, für eine stark verteilte Datenarchitektur. Unternehmen stehen deshalb vor der Herausforderung, den Überblick über ihre Datenressourcen, die in den verschiedenen Geschäftsprozessen genutzt werden, zu behalten und die Bedeutung einer Datenressource im Sinne eines Ausfallrisikos zu bewerten. Dabei reicht es nicht aus, nur auf die unternehmenseigenen Datenressourcen zu schauen. Auch unternehmensfremde Datenressourcen, die von Extern in das unternehmensinterne Datennetzwerk fließen, spielen bei der Bewertung von möglichen Risiken eine Rolle.

Diese Arbeit stellt im Rahmen der Design Science [HMPR04] einen Lösungsansatz für Unternehmen vor, welcher auf der Verwaltung von internen und externen Datenressourcen in Informationssystemen mit Katalogisierungsfunktion und der Nutzung von Metadaten basiert. Hierzu werden in einem ersten Schritt die notwendigen Hintergründe zur aktuellen Forschung im Bereich von Datenkatalogsystemen und Metadatenmodellen vorgestellt. Aufbauend auf dem aktuellen Stand der Technik werden die Möglichkeiten und Herausforderungen einer Überführung dieser auf verteilte, unternehmensübergreifende Datennetzwerke analysiert und vorgestellt. Als Ergebnis wird ein Konzept präsentiert, welches die Erkenntnisse der vorangegangenen Analyse nutzt, um einen Lösungsansatz für die Umsetzung in Informationssystemen zu geben. Die Evaluation des Konzeptes erfolgt anhand einer prototypischen Implementierung eines Datenkatalogs mit den erweiterten Funktionalitäten, welche für eine übergreifende Risikobetrachtung notwendig sind.

2 Stand der Forschung und Technik

Durch die immer weiter ansteigende Menge an Daten wird es zunehmend schwerer, den Überblick über diese zu behalten. Datenkataloge unterstützen hierbei, indem sie den Nutzern Werkzeuge an die Hand geben, um die Daten einheitlich zu beschreiben und zu organisieren. Zaidi et al. definieren einen Datenkatalog dabei wie folgt:

„A data catalog maintains an inventory of data assets through the discovery, description and organization of datasets. The catalog provides context to enable data analysts, data scientists, data stewards and other data consumers to find and understand a relevant dataset for the purpose of extracting business value.“ [ZDSED17]

Wie in der Definition beschrieben, fokussieren sich Datenkataloge in erster Linie auf die Inventarisierung von Daten. Dafür werden die Metadaten von Datenressourcen, meistens aufbauend auf einem standardisierten Metadatenmodell, im Datenkatalogsystem hinterlegt. Nutzer können dann, beispielsweise über eine Google-ähnliche Suchoberfläche, nach für sie passenden Daten suchen. Datenkataloge bieten typischerweise zusätzliche Funktionalitäten an. Man kann hinterlegen, wer für die Daten verantwortlich ist, wer die Daten verwendet und von wo die Daten bezogen werden. Hierdurch kann das Vertrauen in die genutzten Datenressourcen gesteigert werden [ZDSED17]. Datenkataloge werden als Teil einer ganzheitlichen Digitalisierungsstrategie angesehen und sind dementsprechend eine unverzichtbare Komponente in der zukunftsorientierten Datenarchitektur eines Unternehmens [MW13].

Um einen Überblick über die existierenden Datenkataloglösungen zu bekommen, wurde eine heuristische Recherche durchgeführt. Als Ausgangspunkt diente dabei die Arbeit von Zaidi et al. [ZDSED17]. Insgesamt konnten während der Recherche 26 Anwendungen identifiziert werden. Anwendungen wie Comprehensive Knowledge Archive Network (CKAN) [CKA18], Collibra Catalog [Col18] oder Waterline Data Catalog [Wat18] fungieren dabei primär als Datenkatalogsanwendungen. Sie sind auf das Management der Metadaten von Datenressourcen spezialisiert und bieten darüber hinaus Funktionalitäten, die auf der Nutzung der vorhandenen Metadaten aufbauen. Andere Anwendungen wie DSpace [Dur18, TBS⁺03] und Dataverse [Dat18, Kin07] bringen nur eine implizite Katalogisierungsfunktion mit, da ihre Hauptfunktionalität im Bereich von Datenrepositories positioniert ist. In einem weiteren Schritt wurde untersucht, inwiefern die recherchierten Anwendungen ihre Nutzer dabei unterstützen neben den eigenen auch externe Datenressourcen, die in den Geschäftsprozessen des Unternehmens verwendet werden, zu verwalten. Hierbei konnte zum einen festgestellt werden, dass sich Da-

tenkataloge primär auf die intern genutzten Datenressourcen eines Unternehmens beschränken und entstehende verteilte Systeme und Datenaustausch zwischen Unternehmen in Netzwerken nicht ausreichend berücksichtigen. Das Themenfeld einer Bewertung von Sicherheits- und Risikoaspekten in Bezug auf die in den Unternehmen verwendeten Daten nennen nur wenige der betrachteten Datenkataloge explizit.

Datenkataloge speichern Metadaten über die zu katalogisierenden Ressourcen. Welche Metadaten abgespeichert werden, ist von der Implementierung des betreffenden Datenkataloges abhängig. Um ein einheitliches Verständnis über die Informationen zu erlangen, haben sich diverse Metadatenmodelle zur Beschreibung von Daten und den dazugehörigen Datenquellen herausgebildet. Einen weit verbreiteten Standard zur Beschreibung von Attributen in einem Metadatenmodell stellt das Dublin Core Metadata Element Set (DCE) [WKLW98] dar. Dieses wurde im Jahr 2003 als ISO Standard 15836 [Int03] festgehalten. DCE bietet 15 Kern-Attribute zur Beschreibung an, welche durch andere Vokabulare wiederverwendet werden können. Die Attribute sind generisch gehalten, sodass sie in vielen Anwendungsfällen genutzt werden können. Beispielhafte Attribute sind *title*, *description* und *date*. Ein Metadatenmodell, welches auf DCE aufbaut ist das Data Catalog Vocabulary (DCAT) [MCP10]. Dieses Metadatenmodell wurde explizit zur Beschreibung von Datenressourcen sowie Datenkatalogen entwickelt und wird beispielsweise von Anwendungen wie CKAN eingesetzt. Mit DCAT soll die Interoperabilität zwischen verschiedenen Katalogen verbessert werden. DCAT erlaubt die Modellierung von verschiedenen Distributionen der Unternehmensdaten. Aufbauend auf DCAT wurde von der Interoperability Solutions for European Public Administrations (ISA) das Asset Description Metadata Schema (ADMS) geschaffen [SLPS13]. Dieses soll die semantische Interoperabilität zwischen den Behörden der verschiedenen Staaten der Europäischen Union verbessern. Hierdurch wird dem Gedanken von verteilten Datenressourcen Rechnung getragen. ADMS erweitert DCAT um eine Reihe von Klassen, die beispielsweise Kontaktpersonen modellieren oder Informationen über den Zeitraum der Gültigkeit des Assets geben. Die hinzugefügten Klassen des ADMS reichen jedoch nicht aus um der Verwaltung von Datenressourcen in der digitalen Ökonomie gerecht zu werden. Mit dem Metadata Model for Data Goods (M4DG) wurde von Spiekermann et al. ein Metadatenmodell entwickelt, welches Unternehmensdaten im Sinne eines Wirtschaftsgutes [STWO18] und Datenressourcen in verteilten Datennetzwerken beschreibt. So beinhaltet das M4DG unter anderem Klassen für die Definition von Preismodellen, Servicevereinbarungen und Nutzern der Datenressource. Durch die Abbildung letzterer können Relationen zu Nutzern von spezifischen Datenressourcen hergestellt werden. Auch die Verknüpfung von Prozessen, die auf eine Datenressource angewiesen sind, ist möglich.

3 Erstellung von Metriken in Datenkatalogen

Daten spielen eine immer wichtigere Rolle für den Erfolg eines Unternehmens. Durch den Einsatz eines Datenkataloges kann erfasst werden, welche Datenressourcen in einem Unternehmen vorhanden sind. Datenkataloge, so wie es sie heute gibt, adressieren aber nur unzureichend den Wunsch nach einer Risikobewertung für Datenressourcen. Die Daten können den Erfolg eines Unternehmens steigern oder gar einen wichtigen Teil des Geschäftsmodells ausmachen [Ott15]. Deshalb ist es wichtig, dass die Unternehmen mögliche Risiken in Bezug auf ihre genutzten Datenressourcen im Blick behalten. Da Datenkataloge schon jetzt als ein essentieller Teil der Digitalisierungsstrategie von Unternehmen angesehen werden [ZDSED17], bieten sie somit auch einen möglichen Raum, um auf Risiken, in Bezug auf die genutzten Datenressourcen, aufmerksam zu machen. Unternehmen bewerten mögliche Risiken individuell und haben

unterschiedliche Auffassungen davon, was ein potentielles Risiko für sie darstellt. Um dieser Individualität gerecht zu werden, wurde ein Web-basierter Datenkatalog in Node.js implementiert, in denen Unternehmen eigene Risikometriken hinzufügen können. Durch diese Flexibilisierung sollen mehr Unternehmen in die Lage versetzt werden, ihre eigenen Anforderungen an eine Risikobewertung für Datenressourcen umzusetzen.

Der implementierte Datenkatalog nutzt M4DG zur Beschreibung der dort hinterlegten Datenressourcen. M4DG bietet umfangreiche Möglichkeiten, eine Datenressource als ein Wirtschaftsgut zu beschreiben [STWO18]. Die dort angebotenen Attribute können für eine Bewertung der Datenressource, im Sinne eines unternehmerischen Risikos, herangezogen werden. Die Metriken müssen nicht zwangsweise generisch formuliert werden, sodass sie auf alle Datenressourcen, die im Datenkatalog abgelegt werden können, passen. Durch eine Gruppierungsfunktion können Metriken thematisiert werden. Beispielsweise können auf unternehmenseigene Datenressourcen andere Metriken angewendet werden als auf extern hinzugekaufte. Die Flexibilisierung wird durch einen Code-Editor ermöglicht. In diesem können eigene Metriken in der Programmiersprache JavaScript formulieren werden. Dabei stehen einem alle Funktionalitäten der Programmiersprache zur Verfügung. Der Entwickler einer Metrik kann über die vordefinierte Variable *asset* auf die vorhandenen Attribute des M4DG zugreifen. Mit *asset.distributions.length* kann beispielsweise die Anzahl der vorhandenen physikalischen Repräsentationen der Datenressource bestimmt werden. Geschriebene Metriken können außerdem in einer Sandbox getestet werden, um während des Entwicklungsprozesses die Korrektheit der Implementierung prüfen zu können. Je nach Anforderung können Metriken sowohl numerische als auch textuelle Werte als Ergebnis haben. Wählt man im Datenkatalog die Detailansicht einer Datenressource aus, so werden dort die Ergebnisse der Metrikkalkulationen angezeigt. Die Ausführung des JavaScript Programms erfolgt gekapselt, sodass fehlerhafter oder böswilliger Code keine Auswirkungen auf den Rest der Datenkatalogsanwendung hat. Hierfür werden Methoden eingesetzt, wie sie auch von Online-IDEs wie *JS Bin* (<https://jsbin.com>), *JS-Fiddle* (<https://jsfiddle.net>) oder *CodePen* (<https://codepen.io/>) angewendet werden. Der Code wird in einem *sandboxed iFrame* innerhalb eines *Web Workers* ausgeführt, wobei das Interface klare Restriktionen aufweist. Es können beispielsweise keine Funktionen als Ergebnis zurück gegeben werden.

Zum besseren Verständnis wird im Folgenden eine Beispielmetrik erstellt. Die Metrik soll Unternehmen auf Datenressourcen aufmerksam machen, welche durch eine geringe Verfügbarkeit die Durchführung von unternehmensinternen Prozessen blockieren könnten. Über das *subscribers* Attribut des M4DG lassen sich die Relationen von einer Datenressource zu Unternehmensprozessen modellieren. Die Prozesse wiederum modellieren ihre Verfügbarkeitsanforderungen an Datenressourcen durch einen prozentualen Wert. Andere Konzepte wie *Mean Time Between Failures* (MTBF) oder *Mean Time To Repair* (MTTR), die bei Verfügbarkeitsanalysen herangezogen werden können, werden zu Gunsten eines einfachen Beispiels hier nicht betrachtet. Für jede physikalische Repräsentation einer Datenressource (in M4DG: *distributions*) lässt sich ebenfalls die Uptime als prozentualer Wert bestimmen. Für jede Relation von Datenressource und Prozess lässt sich nun sagen, ob die Minimalanforderungen an die Uptime gegeben sind. Abbildung 1(a) zeigt, wie die beispielhafte Metrik im Datenkatalog implementiert wurde und Abbildung 1(b) wie das Ergebnis auf der Detailseite einer Datenressource dargestellt wird. Die Metrik lässt sich beliebig verfeinern, sodass beispielsweise die zuvor genannten Aspekte wie MTBF und MTTR mitberücksichtigt werden können. Die einzige Voraussetzung hierfür ist, dass das genutzte Metadatenmodell zur Beschreibung von Datenressourcen die notwendigen

Informationen modellieren kann. Das hier vorgestellte Konzept zur Flexibilisierung von Metriken lässt sich prinzipiell in andere Datenkataloge überführen und ist nicht auf ein bestimmtes Metadatenmodell zur Beschreibung von Datenressourcen angewiesen. Neben Risikometriken lassen sich auch beliebige andere Arten von Kennzahlen anhand der im Metadatenmodell hinterlegten Werte berechnen.

Metric Development

Name

Show in following groups

Metric calculation

```

1 let res = "Requirements fulfilled";
2 asset.subscribers.forEach(function(subscriber) {
3   asset.distributions.forEach(function(distribution) {
4     if (subscriber.min_uptime > distribution.uptime) {
5       res = "Requirements not fulfilled!";
6     }
7   });
8 });
9 return res;

```

Default Metrics

Metric	Result
Completeness of Metadata	Good

Specific Metrics

Metric	Result
Uptime Requirements by Processes	Requirements fulfilled
Business Impact	3

(a) Entwicklungsumgebung für Metriken

(b) Metrikauswertung für eine Datenressource

Abb. 1: Screenshots aus Datenkataloganwendung

4 Einsatz von Datenkatalogen

Datenkataloge unterstützen die Verwaltung von Datenressourcen, indem sie Metainformationen sammeln und bereitstellen, die von den Mitarbeitern eines Unternehmens genutzt werden können. So erleichtern Datenkataloge unter anderem das Auffinden passender Daten, sowie die Qualitätskontrolle und die Kollaboration zwischen den Nutzern einer Datenressource. Datenkataloge sind ebenfalls dazu geeignet, Datensilos zu vermeiden, indem Informationen zu Ablageorten, verantwortlichen Personen, Aktualisierungszyklen und Alternativen an einer zentralen Stelle bereitgestellt werden. Ein zentral genutzter Datenkatalog unterstützt schon heute Unternehmen dabei, aus ihrem eigenen Datenpool heraus, passende Daten für ihre Anwendungsfälle zu finden. Auch können Konformitätsanforderungen durch die Definition von Metriken und der Anwendung automatisierter Analysen unterstützt werden. Im vorherigen Kapitel wurde gezeigt, wie Unternehmen, durch den Einsatz eines flexibilisierten Systems zur Erstellung eigener Metriken, Risiken in Bezug auf ihre Datenressourcen aufdecken können. Diese durch den Datenkatalog bereitgestellten Fähigkeiten müssen für den unternehmensübergreifenden und weltweiten Datenaustausch adaptiert und hochskaliert werden. Datenkataloge müssen die Sicht auf einzelne Unternehmen und Institutionen verlassen und eine globale Perspektive einnehmen können. Dadurch soll es ermöglicht werden, über Unternehmensgrenzen hinweg, und perspektivisch auch automatisiert, nach passenden Datensätzen zu suchen und diese in die eigenen Wertschöpfungsketten und Datennetzwerke zu integrieren. Durch das Management der Metadaten externer Datenressourcen, im Datenkatalog des eigenen Unternehmens, lassen sich Funktionalitäten, wie beispielsweise die durch das Unternehmen definierten Risikometriken, auf nicht unternehmensinterne Datenressourcen anwenden.

Es stellt sich die Frage, wie ein System aussehen muss, in welchem die Metadaten der verschiedenen Unternehmen ausgetauscht werden können. Eine Möglichkeit wäre die Bereitstellung eines einzelnen, globalen Datenkatalogs, in welchem alle Teilnehmer ihre Datenangebote

eintragen können. Diese Lösung ist jedoch nicht praktikabel. Der Datenkatalog müsste von einem der Unternehmen oder einer dritten Instanz bereit gestellt werden, wodurch ungewollte Abhängigkeiten zum Betreiber des Systems entstehen würden. Auch müssten die Metadaten der Datenressourcen doppelt gepflegt werden. Einmal für interne Daten im Datenkatalog des Unternehmens und einmal für zu teilende Daten im globalen Datenkatalog. Das Unternehmen ihren gesamten Datenbestand in einem von Dritten organisierten Datenkatalog verwalten ist ebenfalls unwahrscheinlich. Unternehmen haben erkannt, dass Daten ein wertvolles Gut sind und wollen vermeiden, dass Mitbewerber ungewollt Einblicke bekommen [SJ18, FPFS⁺17].

Im Rahmen dieser Arbeit wurde daher die Überlegung bevorzugt, die einzelnen Datenkataloge aus den Unternehmen direkt miteinander zu verknüpfen. Die Datenkataloge können auf diese Weise ein weltweites Datennetzwerk aufspannen. Dadurch, dass die Datenkataloge innerhalb der Unternehmen verweilen, bleibt die Datensouveränität gewahrt. Unternehmen behalten die Kontrolle über die eigenen, internen Metadaten und geben nur die Metadaten nach außen, die sie wirklich teilen möchten. Abbildung 2 zeigt schematisch den Zusammenschluss mehrerer Datenkataloge zu einem großen Verbund.

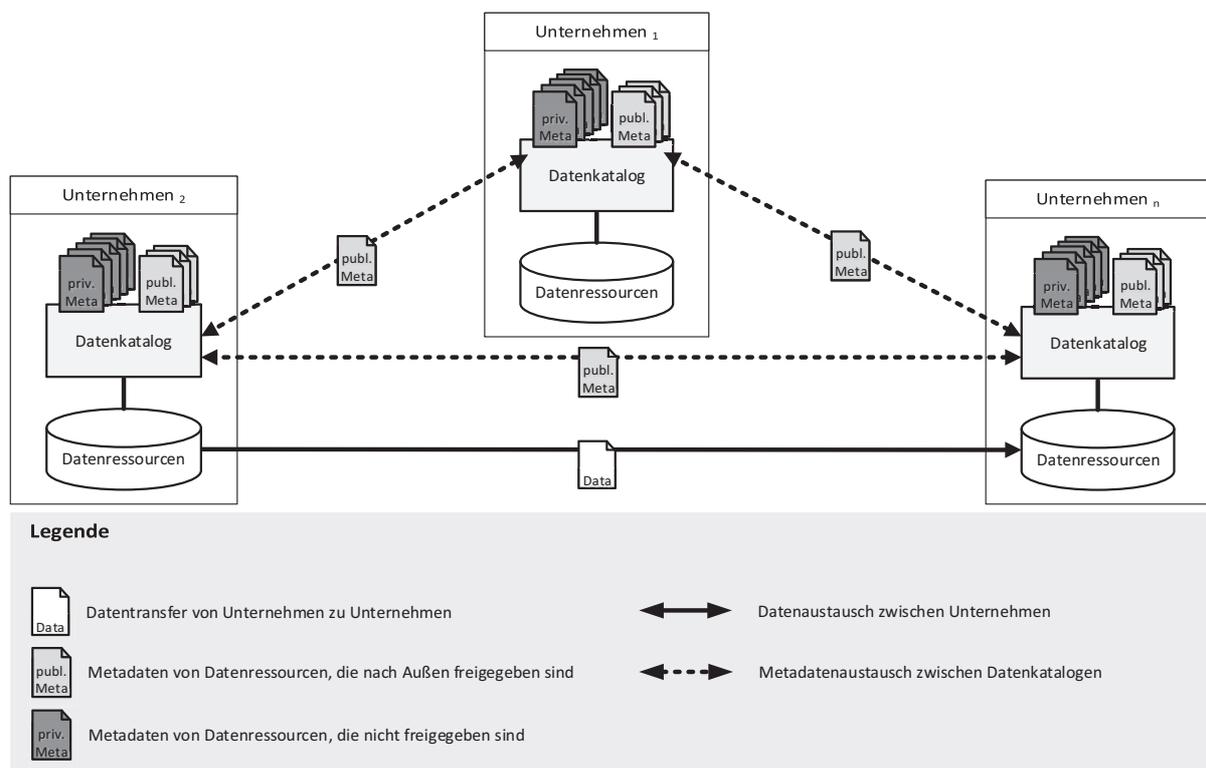


Abb. 2: Zusammenschluss von einzelnen Datenkatalogen zu einem Datenkatalognetzwerk

Teilnehmer des Datenkatalognetzwerkes können in angebundenen Datenkatalogen nach für sich passenden Daten suchen und dann einen beliebigen Kanal für den eigentlichen Datenaustausch zwischen den Unternehmen aushandeln. Die prototypische Implementierung zur Evaluation der Idee nutzt das M4DG um ein einheitliches Verständnis einer Datenressource über alle Datenkataloge zu erzeugen. Die zugrunde liegende Idee hat sich bereits im Kontext des Industrial Data Spaces etabliert [OLA⁺17]. Bei diesem handelt es sich um einen virtuellen Datenraum,

in dem die Teilnehmer ihre Daten, unter Wahrung ihrer Datensouveränitätsansprüche, austauschen können. Dafür werden nur diejenigen Informationen nach außen gegeben, die für einen Datenaustausch notwendig sind. Der Austausch der eigentlichen Daten findet im Industrial Data Space ebenfalls ohne Mittelsmann und somit direkt zwischen den Unternehmen statt.

5 Risikobewertung und Ableitung von Empfehlungen

Ein wichtiger Aspekt bei der Betrachtung von Datenressourcen des Unternehmens und der Einbindung in die Unternehmensprozesse und Geschäftsmodelle stellt die Bewertung von Sicherheits- und Risikoaspekten dar. Aspekte der Sicherheit und des Risikos betreffen dabei nicht nur die reine Speicherung der Daten und den Schutz vor unerlaubtem Zugriff, sondern beeinflussen maßgeblich die Governance, Compliance und die Datensouveränität eines Unternehmens. Diese Arbeit stellt Datenkataloge mit ihrer Funktionalität und passenden Informationsmodellen als eine Grundlage für die Bewertung dieses Aspektes vor. Die Katalogkomponenten sind dabei durch ein geeignetes Informationsmodell für Datengüter im Unternehmen zu ergänzen, welches den Herausforderungen und Potentialen des Datenmanagements in der digitalen Ökonomie genügt [STWO18].

Aufgrund der Dreiecksbeziehung zwischen Information, Geschäftsprozessen und Datenressourcen, in der die Ressourcen die Geschäftsprozesse unterstützen, wird die Komplexität durch digitale Ökosysteme zunehmend größer. Da die Risiken einer Nichtverfügbarkeit über die Datenressourcen direkt die Geschäftsprozesse eines Unternehmens beeinflussen, ist es entscheidend das Risiko für jede Ressource zu kennen, welche diese für die unterstützen Geschäftsprozesse darstellt [Sac08]. Eine Bewertung des Risikos kann dabei über quantitative und qualitative Methoden eruiert werden, welche jeweils Vor- und Nachteile aufweisen [Rot08]. Um eine erste Umsetzung im Rahmen der Überlegungen einer Risikobewertung über die Abbildung von verknüpften Datenkatalogen zu betrachten, wird im Folgenden eine qualitative Bewertung des Risikos beschrieben. Gerade durch die digitale Transformation in den Unternehmen nimmt der unternehmensübergreifende Datenaustausch eine immer wichtigere Rolle ein, in der Daten zwischen den Unternehmen gehandelt werden und Unternehmen eigene Datenressourcen für Partner einer gemeinsamen digitalen Wertschöpfung oder für andere Unternehmen kostenpflichtig als eigenständiges Wirtschaftsgut bereitstellen. Hierdurch ist das IT-Risiko, beispielsweise durch eine Nichtverfügbarkeit, unternehmensübergreifend zu betrachten.

In einem ersten Schritt zur Evaluation des vorgestellten Ansatzes wird daher der Risikobewertungsprozess näher betrachtet und auf den neuen Kontext übertragen. Aufgrund der einfachen Implementierung und Methodik wird in einem ersten Schritt eine qualitative Bewertung betrachtet. Für die qualitative IT-Risikobewertung existieren mehrere Modelle und Methoden, als Beispiele werden in der Folge FMEA/FMECA, CRAMM und NIST SP 800-30 erläutert. Die FMEA (Failure Mode and Effects Analysis) und FMECA (Failure Mode and Effects Criticality Analysis) sind Methoden, deren Ziel die Analyse von Auswirkungen potenzieller Fehler auf die Funktionalität eines Gesamtsystems sind. Im Falle von FMECA ist dies die Unterscheidung verschiedener Stufen der Kritikalität von potenziellen Fehlern. Die FMEA/FMECA Methodik erfordert spezifisches Domänenwissen und die Unterstützung durch spezielle Tools. CRAMM (CCTA's Risk Analysis and Management Methodology), stellt ein standardisiertes Modell für die Analyse und Verwaltung von Risiken dar. CRAMM besteht aus drei Phasen, die bei der Bewertung eines Risikos zu berücksichtigen sind [Ryb06]. Dabei handelt es sich um:

1. Identifizierung und Bewertung der Ressourcen

2. Bewertung der Bedrohung und Anfälligkeiten
3. Handlungsempfehlung für Mechanismen zur Kontrolle und Schutz

Die durch das National Institute for Standards and Technology entwickelte Methode SP 800-30 besteht aus 9 Phasen und ordnet potenzielle Risiken in eine Risikomatrix ein. Diese bildet die Eintrittswahrscheinlichkeit und die Kritikalität, wie in Tabelle 1 dargestellt, ab.

Aufgrund der festgelegten Matrix zur Bewertung der IT-Risiken wird im weiteren Verlauf die SP 800-30 Methode für die weiteren Überlegungen näher betrachtet. Überträgt man die Phasen der NIST Methode auf den in diesem Artikel betrachteten Kontext der verteilten Datenkataloge, ergibt sich folgender Prozess:

1. Identifizierung der im Unternehmen genutzten Datenressourcen und Verknüpfung mit den angeschlossenen Informationssystemen und Prozessen
2. Definition des Evaluierungsumfangs anhand der Limitierung von betrachteten und Sammlung benötigter Informationen
3. Identifizierung potenzieller Risiken für die Systeme
4. Analyse vorhandener und geplanter Mechanismen zur Risikokontrolle
5. Bewertung der Eintrittswahrscheinlichkeiten (niedrig, mittel, hoch) durch die Nutzung von individualisierbaren Metriken innerhalb der Datenkataloge und dem Zugriff auf die hinterlegten Metainformationen
6. Bewertung der Auswirkungen eines potenziellen Risikos (niedrig, mittel, hoch) durch die Definition individueller Metriken in den Datenkatalogen
7. Gesamtbewertung einer Datenquelle durch die Bewertung des potentiellen Risikos durch Einordnung in die durch die Methode definierte Bewertungsmatrix

Die Vorteile einer Risikobewertung in Datennetzwerken wird in den Phasen 8 und 9 sichtbar. Hierbei handelt es sich um die Phasen:

8. Erarbeitung von Schutz- und Kontrollmechanismen zur Minimierung des Risikos
9. Dokumentation von Ergebnissen und Reporting an das Management

Ausgehend von den beschriebenen Möglichkeiten der Datenkataloge und deren Zusammenschluss zu Datennetzwerken kann durch die Definition von Metriken zur Bestimmung eines Risikos nach NIST (niedrig, mittel, hoch) ein Risiko für das gesamte Datennetzwerk ermittelt und kontinuierlich überprüft werden. Dabei müssen keine Details der Verwendung einer Datenquelle in den Unternehmen nach außen bekannt gegeben werden. Hierzu werden die Ergebnisse der Risikobewertung für eine Datenquelle mit dem Datennetzwerk geteilt. Aus dem Durchschnitt der so ermittelten Risiken, lässt sich dann eine Risikobewertung für das gesamte Datennetzwerk ableiten.

Tab. 1: Bewertungsmatrix nach NIST SP 800-30

Eintrittswahrscheinlichkeit	Ergebnis		
	Niedrig (10)	Mittel (50)	Hoch (100)
Hoch (1,0)	10	50	100
Mittel (0,5)	5	25	50
Niedrig (0,1)	1	5	10

Dies kann dazu führen, dass Datenressourcen, die für einige Unternehmen nur ein geringes Risiko aufweisen, für das gesamte Datennetzwerk und andere Unternehmen eine hohe Priorität oder Kritikalität aufweisen. Hiervon muss der Data Owner, als Ersteller und Anbieter der Datenressource, nicht zwangsläufig Kenntnis haben. Durch den Zusammenschluss der Datenkataloge und der möglichen Definition von Metriken zur Berechnung und Darstellung der Risiken im gesamten Netzwerk wird dies, wie in Abbildung 3 dargestellt, jedoch möglich.

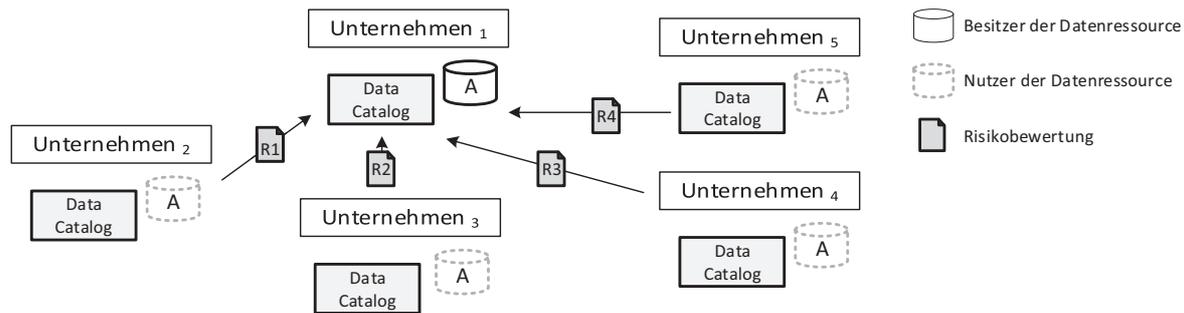


Abb. 3: Zusammenführung individueller Risikobewertungen

Aus den neuen Erkenntnissen, die ein solches Vorgehen ermöglicht, können kollaborativ notwendige Handlungsanweisungen abgeleitet werden. Im Sinne einer hohen Kritikalität einer bestimmten Datenressource wäre dies beispielsweise die redundante Bereitstellung der Datenquelle oder besonderer Absicherungsmechanismen in dem datenhaltenden Unternehmen. Hierbei können anfallende Kosten für diese Maßnahmen an Nutzer, gewichtet nach deren individuellen Risikobewertungen, verteilt werden.

6 Ergebnis der Arbeit

Die Arbeit zeigt ein Konzept zur Verwendung von Datenkatalogen für den Aufbau von unternehmensübergreifenden Datennetzwerken. Die aktuelle Version des Assets zeichnet sich durch seine dezentrale Topologie und den standardisierten Umgang mit den Metainformationen der Unternehmen aus. Hierzu sind bestehende Datenkatalogsysteme zu erweitern und geeignete Metadatenmodelle auszuwählen. Dieses Vorgehen ermöglicht die Bestimmung von Metriken anhand von katalogisierten Metadaten. Durch den beschriebenen Ansatz zur Flexibilisierung, lassen sich Metriken an die Bedürfnisse von Unternehmen anpassen. Am Beispiel einer Analyse über die Nutzung bestimmter Datenressourcen und deren Bewertung innerhalb des Datennetzwerkes wurde dann die Bestimmung eines Ausfallrisikos unter Nutzung dieser Metriken betrachtet. Hierdurch konnte der Mehrwert eines solchen Ansatzes innerhalb des gesamten Datennetzwerkes und die damit einhergehende Möglichkeit der Ableitung von Handlungsempfehlungen aufgezeigt werden. In dieser Arbeit wurde eine qualitative Bewertung eines Risikos betrachtet. Eine weitergehende Evaluierung des Ansatzes sollte auch die Möglichkeiten einer quantitativen Analyse in Betracht ziehen. Das Konzept wurde für die Evaluation des Ansatzes prototypisch implementiert um den grundlegenden Prozess und Nutzung der einzelnen Komponenten des Ansatzes aufzuzeigen. Zukünftige Forschung wird sich mit der weiteren Evaluation der Architektur und dem Abgleich der genannten Ziele für einen globalen Zusammenschluss von Datenkatalogen beschäftigen. Hierbei steht vor allem die Praktikabilität im Vordergrund.

Darüber hinaus wird der Mehrwert unseres Ansatzes für ein unternehmensübergreifendes Datenmanagement anhand von weiteren Praxisbeispielen und Kennzahlen aufzuzeigen sein.

Danksagung

Diese Arbeit wurde durch das Bundesministerium für Bildung und Forschung (BMBF) (InDaSpacePlus – Industrial Data Space Plus, Projekt-Nr. 01IS17031) gefördert.

Literatur

- [AYH⁺17] E. Ahmed, I. Yaqoob, I.A. Targio Hashem, I. Khan, Abdelmuttlib Ibrahim Abdalla Ahmed, Muhammad Imran, Athanasios V Vasilakos: The role of big data analytics in internet of things. *Computer Networks*, 129:459–471, 2017.
- [CKA18] CKAN Association. Comprehensive Knowledge Archive Network (CKAN), 2018. URL: <https://ckan.org/>.
- [Col18] Collibra, Inc.: Collibra Catalog, 2018. URL: <https://www.collibra.com/data-governance-solutions/data-catalog/>.
- [Dat18] Dataverse: The Dataverse Project, 2018. URL: <https://dataverse.org/>.
- [Dur18] DuraSpace: DSpace, 2018. URL: <http://www.dspace.org/>.
- [FPFS⁺17] T. Fedkenhauer, A. Pauer, Y. Fritzsche-Sterr, L. Nagel, Aleksei Resetko: Datenaustausch als wesentlicher Bestandteil der Digitalisierung, 2017. Studie der PricewaterhouseCooper GmbH.
- [GR12] J. Gantz, D. Reinsel: The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future*, 2007(2012):1–16, 2012.
- [HMPR04] A. R. Hevner, S. T. March, J. Park, S. Ram: Design science in information systems research. *MIS quarterly*, 28(1):75–105, 2004.
- [Int03] International Organization for Standardization: Iso 15836:2003 – information and documentation – the dublin core metadata element set, 2003. URL: <https://www.iso.org/standard/37629.html>.
- [Kin07] G. King: An introduction to the dataverse network as an infrastructure for data sharing, 2007.
- [MCP10] F. Maali, R. Cyganiak, V. Peristeras: Enabling interoperability of government data catalogues. In *International Conference on Electronic Government*, pages 339–350. Springer, 2010.
- [MW13] P. Mertens, H.W. Wiczorrek: *Data X Strategien: data warehouse, data mining und operationale Systeme für die Praxis*. Springer-Verlag, 2013.
- [OLA⁺17] B. Otto, S. Lohmann, S. Auer, G. Brost, J. Cirullies, A. Eitel, T. Ernst, C. Haas, M. Huber, C. Jung, J. Jürjens, C. Lange, C. Mader, N. Menz, R. Nagel, H. Pettenpohl, J. Pullmann, C. Quix, J. Schon, D. Schulz, J. Schütte, M. Spiekermann, S. Wenzel: Reference architecture model for the industrial data space, 2017. URL: <http://www.industrialdataspace.org/en/publications/industrial-data-space-reference-architecture-model-2017/>.
- [OÖ16] B. Otto, H. Österle: *Corporate Data Quality*. Springer, 2016.

- [Ott15] B. Otto: Quality and value of the data resource in large enterprises. *Information Systems Management*, 32(3):234–251, 2015.
- [Rot08] A. Rot: IT Risk Assessment : Quantitative and Qualitative Approach. *Proceedings of The World Congress on Engineering and Computer Science 2008*, pages 1073–1078, 2008.
- [Ryb06] M. Ryba: Multidimensional methodology of analysis and management of it systems risk–mir-2m. *doctoral thesis*, 2006.
- [Sac08] S. Sackmann: A Reference Model for Process-Oriented IT Risk Management. *Ecis*, pages 1346–1357, 2008.
- [SJ18] M. Sylverstensen, H.A. Johanson: City data exchange: Lessons learned from a public/private data collaboration, 2018.
- [SLPS13] G. Shukair, N. Loutas, V. Peristeras, S. Sklarß: Towards semantically interoperable metadata repositories: The asset description metadata schema. *Computers in Industry*, 64(1):10–18, 2013.
- [STWO18] M. Spiekermann, D. Tebernum, S. Wenzel, B. Otto: A metadata model for data goods. In *Tagungsband Multikonferenz Wirtschaftsinformatik*, pages 326–337, 2018.
- [TBS⁺03] R. Tansley, M. Bass, D. Stuve, M. Branschofsky, D. Chudnov, G. McClellan, M. Smith, : *The DSpace institutional digital repository system: Current functionality*. IEEE Computer Society, 2003.
- [Wat18] Waterline Data, Inc.: Waterline Data Catalog, 2018. URL: <https://www.waterlinedata.com/data-catalog-details/>.
- [WKLW98] S. Weibel, J. Kunze, C. Lagoze, M. Wolf. Dublin core metadata for resource discovery. Technical report, 1998.
- [ZDSED17] E. Zaidi, G. De Simoni, R. Edjlali, A.D. Duncan: Data catalogs are the new black in data management and analytics, 2017.