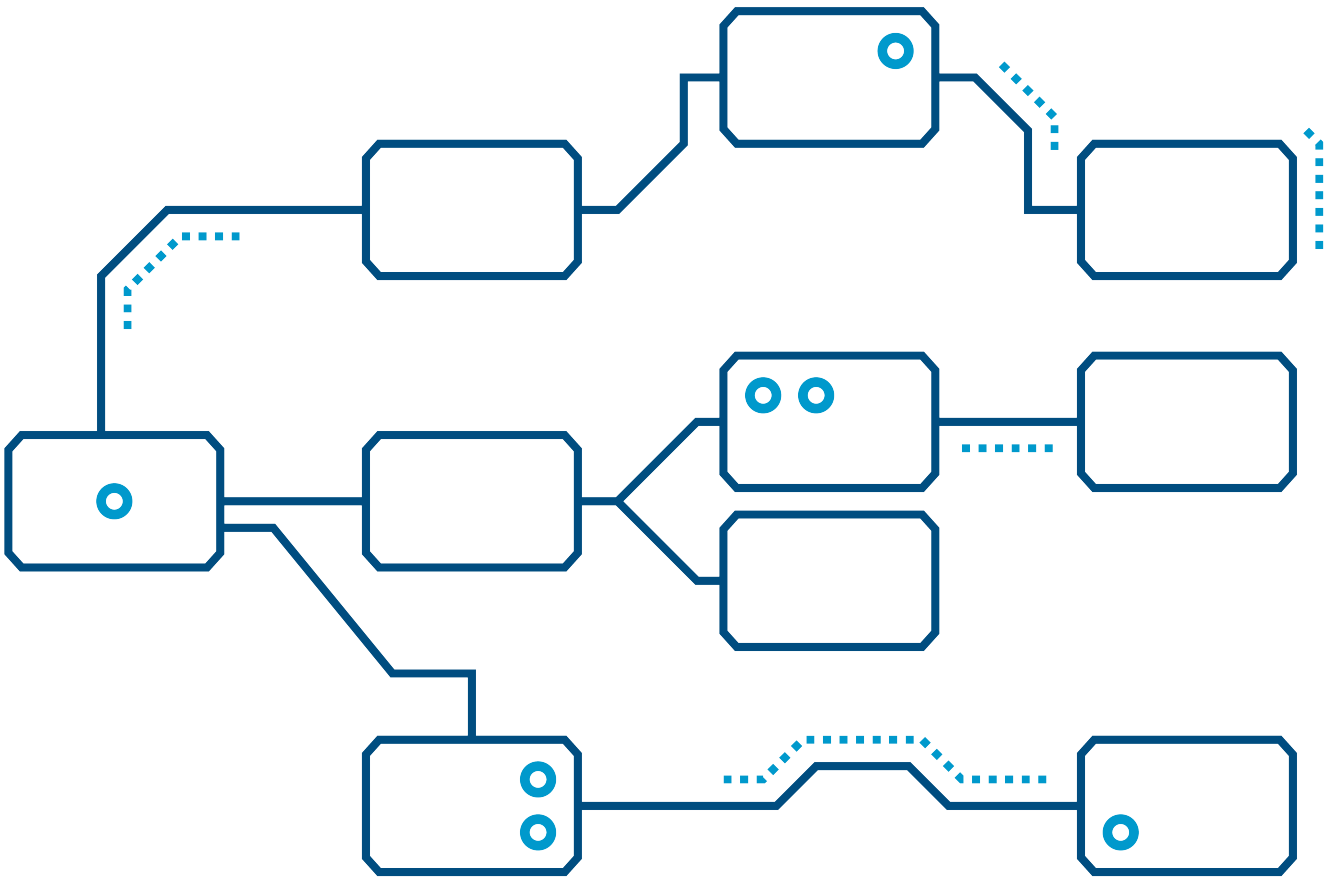
DFN-CERT

DEVISE-MICE Sachbericht Teil 1

Vorhaben: DEVISE
Teilvorhaben: MICE
Laufzeit: 01.04.2021 - 31.09.2024
Förderkennzeichen: 16KIS1351

März 2025



DEVISE-MICE Sachbericht Teil 1

Ursprüngliche Aufgabenstellung

DEVISE-MICE war ein Teilprojekt des Verbundprojekts DEVISE. Ziel von DEVISE war die Messung und Verbesserung der Qualität sicherheitsrelevanter Daten aus den Bereichen Identity und Access Management (IAM) sowie Cyber Threat Intelligence (CTI) mittels eines umfassenden Reifegradmodells und der Anwendung von Datenqualitätsmethoden für die IT-Sicherheit. Das Teilvorhaben MICE fokussierte auf die Messung der Qualität von CTI-Daten und darauf aufbauender Prozesse. Dies sollte zu einem Reifegradmodell inklusive Metriken zur Messung der Qualität sicherheitsrelevanter Daten führen.

Zur Messung der CTI-Datenqualität sollten bestehende Metriken auf ihre Eignung zur Integration in ein Managementsystem für Datenqualität im DEVISE-Reifegradmodell untersucht werden. Hierbei sollten Anforderungen aus den Erkenntnissen von Design und Instanziierung des DEVISE-Reifegradmodells einfließen.

Das Teilvorhaben MICE sollte die Implementierung der entsprechenden Metriken im DEVISE-Reifegradmodell unterstützen. Dabei war zu klären, ob und wie die technische Bewertung der CTI-Datenqualität verlässlich automatisiert und in ein Tool integriert werden kann.

Anknüpfung an den wissenschaftlichen und technischen Stand

Eine im Rahmen des Projekts durchgeführte Literaturrecherche hat gezeigt, dass es zwar einige Arbeiten zum Thema Datenqualität im Allgemeinen gibt, CTI-spezifische Betrachtungen aber nur vereinzelt vertreten sind. Es wurde festgestellt, dass grundlegende Betrachtungen des Feldes fehlen und so die vorhandenen Arbeiten teilweise untereinander inkonsistent oder widersprüchlich sind. Zudem zeigte sich eine große Diskrepanz zwischen den in der Literatur gut vertretenen Dimensionen und Metriken und solchen, die in der Praxis als relevant befunden werden.

An den wissenschaftlichen und technischen Stand konnte so mit strukturgebenden Arbeiten, sowie Methoden zur Auswahl von Qualitätsdimensionen und -metriken angeknüpft werden.

Ablauf des Vorhabens

Die Arbeiten im Projekt DEVISE-MICE verteilen sich auf verschiedene Arbeitspakete: *Geltungsbereich & Design des Reifegradmodells*, *Instanziierung des Reifegradmodells*, *Messung der CTI-Datenqualität*, sowie *Implementierung & Veröffentlichung*.

Für die Analyse des *Geltungsbereiches & Design* des allgemeinen DEVISE-Reifegradmodells wurde eine zielgruppengerechte Erhebung des Geltungsbereiches für CTI-Daten durchgeführt sowie Anforderungen aus der Literatur für diesen Bereich beigetragen.

Im Arbeitspaket *Instanziierung des Reifegradmodells* wurde das allgemeine DEVISE-Reifegradmodell für den Bereich CTI instanziiert. Hierfür mussten die allgemeinen Reifegrade im Kontext von CTI-Daten interpretiert und relevante Qualitätsdimensionen identifiziert werden. Anschließend wurden Indikatoren für die allgemeinen Fähigkeitsgrade der DEVISE-Reifegrade bestimmt und mit dazu passenden konkreten Metriken messbar gemacht. Da

die konkrete Instanziierung von der initialen Zielsetzung des Modells sowie den äußeren Rahmenbedingungen abhängt, wurde für diese Schritte ein allgemeines Vorgehen mit verschiedenen Optionen entworfen und exemplarisch durchgeführt.

Zur *Messung der CTI-Datenqualität* wurden konkrete Qualitätsdimensionen und -metriken aus der Literatur identifiziert und analysiert. Hierbei wurden die genannten Lücken offenbar, sodass eine vereinheitlichte Sammlung von Dimensionen und Metriken erstellt wurde. Da verschiedene Dimensionen nur messbar sind, wenn die verarbeiteten Daten grundlegende Eigenschaften erfüllen, wurde diese Sammlung um entsprechende Anforderungen ergänzt, die bei einer konkreten Instanziierung des Reifegradmodells bereits berücksichtigt werden können. Um der unterschiedlichen Qualität der beschriebenen Metriken Rechnung zu tragen, wurde ein separates Reifegradmodell für diese entwickelt und die resultierende Einstufung in die Sammlung aufgenommen.

Zur Evaluation der vorherigen Schritte wurde ein Demonstrator erstellt, der die Schritte der Instanziierung mittels der Sammlung geeigneter Dimensionen und Metriken unterstützt, die Messung der Datenqualität exemplarisch durchführt und die Ergebnisse visualisiert.

Wesentliche Ergebnisse und Zusammenarbeit mit anderen Forschungseinrichtungen

Während der Projektzeit wurde erfolgreich im Verbund mit den anderen Partnern ein Rahmenwerk für das gemeinsame Reifegradmodell entwickelt. Hierbei wurden wesentliche Ansichten aus der Praxis beigesteuert, um den Geltungsbereich definieren und sowohl Reifegrade als auch Fähigkeitsgrade des übergeordneten Reifegradmodells bestimmen zu können.

Ein zentrales Ergebnis dieser Arbeiten ist, dass der Bereich CTI so heterogen in der Nutzung von Daten ist, dass das Reifegradmodell auf eine Vielzahl von Arten angewendet werden kann und je nach konkretem Einsatz anders instanziiert werden muss. Entsprechend wurde ein Vorgehen entwickelt, welches als Handreichung für eine konkrete Instanziierung dienen und flexibel angepasst werden kann.

Um Nutzer bei der konkreten Instanziierung für ihren Anwendungsfall zu unterstützen, wurde eine Datenbank geschaffen, die Qualitätsdimensionen und -metriken mit ihren jeweiligen Anforderungen an die vorliegenden Daten aufführt. Die Bewertung weiterer, sowie die Definition neuer Metriken, wird durch das neu definierte Reifegradmodell für Metriken unterstützt. In Kombination zeigen die Anforderungen und Reifegradmodelle Wege auf, um die Datengrundlage zu erweitern, die Datenerfassung zu ergänzen und die Messung der Datenqualität selbst zu verbessern mit dem Ziel höhere Reifegrade zu erreichen und damit messbar die Datenqualität zu erhöhen.

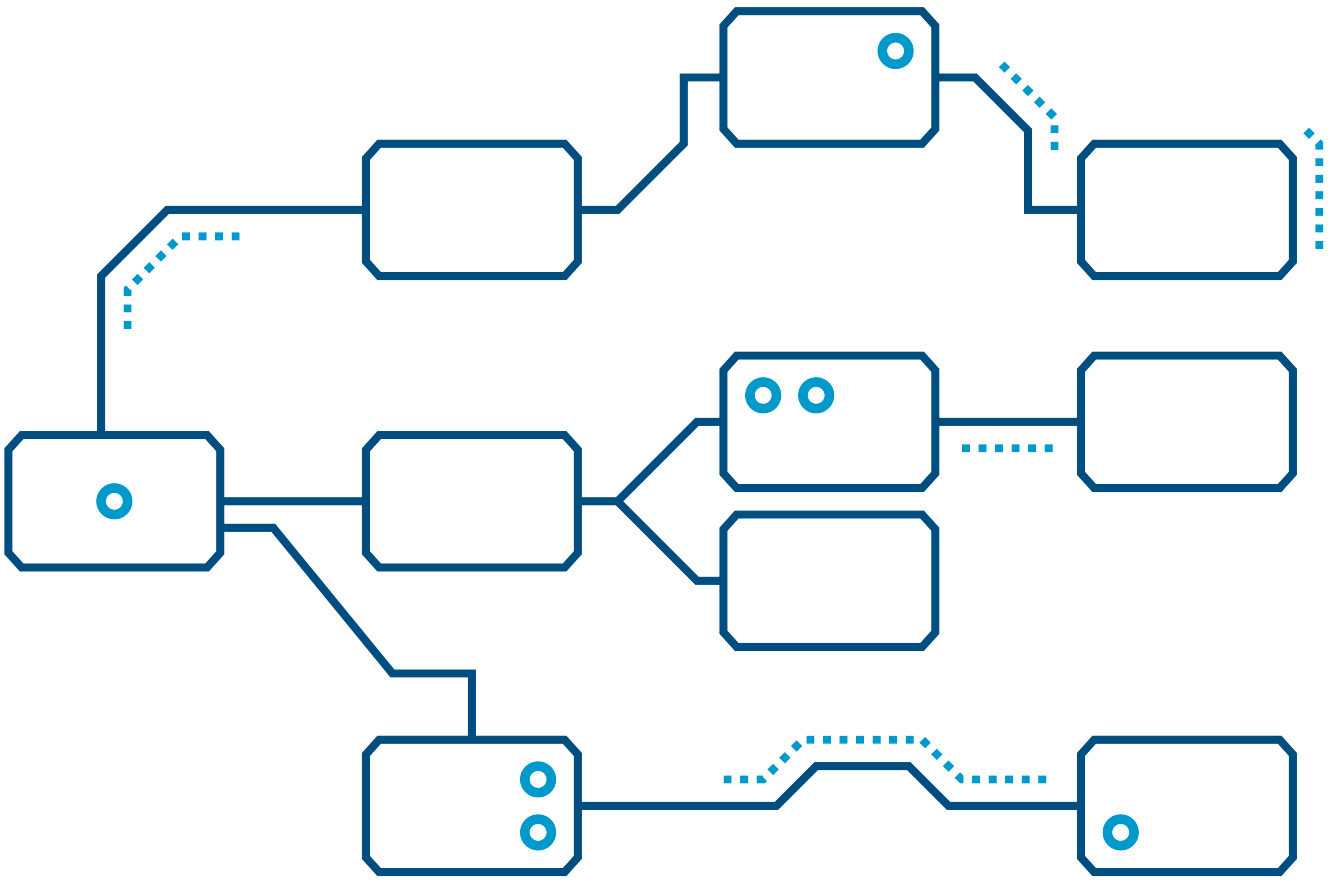
Der implementierte Demonstrator zeigt, wie eine technische Unterstützung der Messung der Datenqualität aussehen kann. Angefangen mit einer interaktiven Datenbank von Qualitätsdimensionen und -metriken, werden auch die frühen Schritte zur konkreten Instanziierung durch den Nutzer unterstützt. Mit Werkzeugen zur automatischen Ausführung von individuellen Pipelines werden Schritte zur Beschaffung von CTI-Daten, Transformation und Speicherung, sowie der Berechnung der Qualitätsmetriken angeboten. Nach der persistenten Speicherung der Ergebnisse, können diese in frei gestaltbaren Dashboards präsentiert werden, um die Ergebnisse des Reifegradmodells zu visualisieren.

DFN-CERT

DEVISE-MICE Sachbericht Teil 2

Vorhaben: DEVISE
Teilvorhaben: MICE
Laufzeit: 01.04.2021 - 31.09.2024
Förderkennzeichen: 16KIS1351

März 2025



Ziel des Vorhabens

DEVISE-MICE war ein Teilprojekt des Verbundprojekts DEVISE. Ziel von DEVISE war die Messung und Verbesserung der Qualität sicherheitsrelevanter Daten aus den Bereichen Identity und Access Management (IAM) sowie Cyber Threat Intelligence (CTI) mittels eines umfassenden Reifegradmodells und der Anwendung von Datenqualitätsmethoden für die IT-Sicherheit. Das Teilvorhaben MICE fokussierte auf die Messung der Qualität von CTI-Daten und darauf aufbauender Prozesse. Dies sollte zu einem Reifegradmodell inklusive Metriken zur Messung der Qualität sicherheitsrelevanter Daten führen.

Zur Messung der CTI-Datenqualität sollten bestehende Metriken auf ihre Eignung zur Integration in ein Managementsystem für Datenqualität im DEVISE-Reifegradmodells untersucht werden. Hierbei sollten Anforderungen aus den Erkenntnissen von Design und Instanziierung des DEVISE-Reifegradmodells einfließen.

Das Teilvorhaben MICE sollte die Implementierung der entsprechenden Metriken im DEVISE-Reifegradmodell unterstützen. Dabei war zu klären, ob und wie die technische Bewertung der CTI-Datenqualität verlässlich automatisiert und in ein Tool integriert werden kann.

Durchgeführte Arbeiten im Vergleich zur Vorhabensbeschreibung

Die Arbeiten im Projekt DEVISE-MICE verteilen sich auf verschiedene Arbeitspakete: *Geltungsbereich & Design des Reifegradmodells*, *Instanziierung des Reifegradmodells*, *Messung der CTI-Datenqualität*, sowie *Implementierung & Veröffentlichung*.

Zur Unterstützung der Definition von *Geltungsbereich & Design* des allgemeinen DEVISE-Reifegradmodells sollte eine zielgruppengerechte Erhebung des Geltungsbereiches für CTI-Daten durchgeführt sowie Anforderungen aus der Literatur für diesen Bereich beigetragen werden.

Mittels ausführlicher Literaturrecherche wurde eine Literaturübersicht zum Projekt beigetragen, die die Basis für die Analyse von Dimensionen und Metriken stellt und in die Metrikübersichten in Anhang A und Anhang B mündet. Die Betrachtungen des Geltungsbereiches werden im Abschnitt *Instanziierung des Reifegradmodells* weiter ausgeführt.

In dem Arbeitspaket *Instanziierung des Reifegradmodells* sollte das allgemeine Reifegradmodell für die geplanten Ausprägungen instanziiert werden. Dies erforderte die *Identifikation und Verifikation der Fähigkeitsgrade* sowie die *Instanziierung des Reifegradmodells für CTI-Daten*.

Die wesentlichen Arbeiten sind hier in der koordinierten Fassung der Fähigkeitsgrade wie im Abschnitt *DEVISE-Reifegradmodell* dargestellt und der Instanziierung des Reifegradmodells für den Bereich CTI wie ausführlich im Abschnitt *Instanziierung des Reifegradmodells* beschrieben.

Zur *Messung der CTI-Datenqualität* sollten Indikatoren zur Bestimmung der Fähigkeitsgrade im Bereich CTI erhoben werden. Hierfür war eine *Analyse relevanter Dimensionen, Metriken und Verfahren zur Messung der CTI-Datenqualität*, sowie die *Entwicklung von Indikatoren* und *Dimensionen, Metriken und Verfahren* erforderlich. Ergänzend sollte eine *Evaluation der CTI-bezogenen Dimensionen, Metriken und Verfahren* erfolgen.

Für dieses Arbeitspaket wurden wesentliche Arbeiten geleistet, um geeignete Messverfahren zu implementieren. Es zeigte sich, dass der CTI-Bereich eine hohe Heterogenität und eine daraus folgende hohe Varianz im Einsatzzweck eines Reifegradmodells für die Datenqualität von CTI aufweist. Aufgrund dessen war ein flexiblerer Ansatz gefordert, um für konkrete Anwendungsfälle angepasste Ergebnisse zu erzielen. Somit führten die Analysen zu einer *Ausführlichen Sammlung von CTI-Qualitätsdimensionen und Metriken*, welche im gleichnamigen Abschnitt vorgestellt wird. Um die konkrete Implementierung von Messverfahren zu unterstützen, bieten die Verfahren aus Abschnitt *Bewertung von Metriken mittels eines eigenen Reifegradmodells* und Abschnitt *Identifikation von Anforderungen an Dateneigenschaften von Qualitätsdimensionen* Ansätze.

Das Arbeitspaket *Implementierung & Veröffentlichung* umfasst Arbeiten zur Implementierung des Reifegradmodells und die Veröffentlichung der Projektergebnisse. Hier sollte die *Integration CTI-bezogener Indikatoren, Dimensionen, Metriken und Verfahren in die Referenzimplementierung* und eine *Validierung der Referenzimplementierung für CTI-Daten* erfolgen. Die Ergebnisse sollten in *Standardisierung und wissenschaftliche Veröffentlichungen* einfließen.

Die in den vorausgehenden Arbeitspaketen entwickelten Verfahren wurden in einem Demonstrator umgesetzt, der im Abschnitt *Framework for Assessing CTI Quality (FACQ)* näher dargestellt wird. Die wissenschaftliche Veröffentlichung von Teilen der Arbeit ist zum Zeitpunkt des Projektendes in Vorbereitung.

DEVISE-Reifegradmodell

Das DEVISE-Reifegradmodell für die Qualität von sicherheitsrelevanten Daten soll Aufschluss über den Entwicklungsstand dieser Daten liefern. Es soll ein generelles Reifegradmodell sein, das für verschiedene Ausprägungen sicherheitsrelevanter Daten nutzbar ist.

Als Basis für die spezialisierten Instanziierungen bietet das vom Konsortium als Rahmenwerk entwickelte DEVISE-Reifegradmodell grundlegende Strukturen, um die Reife des Datenqualitätsmanagements sicherheitsrelevanter Daten zu evaluieren. Zur Instanziierung konkreterer Ausprägungen für CTI und IAM müssen die Basisstrukturen weiter ausgearbeitet und entsprechend dem Anwendungsbereich interpretiert werden.

Das Reifegradmodell besteht aus mehreren Ebenen, die in der Übersicht in Abbildung 1 dargestellt sind. Auf oberster Ebene befinden sich die Reifegrade (Security Quality Maturity - SQM) “Initial”, “Reaktiv”, “Proaktiv” und “Kollaborativ”. Reifegrade spiegeln den groben Entwicklungsstand des betrachteten Bereichs wider und sind aufeinander aufbauend. Somit ist ein möglichst hoher Reifegrad anzustreben. Die Reifegrade sind eine Ebene tiefer in Fähigkeitsgrade (Security Quality Capabilities - SQC) strukturiert. Jeder Reifegrad bündelt hier die Fähigkeitsgrade “State Recognition”, “Assessment” und “Improvement”. Die Erreichung von Fähigkeitsgraden wird über Indikatoren (Security Quality Indicators - SQIs) angezeigt, die wiederum mittels Qualitätsdimensionen oder Merkmalen gemessen werden. Qualitätsdimensionen sind Qualitätseigenschaften der Daten wie zum Beispiel Aktualität oder Korrektheit. Weitere Merkmale können Kriterien sein wie zum Beispiel ob CTI-Daten automatisiert verarbeitet werden. Indikatoren sind Eigenschaften, die als Anforderungen zur Erreichung von Reifegraden dienen. Sind die definierten Schwellwerte für alle Indikatoren eines Fähigkeitsgrades erreicht, so kann dieser Bereich als erfüllt angesehen werden. Werden alle Fähigkeitsgrade eines Reifegrades abgedeckt, so gilt dieser Reifegrad als erreicht.

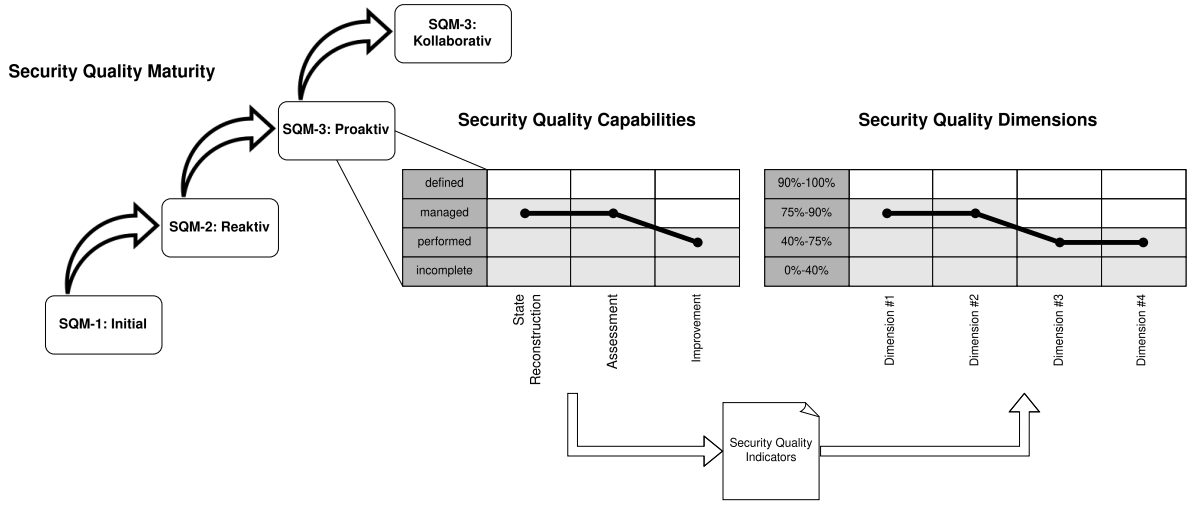


Abbildung 1: DEVISE-Reifegradmodell Übersicht

Zur Instanziierung des Reifegradmodells hat das Konsortium einen Prozess entwickelt, der den Schritt vom übergeordneten Modell zu einer konkreten Ausprägung definiert. Dieser Prozess ist in Abbildung 2 gezeigt.

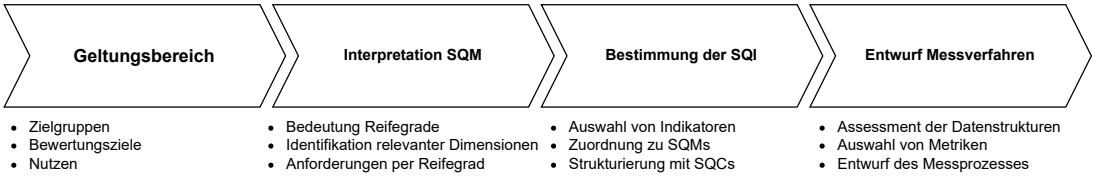


Abbildung 2: DEVISE-Reifegradmodell Instanziierungsprozess

Der Prozess zur Instanziierung sieht vor, dass zunächst der Geltungsbereich weiter definiert wird. Hierzu zählt die Bestimmung von Zielgruppen, an die sich das Reifegradmodell wenden, sowie die Bestimmung von Bewertungszielen und Nutzen, der aus dem Modell gezogen werden soll. Im zweiten Schritt werden die Reifegrade (SQM) interpretiert. Hierbei ist klarzustellen, was die SQM für den Anwendungsbereich bedeuten. Außerdem werden relevante Dimensionen identifiziert und die Anforderungen pro Grad erhoben. Der nächste Schritt dient dazu Indikatoren (SQI) den Dimensionen und Anforderungen zuzuordnen. Hierfür werden SQI ausgewählt, die für die Messung der Dimensionen geeignet sind und so den Entwicklungsstand hinsichtlich der gestellten Anforderungen anzeigen. Diese SQI werden dann mittels der Fähigkeitsgrade (SQC) aggregiert und konkret den SQM zugewiesen. Im finalen Schritt werden Messverfahren entwickelt, welche die SQI abdecken können. Eine Analyse der vorliegenden Daten und ihrer Strukturen liefert eine Reihe nutzbarer Metriken, aus denen konkrete Metriken zur Abdeckung der SQI ausgewählt werden. Hierauf aufbauend können schließlich Prozesse zur Messung der Datenqualität entworfen werden.

Instanziierung des Reifegradmodells

Der Bereich CTI weist eine sehr hohe Heterogenität auf, sowohl in der Art der verwendeten Daten als auch in der Zielsetzung der etablierten Prozesse verschiedener Anwender. Somit sind auch die Einsatzmöglichkeiten des anvisierten Reifegradmodells sehr variabel. Um dieser Vielfalt gerecht zu werden, wird ein flexibles Modell präsentiert, welches auf den konkreten Anwendungszweck parametrisiert werden kann. Eine beispielhafte Darstellung demonstriert dabei den Prozess der Anpassung.

Die nachfolgenden Abschnitte folgen dem Prozess der Instanziierung. Zunächst erfolgt die Definition des Geltungsbereich, sowie die Interpretation der SQMs. Anschließend wird eine Auswahl und Gewichtung der SQIs getroffen und der Entwurf der Messverfahren für die SQIs entwickelt.

Definition des Geltungsbereich

Bei der Definition des Geltungsbereiches werden Zielgruppen, Bewertungsziele und Nutzen des Reifegradmodells betrachtet.

Die Auswahl der **Zielgruppen** bietet zwei Ausprägungen: strategisch und technisch.

Strategisch ausgerichtet wird das Reifegradmodell reifezentriert genutzt – sprich die finalen Reifegrade sind von wesentlicher Bedeutung – und unterstützt vor allem Anwender in leitender Position (z. B. CISOs, Datenschutzbeauftragte und das Management).

Bei technischer Ausrichtung wird das Reifegradmodell dagegen metrikzentriert genutzt, sodass einzelne Messungen oder Teilbereiche von höherer Bedeutung sind und hieraus konkrete Maßnahmen zur Verbesserung abgeleitet werden sollen. Dies kann etwa von Administratoren, SOCs oder Vorfallsbearbeitungsteams genutzt werden.

Als **Bewertungsziele** bieten sich im Anwendungsfall CTI mehrere Bereiche an. Hierzu zählen Prozesse (Quellen aktuell halten, CTI Filtern, neue CTI sichten, usw.), die Verarbeitung von CTI-Daten (Datenfeeds beziehen, Priorisierung von CTI-Daten, Korrelation, usw.), die CTI-Daten selbst (Qualität vorhandener Daten, Anwendbarkeit von Daten, Abdeckung verschiedener CTI-Kategorien, usw.) sowie der Wissensstand des Personals (Kompetenz, Awareness, usw.).

Je nach Zielsetzung können verschiedene **Nutzen** aus dem Reifegradmodell gezogen werden:

Entscheidungen Mithilfe des Reifegradmodells können Entscheidungen oder Ressourcenverteilungen unterstützt werden. Es kann zur Argumentation der Notwendigkeit herangezogen werden oder Maßnahmen zielgerichteter eingesetzt werden.

Verbesserungen Das Reifegradmodell kann zur Verbesserung im Umgang mit sicherheitsrelevanten Informationen führen. Es kann genutzt werden, um Qualitätsmanagement zu betreiben und Potenziale aufzudecken. Schließlich kann der Erfolg der Maßnahmen anhand des Reifegradmodells gemessen werden.

Zertifizierung In entsprechenden Bereichen kann das Reifegradmodell als Teil einer Zertifizierung herangezogen werden.

Rechtfertigung Das Reifegradmodell kann zur Rechtfertigung genutzt werden, um Gelder für Projekte zur Verbesserung der Nutzung von sicherheitsrelevanter Daten zu bekommen oder Managemententscheidungen zu begründen.

Vergleichbarkeit Das Reifegradmodell kann genutzt werden, um verschiedene Vergleiche zu führen. So können die Quellen für sicherheitsrelevante Informationen im Hinblick auf ihren Beitrag zum Reifegradmodell verglichen werden. Auch Maßnahmen können dahingehend verglichen werden wie sie sich auf die Einstufung im Reifegradmodell auswirken. Zuletzt kann man die eigene Einstufung mit der Konkurrenz vergleichen, sofern deren Einstufung bekannt ist (z. B. wenn als Werbung oder zur Zertifizierung eingesetzt).

Vertrauen Das Reifegradmodell kann zur Schaffung von Vertrauen verwendet werden. Zum einen in die eigene Nutzung der sicherheitsrelevanten Daten, zum anderen als Werbung, um Vertrauen gegenüber Kunden zu erzeugen.

Lastreduzierung Wenn ein Reifegradmodell Verbesserungspotentiale aufzeigt oder effizientere Prozesse in der Verarbeitung von Daten ermöglicht, kann dies zu einer Lastreduzierung führen. So gibt einem ein gut definiertes Reifegradmodell die Möglichkeit nicht notwendige Arbeiten zu erkennen.

Als Fallbeispiel betrachten wir nun ein Unternehmen, welches aufgrund von CTI-Daten und eigener Sensorik Warnmeldungen an seine Kunden verschickt. Zur Definition des Geltungsbereichs sind zuerst grundlegende Entscheidungen zu treffen: Das Modell soll in diesem Fall **technisch** genutzt werden, um die **Qualität der CTI-Daten**, die die Grundlage der Warnmeldungen sind, zu **verbessern**.

Interpretation der SQMs

Die im allgemeinen Reifegradmodell definierten Reifegrade (SQM) müssen für die Instanziierung nun weiter interpretiert werden. Gleichzeitig werden für die Reifegrade relevante Qualitätsdimensionen identifiziert. Die Charakterisierung der Dimensionen ist hierbei qualitativer und nicht quantitativer Natur. Die in diesem Abschnitt identifizierten Qualitätsdimension sind für die gewählte Interpretation der SQM direkt relevant, sie sind allerdings nicht abschließend und können je nach konkretem Einsatzgebiet und -zweck erweitert werden. Weitere Dimensionen mit näherer Beschreibung und Definition finden sich in Anhang A & Anhang B.

Bevor die Qualitätsdimensionen den SQM zugeordnet werden, sollen diese kurz vorgestellt werden. *Accuracy* misst die Genauigkeit der Daten meist bezüglich syntaktischer oder semantischer Korrektheit, das heißt ob Formate eingehalten werden oder die Realität korrekt abgebildet wird. *Actionability* sagt aus, ob auf Grundlage allein der Daten direkt Handlungen abgeleitet werden können. *Believability* spiegelt den Grad an Glaubwürdigkeit oder Vertrauen in die Daten wider. *Completeness* bewertet die Vollständigkeit der Daten, meist bezogen auf die Menge verfügbarer Daten oder die Abdeckung ausgefüllter Datenfelder. *Relevance* beschreibt wie relevant, wichtig oder treffend die Daten für die Organisation sind. *Timeliness* bewertet die Verzögerung mit der Daten zur Verfügung stehen.

Die Reifegrade können nun wie folgt interpretiert werden:

- **SQM-0 Initial**: Datenqualitätsmanagement für CTI ist nicht vorhanden. Es werden zwar Daten gesammelt, allerdings ist unklar in welchem Umfang und wie die Qualität zu bewerten

ist. Es findet keine Automatisierung statt und es ist kein Prozess zur Qualitätsverbesserung von CTI etabliert.

- **SQM-1 Reaktiv**: Der Reifegrad SQM-1 Reaktiv zeugt davon, dass CTI-Daten umfassend gesammelt werden und zugänglich sind. Die Datenqualität ermöglicht reaktive (meist manuelle) Maßnahmen unter Verwendung der verfügbaren Daten. Hierbei werden vor allem leichte Ansprüche an die Relevanz und die Aktualität der Daten gestellt. Aufseiten der Prozesse ist der Zugriff auf die gesammelten CTI-Daten möglich ist, sodass diese manuell verarbeitet werden können. Es werden erste Initiativen zur Verbesserung der Datenqualität und Nutzung von CTI angestrengt.

  - Relevante Dimensionen: Relevance und Timeliness
  - Relevante Prozesse: manuelle Verarbeitung von CTI

- **SQM-2 Proaktiv**: Ist die Sammlung und Nutzung von CTI auf einen erheblichen Teil des Geltungsbereichs ausgeweitet und hat die Qualität der Daten ein Niveau erreicht, auf dem diese automatisiert genutzt werden können, ist der Reifegrad SQM-2 Proaktiv erreicht. Hierbei ist die Datenqualität so hoch, dass neben reaktiven Maßnahmen auch proaktive unterstützt werden können. Dafür werden strengere Anforderungen an die Aktualität und Relevanz gestellt als zuvor, da die Daten nicht durch einen Experten manuell verarbeitet werden. Es kommen Anforderungen an die Genauigkeit, Glaubwürdigkeit und Aktionsfähigkeit hinzu, sodass automatisierte Maßnahmen ermöglicht werden. Um effektive Maßnahmen zu ermöglichen, ist auch ein gewisses Maß an Vollständigkeit erforderlich. Die iterative Verbesserung der Datenqualität wird als Prozess stärker umgesetzt.

  - Relevante Dimensionen: Relevance, Timeliness, Accuracy, Actionability, Believability, Completeness
  - Relevante Prozess: automatisierte Verarbeitung von CTI

- **SQM-3 Kollaborativ**: Die vorhandenen Daten für CTI sind so umfangreich und qualitativ hochwertig, dass diese Dritte in ihren Sicherheitsmaßnahmen unterstützen können. Entsprechend kommt neben der Verbesserung der bisherigen Anforderungen eine Öffnung gegenüber Externen hinzu, was zu Synergien innerhalb der Community führt. Hierbei wird die Kollaboration mit anderen Unternehmen zur gemeinsamen Verbesserung der übergreifenden Sicherheit angestrebt. Es werden stärkere Anforderungen an Accuracy und Timeliness gestellt als vorher, damit diese Daten auch nach dem Teilen noch für andere nützlich sind.

  - Dimensionen: Accuracy, Timeliness
  - Prozess: Teilen

Die Interpretation der Reifegrade lässt sich für unser Fallbeispiel weiter konkretisieren. So charakterisieren die SQM hier qualitativ die genutzten Daten.

SQM-0 Initial Der initiale Reifegrad bringt keine weiteren Anforderungen mit sich.

SQM-1 Reaktiv Es existiert eine Übersicht verfügbarer und potenziell interessanter Datenquellen. Daten werden nach Aktualität gefiltert und regelmäßig manuell auf ihre Nützlichkeit für die Kunden evaluiert.

SQM-2 Proaktiv Eine Vielzahl relevanter CTI-Bereiche können zeitnah abgedeckt werden, sodass eine automatisierte Verarbeitung möglich wird. Die Daten werden auf ihre Tauglichkeit und Glaubwürdigkeit evaluiert, um einen möglichst hohen Nutzen für die Kunden auch bei automatischer Verarbeitung zu garantieren.

SQM-3 Kollaborativ Die verarbeiteten CTI sind von so hoher Qualität, dass diese auch mit anderen Einrichtungen geteilt werden können. Zusätzlich werden eigene CTI erhoben, um das gesamte Feld zu ergänzen. Es wird aktiv mit anderen Organisationen zusammen gearbeitet.

Die beschriebenen Fähigkeitsgrade zeigen, wie sich eine Dienstleistung der gewünschten Art je nach Reifegrad differenzieren würde. Während ein Reifegrad SQM-0 einer eher wahllosen und unregelmäßigen Weiterleitung von Informationen an die Kunden entsprechen würde, die kaum als separate "Dienstleistung" zu erkennen ist, zeigen die höheren Reifegrade eine sukzessive Verbesserung der inhaltlichen Qualität der versendeten Daten an.

Auswahl und Gewichtung der SQIs

Zur weiteren Ausgestaltung des Reifegradmodells müssen Indikatoren (SQIs) definiert werden. Pro Reifegrad werden die SQIs den SQCs zugewiesen, um eine weitere Ebene der Strukturierung zu schaffen.

Die hohe Heterogenität in Verwendung und Nutzen von CTI schafft eine hohe Bandbreite von möglichen Lösungen im konkreten Fall. Um dieser hohen Variabilität gerecht zu werden, präsentieren wir eine Liste möglicher generalisierter Indikatoren, welche für konkrete Instanziierungen ausgewählt werden können. Hierbei werden Platzhalter wie beispielsweise `<Dimension>` für Datenqualitätsdimensionen verwendet, um eine kompakte, generalisierte Darstellung zu erreichen. Neben den Erfüllungskriterien werden auch deren Typen in Klammern angegeben. Hierbei bedeutet boolesch, dass es nur zwei Werte für erfüllt oder nicht erfüllt gibt, was häufig durch das Vorhandensein eines Dokumentes oder per Fragebogen bestimmt werden kann. Hingegen bedeutet quantitativ, dass eine nummerische Messung erforderlich ist und ein Schwellwert zur Erreichung festgesetzt werden muss. Ist dies nicht möglich, so können die quantitativen Indikatoren auch durch schwächere, qualitative ersetzt werden.

SQI	Erfüllungskriterium
Anforderungen an `<Dimension>` für `<Zweck>` sind bekannt	Dokument mit Anforderungen ist vorhanden (boolesch)
Manuelle Bestimmung von `<Dimension>`	Dokumentierte Schritte zur Messung der Dimension (boolesch)
`<Dimension>` ist über/unter `<X>`%	Wert der Messung liegt über/unter Schwellwert (quantitativ)
`<Dimension>` wir durch Feedbackschleife evaluiert	Technik- oder prozessgestützte Feedbackschleife vorhanden (boolesch)
Benötigte CTI-Arten und Bereiche sind identifiziert	Übersicht über benötigte Arten und Bereiche vorhanden (boolesch)
`<X>`% der CTI-Bereiche sind abgedeckt	Übersicht über Quellen und Bereiche `#Abgedeckt / #Benötigt` (quantitativ)

SQI	Erfüllungskriterium
Manueller Zugriff auf CTI ist möglich	Dokumentierter Zugriff auf CTI vorhanden (boolesch)
Quellen werden auf Tauglichkeit evaluiert	Prozess zur Evaluation ist definiert (boolesch)
Automatisierung der Verarbeitung von CTI	Tools zur automatischen Verarbeitung etabliert (boolesch)
Automatische Prozesse zur Bestimmung relevanter CTI	Tools zur automatischen Bestimmung der Relevanz (boolesch)
Es werden weitere Quellen aktiv erschlossen	Prozesse sind definiert, um neue Quellen zu erschließen (boolesch)
Es werden eigene CTI-Daten erhoben	CTI-Daten werden durch eigenen Betrieb erzeugt (boolesch)
Organisation ist Teil von Sharing Communities	Organisation beteiligt sich an Sharing Groups (boolesch)
Es existieren Techniken zur Kollaboration	Technik- oder prozessgestützte Kollaboration ist etabliert (boolesch)
Es werden CTI-Daten geteilt	Es werden relevante CTI-Daten mit anderen geteilt (boolesch)
Es werden CTI-Daten korreliert	Es existieren Techniken zur Korrelation von CTI-Daten (boolesch)

Zurück zu unserem Fallbeispiel und einer möglichen Auswahl von SQIs und SQCs. Die Anpassung an andere Anwendungsfälle erfolgt analog unter Verwendung der generischen Liste der SQI und der Sammlung der identifizierten Dimensionen in Anhang B.

SQM-0 Initial:

Der initiale Reifegrad bringt keine Anforderungen mit sich.

SQM-1 Reaktiv:

SQC-1	**State Reconstruction** Benötigte CTI-Arten und -Bereiche sind identifiziert 20% der CTI-Bereiche sind abgedeckt Manueller Zugriff auf CTI möglich
SQC-2	**Assessment** Anforderungen an die Aktualität zur effektiven Verarbeitung sind bekannt *Timeliness* muss unter 5 Tagen liegen Manuelle Bestimmung der *Relevance* ist möglich *Relevance* der CTI liegt bei über 40%
SQC-3	**Improvement** Quellen werden bezüglich ihrer Tauglichkeit evaluiert

SQM-2 Proaktiv:

SQC-1	**State Reconstruction** Datensammlung deckt 85% der definierten Bereiche ab Automatisierte Verarbeitung von CTI *Completeness* über 60%
SQC-2	**Assessment** Anforderungen an *Timeliness* zur automatischen Verarbeitung bekannt *Timeliness* besser als 3 Tage zur automatischen Verarbeitung Automatische Prozesse zur Bestimmung relevanter CTI *Relevance* muss mindestens 60% betragen *Believability* ist bestimmbar für Quellen Proaktive genutzte CTI haben *Believability* über 80% *Actionability* ist bestimmbar Über 50% der CTI sind *actionable* *Semantic Accuracy* ist hoch, vor allem *False Positive Rate* unter 15%
SQC-3	**Improvement** Es werden weitere Quellen aktiv erschlossen *Believability* wird durch Feedbackschleifen weiter evaluiert

SQM-3 Kollaborativ:

SQC-1	**State Reconstruction** Gebiete von CTI sind nahezu komplett abgedeckt Es werden eigene CTI erhoben
SQC-2	**Assessment** Die *Timeliness* ist besser als 2 Tage — selbst nutzbar nach der Verzögerung durch den Datenaustausch *Semantic Accuracy* ist über 80% *Syntactic Accuracy* eigener CTI ist nahe 100%
SQC-3	**Improvement** Unternehmen ist Teil von Sharing Communities Es existieren Techniken zur Kollaboration Es werden CTI mit anderen geteilt Es werden Techniken zur Korrelation genutzt

Entwurf der Messverfahren für die SQIs

Der breite Raum an konkreten Szenarien steht auch hier einer allgemeingültigen Definition der Messverfahren entgegen. Um diesem Problem Rechnung zu tragen, werden in der Literatur verfügbare Metriken untersucht und auf ihre Reife analysiert. In Verbindung mit der erstellten

Sammlung von Dimensionen und deren Anforderungen an die verarbeiteten Daten entsteht so eine solide Basis für die Auswahl konkreter Messverfahren und damit die finale Instanziierung des Reifegradmodells.

Messung der CTI-Datenqualität

Aufbauend auf den beschriebenen, theoretischen Grundlagen zur Instanziierung beschäftigen sich die folgenden Abschnitte mit der konkreten Messung der CTI-Datenqualität. Zu diesem Zweck werden in der Literatur beschriebene Dimensionen und Metriken identifiziert und gefundene Definitionen vereinheitlicht. Diese Analyse führt zu einem neuen Reifegradmodell für Metriken an sich und offenbart Anforderungen der Metriken an die verarbeiteten Daten.

Ausführliche Sammlung von CTI-Qualitätsdimensionen und Metriken

Für die Auswahl der notwendigen SQIs im Rahmen der Instanziierung bietet eine Übersicht bekannter Dimensionen und Metriken eine signifikante Erleichterung. Um eine aktuelle Sicht auf den Stand der Forschung zu bekommen, wurden die Ergebnisse einer systematischen Literaturanalyse in einer Datenbank gesammelt. Durch die Analyse der Definitionen können mehrere Probleme in der relevanten Literatur festgestellt werden:

- Die Begriffe Qualitätsdimension und Qualitätsmetrik werden teilweise synonym verwendet, obwohl sie unterschiedliche Aspekte der Datenqualität beschreiben.
- Bei den Namen, die Dimensionen in den Quellen bekommen, gibt es Überschneidungen. Dabei sind die Definitionen nicht notwendigerweise identisch, sodass die Namen allein nicht aussagekräftig sind.
- Semantisch identisch beschriebene Dimensionen werden teils unterschiedlich benannt.

Um einer missverständlichen Verwendung vorzubeugen, enthält die Datenbank klare Definitionen von Qualitätsdimension und Qualitätsmetrik. Dimensionen mit identischer Definition werden zusammengefasst und eindeutig benannt, bevor ihnen die relevanten Metriken zugeordnet werden. Da CTI-spezifische Qualitätsdimensionen nicht ausreichend in der Literatur vertreten sind, wurden sie um Dimensionen aus allgemeinen Datenqualitätsquellen ergänzt. Um den Fokus auf den Bereich CTI zu erhalten, wurden dabei lediglich diejenigen Dimensionen in die finale Auswahl übernommen, für die sich wenigstens eine auf CTI-Daten angewendete Metrik in der Literatur findet. Insgesamt umfasst die Datenbank so 146 Qualitätsmetriken und 16 Qualitätsdimensionen, die sich dem Bereich CTI sinnvoll zuordnen lassen. Eine Übersicht über die Einordnung der Metriken in die vereinheitlichten Qualitätsdimensionen findet sich in Anhang B.

Um Anwendern einen einfachen Zugang zu den gesammelten Metriken zu ermöglichen, enthält die DEVISE Datenqualitätsplattform eine nutzerfreundliche Schnittstelle für die Datenbank (siehe Abschnitt FACQ).

Bewertung von Metriken mittels eines eigenen Reifegradmodells

Die erstellte Sammlung bietet bereits einen guten Überblick über den Stand der Literatur und die darin enthaltenen Metriken. Dabei werden deutliche Qualitätsunterschiede bezüglich der

Nutzbarkeit der beschriebenen Metriken offenbar. Um dies weiter untersuchen und quantifizieren zu können, wurde ein separates Reifegradmodell entwickelt welches die Reife der Definition von Metriken bewertet. Die gesammelten Metriken wurden mit dem entwickelten Reifegradmodell bewertet und die Ergebnisse analysiert. Dieses Reifegradmodell für Metriken ist unabhängig vom DEVISE-Reifegradmodell und in sich abgeschlossen. Es dient allerdings als Werkzeug zur Unterstützung bei der Auswahl von geeigneten Metriken und der Entwicklung entsprechenden Messverfahren.

Der Entwurf des Modells folgt dem von Bruin u. a. (2005) beschriebenen Rahmenwerk zur Entwicklung von Reifegradmodellen. Hierbei wurden die im Rahmenwerk definierten Phasen *Scope*, *Design*, *Populate* und *Test* durchlaufen, bei denen jeweils Parameter und Kennwerte für das Reifegradmodell gesetzt werden. Die Einteilung der Reifegrade orientiert sich an den Empfehlungen des weit verbreiteten Capability Maturity Model Integration (CMMI) für die Entwicklung von Produkten und Dienstleistungen (CMMI-DEV)(CMMI Product Team 2010). Das Reifegradmodell sieht 4 Stufen vor mit – bis auf das initiale Level – jeweils zwei Anforderungen, um die Stufe zu erreichen. Die Reifegrade sind in nachfolgender Tabelle dargestellt.

Reifegrad	Anforderungen	Beschreibung
Level 0: Initial	L0.RQ1: Name	Metrik namentlich erwähnt
Level 1: Managed	L1.RQ1: Description	Die gemessenen Qualitätsaspekte werden beschrieben
	L1.RQ2: Data description	Die gemessenen Daten werden beschrieben
Level 2: Defined	L2.RQ1: Formula	Eine mathematische Formel wird gegeben oder kann leicht abgeleitet werden
	L2.RQ2: Data characterized	Anforderungen an Eigenschaften der Daten werden charakterisiert
Level 3: Quantitative	L3.RQ1: Mathematical sophistication	Die Definition der Metrik ist mathematisch ausgereift
	L3.RQ2: Context	Der Kontext der Anwendung ist gegeben

Die Details der Reifegrade und ihrer Anforderungen sind in dem derzeit noch unveröffentlichten Paper “About the Maturity of Data Quality Measurement in Cyber Threat Intelligence” ausgeführt. Eine Übersicht über die bewerteten Metriken findet sich in Anhang B.

Bei der Auswertung der Reife der gesammelten Metriken zeigen sich Diskrepanzen zwischen der Abdeckung in der Literatur und der von Praktikern wahrgenommenen Wichtigkeit von Qualitätsdimensionen. Nicht alle als wichtig wahrgenommenen Dimensionen sind auch gut mit reifen Metriken versorgt. Das Reifegradmodell bietet zudem Anhaltspunkte zur Verbesserung von zukünftigen Definitionen von Qualitätsmetriken mit dem Ziel eine höhere direkte Nutzbarkeit der Metriken zu erreichen.

Im Rahmen des DEVISE-Reifegradmodells wird das Reifegradmodell für Metriken als Hilfsmittel genutzt, um aus der Vielzahl von verfügbaren Metriken solche auswählen zu können, die eine benötigte Reife aufweisen, um sinnvoll in der Praxis eingesetzt oder überhaupt implementiert

werden zu können. Es dient somit der expliziten Instanziierung im konkreten Anwendungsfall. Darüber hinaus kann es bei der Entwicklung neuer Metriken eingesetzt werden, um diese von vorneherein mit hoher Reife zu entwerfen und so die automatisierte Anwendung zu erleichtern.

Identifikation von Anforderungen an Dateneigenschaften von Qualitätsdimensionen

Bei der Betrachtung und ersten Umsetzung von Qualitätsdimensionen zur Demonstration im Projekt fiel schnell auf, dass es nicht trivial ist, zu den vorliegenden Daten passende Qualitätsdimensionen und -metriken auszuwählen. So stellt die Berechnung von Messwerten zu Dimensionen vielfach Anforderungen an die Eigenschaften der Daten. Eine Berechnung der Timeliness ist etwa nur möglich, wenn mindestens zwei vergleichbare Zeitstempel vorhanden sind. Häufig können die gesammelten Qualitätsdimensionen nicht ad hoc angewendet werden, da die Daten notwendige Eigenschaften nicht erfüllen.

Um eine weitere Hilfestellung zur Auswahl von geeigneten Qualitätsdimensionen und -metriken zu geben, wurde untersucht welche Eigenschaften Daten erfüllen müssen, um die jeweiligen Messwerte erheben zu können.

Die Untersuchung von relevanten Qualitätsdimensionen identifiziert Eigenschaften, die für eine Erhebung erfüllt sein müssen. Die Zuordnung dieser Anforderungen zu den Dimensionen zeigt auf welche Eigenschaften die verfügbaren Daten erfüllen müssen, um geeignete Messwerte zu erheben. So lässt sich bereits bei der Auswahl der Dimensionen die Kompatibilität mit der Datenbasis berücksichtigen oder gegebenenfalls die Datenerhebung ergänzen.

Die gefundenen Eigenschaften und Anforderungen der Qualitätsdimensionen an die Daten werden in dem noch nicht veröffentlichen Paper “Identifying Common Data Requirements of Cyber Threat Intelligence Quality Dimensions” ausführlich behandelt. Der Anhang A zeigt eine Liste dieser Dimensionen und deren Anforderungen.

Framework for Assessing CTI Quality (FACQ)

Zur Anwendung des DEVISE-Reifegradmodells müssen ausgewählte Qualitätsdimensionen gemessen werden. In den vorherigen Abschnitten wurde der Weg zu Metriken beschrieben, die geeignet sind diese Messwerte zu erhalten. Um die Messung der Datenqualität von CTI durchzuführen ist technische Unterstützung notwendig. Im Rahmen des Projekts wurde ein Demonstrator entwickelt, welcher die technische Machbarkeit verifiziert und Ansätze zur Implementierung des Reifegradmodells vorstellt.

Das Ziel der Implementierung ist eine systemübergreifende und nutzerfreundliche Applikation, die folgende Anforderungen erfüllt:

- **Open Source**: Die Plattform soll individuelle Anpassungen ermöglichen.
- **Lokal**: Die Verarbeitung potenziell sensibler Daten soll lokal erfolgen.
- **Automatisierung**: Die Erreichung höherer Reifegrade erfordert die automatisierte Verarbeitung von Daten.
- **Metriken**: Metriken sollen frei definierbar sein, um unterschiedliche Szenarien zu unterstützen.

- **Abstraktion**: Rohdaten und Implementierungen der Metriken sollen vor dem Anwender verborgen bleiben, damit die Komplexität der Anwendung reduziert werden kann.
- **Gestaltbare Pipelines**: Grundlegende Schritte sollen einfach kombinierbar sein, um neue Verarbeitungsschritte in einem Baukastensystem erstellen zu können.
- **Darstellung**: Metriken und deren Entwicklung über Zeiträume sollen visualisiert werden können.
- **Python**: Die Entwicklung soll in Python möglich sein, da es sich um eine exzellente Prototypensprache handelt und im Projektteam die meiste Erfahrung vorhanden ist.

Folgende Technologien wurden daraufhin ausgewählt und in einer zentralen Plattform integriert. Eine Übersicht ist in Abbildung 3 dargestellt.

- **Orchest**[1]: Orchest ist ein Tool zur Orchestrierung von Arbeitsabläufen, das es ermöglicht, Pipelines in einem visuellen Editor zu erstellen und zu konfigurieren. Alle vordefinierten, in Python implementierten, Schritte können flexibel miteinander kombiniert werden. Dadurch können Nutzer ohne Programmierkenntnisse mit der Oberfläche arbeiten. Weiterhin sorgt die direkte Integration von Jupyter Notebooks dafür, dass erfahrene Nutzer einfach Zwischenergebnisse einsehen und verarbeiten können. Orchest wird als Werkzeug zur Sammlung und Bearbeitung von Roh-Daten sowie für die Berechnung und das Speichern von Metriken genutzt.
- **PostgreSQL**[2]: Bei PostgreSQL handelt es sich um eine etablierte, freie Open-Source-Datenbank. In dieser werden die berechneten Metriken und Zwischenergebnisse gespeichert. Weiterhin dient sie zur zentralen Sammlung der Qualitätsdimensionen und -metriken.
- **Grafana**[3]: Für das Analysieren und Präsentieren von Daten kommt das Open-Source-Tool Grafana zum Einsatz. Dies ermöglicht die Anbindung verschiedenster Datenbanken und die direkte Erzeugung anpassbarer Visualisierungen aus den Rohdaten. Es greift auf die PostgreSQL-Datenbank zu und stellt die Metriken in Dashboards dar.
- **Django**[4]: Während die Metriken von Grafana in Grafiken visualisiert werden können, müssen auch die gesammelten Informationen zu Qualitätsdimensionen und -metriken übersichtlich dargestellt werden. Mithilfe des Django Web-Frameworks wurde eine übersichtliche und lösungsorientierte Sicht in Form einer Website geschaffen, welche sich mit den bestehenden Technologien integrieren ließ.

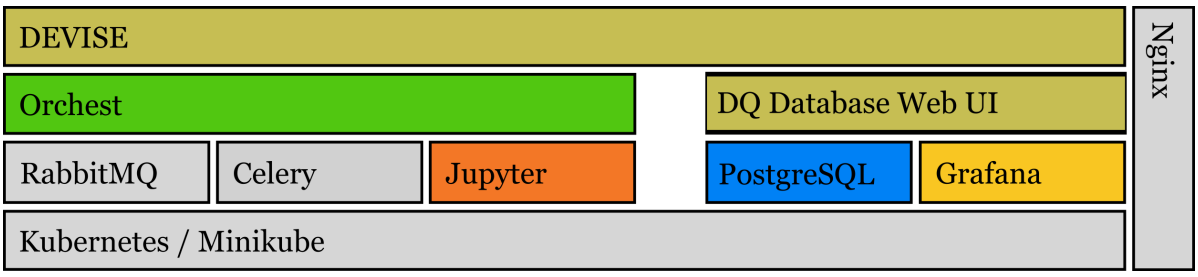


Abbildung 3: FACQ Technologie Übersicht. Unterstützende Technologie ist grau dargestellt.

[1] https://github.com/orchest/orchest, abgerufen am 10.06.2024
[2] https://www.postgresql.org/, abgerufen am 10.06.2024
[3] https://grafana.com/oss/grafana/, abgerufen am 10.06.2024
[4] https://www.djangoproject.com/, abgerufen am 10.06.2024

Mit dem gewählten Technologiestack werden die gesteckten Anforderungen sehr gut abgedeckt:

- **Open Source**: Alle gewählten Projekte sind Open Source.
- **Lokal**: Jedes Projekt bietet die Möglichkeit eigene Server zu betreiben.
- **Automatisierung**: Orchest integriert ein Scheduling zur automatischen Ausführung von Pipelines.
- **Metriken**: Per Jupyter-Scripts oder Quellcodebibliotheken kann das Orchest-Projekt um Metriken erweitert werden.
- **Abstraktion**: Als Bibliothek bereitgestellte Datenquellen, Metriken und Transformationen die in Orchest-Pipelines kombiniert werden können, sorgen für eine hohe Abstraktion und Reduktion der Komplexität für den Anwender.
- **Gestaltbare Pipelines**: Orchest ist explizit darauf ausgelegt Pipelines einfach vom Nutzer gestalten zu lassen.
- **Darstellung**: Grafana ermöglicht die Visualisierungen der Ergebnisse.
- **Python**: Mit Jupyter in Orchest integriert, sind Entwicklungen in verschiedenen Sprachen möglich unter anderem auch Python.

Die Architektur des implementierten Systems findet sich in Abbildung 4. Orchest nimmt die Daten an und schreibt berechnete Metriken nach PostgreSQL. Grafana nutzt diese Daten zu Visualisierung. Die entstehende lose Kopplung der Komponenten reduziert die Abhängigkeiten zwischen den Komponenten und erlaubt den Austausch einzelner Komponenten bei sich ändernden Anforderungen.

Um den Nutzen des von FACQ zu demonstrieren, wurde beispielhaft für IP-Blocklisten[5] der Prozess zur Messung der Datenqualität implementiert.

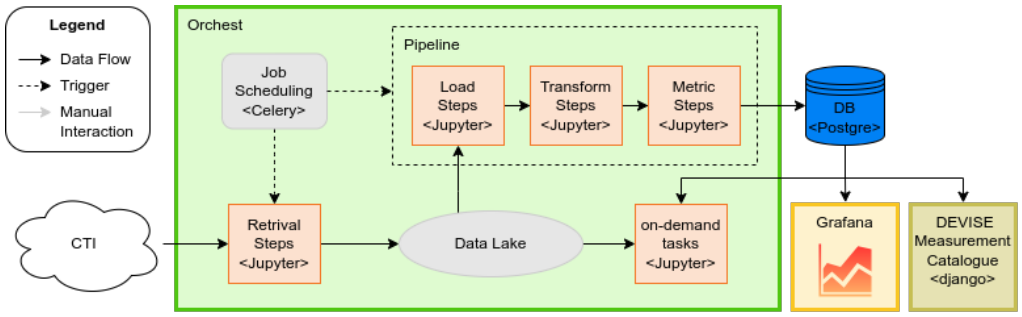


Abbildung 4: FACQ Systemübersicht

DQ Database Web UI

Die ausführliche Sammlung von Qualitätsdimensionen und -metriken stellt an sich bereits eine wesentliche Hilfe dar, um für den konkreten Fall passende Messverfahren zu entwickeln. Um die Sammlung einfacher zugänglich zu machen wurde eine Oberfläche (Abbildung 5) entwickelt, welche die Daten in einer interaktiven Form präsentiert. Hierbei ist es leicht die verfügbaren

[5] In IP-Blocklisten sind IPs enthalten, die mit böswilligen Aktivitäten, wie z. B. das Verschicken von Spam oder dem Hosten von Malware, assoziiert werden können. Sie werden i. d. R. dazu genutzt, Netzwerkverkehr von von diesen IPs direkt zu blocken.

Dimensionen zu durchsuchen, entsprechende Metriken zu finden und bei Bedarf die definierende Literaturreferenz zu erhalten. Dies unterstützt die wesentlichen Schritte der Auswahl von Qualitätsdimensionen bis hin zur Auswahl von konkreten Metriken zur Implementierung von Messverfahren.

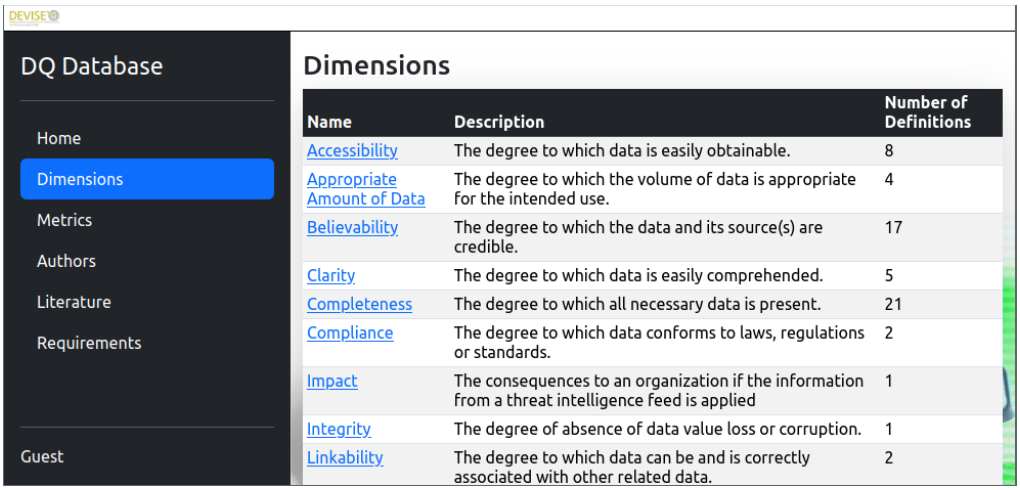


Abbildung 5: Dimensionen dargestellt in der Weboberfläche

Die Oberfläche fasst die gewonnenen Erkenntnisse über CTI-Datenqualitätsdimensionen und -metriken für den Anwender des DEVISE-Reifegradmodells zusammen ohne auf die wissenschaftlichen Details einzugehen, die zur Anwendung des Reifegradmodells nicht vonnöten sind. Des Weiteren integriert sie wesentliche Schritte zur konkreten Instanziierung des Reifegradmodells in den Demonstrator.

Orchest

Orchest wird zu der Definition, Orchestrierung und Ausführung von Pipelines zur Datenverarbeitung genutzt. Mit geeigneten Pipelines können alle Datenverarbeitungsschritte von der Beschaffung der Daten bis zur Berechnung der Metriken abgedeckt werden.

Aus der ausführlichen Sammlung von Qualitätsdimensionen und Metriken wurden für den Demonstrator Metriken umgesetzt und in Pipelines integriert, die den Anwendungsfall für IP-Blocklisten exemplarisch umsetzen.

Zuerst müssen die IP-Blocklisten von externen Quellen bezogen werden. Sie werden von verschiedenen Anbietern zur Verfügung gestellt und können in unterschiedlichen Formaten angeboten werden. Dadurch ist es notwendig, für jedes Format einen eigenen Schritt in der Pipeline zu definieren. Mit diesen können Anwender auswählen, aus welchen Quellen sie die Blocklisten beziehen wollen und in welchen Intervallen dies geschehen soll.

Initial werden die Rohdaten in einem Data Lake gespeichert. Aus diesen können anschließend die unterschiedlichen Metriken berechnet werden. Im Folgenden wird der typische Ablauf einer Berechnungspipeline dargestellt. Bei Bedarf kann dieser durch weitere oder alternative Schritte angepasst werden.

1. **Laden:** Alle für die Berechnung der Metrik notwendigen Daten, die im Data Lake gespeichert sind, müssen im ersten Schritt geladen werden.
2. **Transformieren:** Um nicht für jedes Datenformat, das eine Quelle liefert, individuelle Metriken schreiben zu müssen ist es notwendig die Daten in ein einheitliches Format zu transformieren. Für hohe Flexibilität wurden allgemeine, kombinierbare Transformationsschritte im Demonstrator implementiert.
3. **Metriken berechnen:** Die Berechnung von Metriken kann in einem oder mehreren Schritten erfolgen. Die Berechnung in mehreren Schritten bietet sich etwa an, wenn mehrere Metriken auf den gleichen Daten parallel berechnet werden oder Metriken selbst Eingabedaten für weitere Metriken sind.
4. **Ergebnisse speichern:** Nachdem alle Metriken berechnet sind, werden die Ergebnisse in eine Datenbank gespeichert, um eine Vergleichbarkeit über definierte Zeiträume zu ermöglichen.

Die Pipelines werden in einem vom Nutzer definierten Intervall regelmäßig ausgeführt. Geeignete Intervalle hängen dabei von der Art der Daten, dem Anwendungsfall und dem Aktualisierungsintervall der Quellen ab.

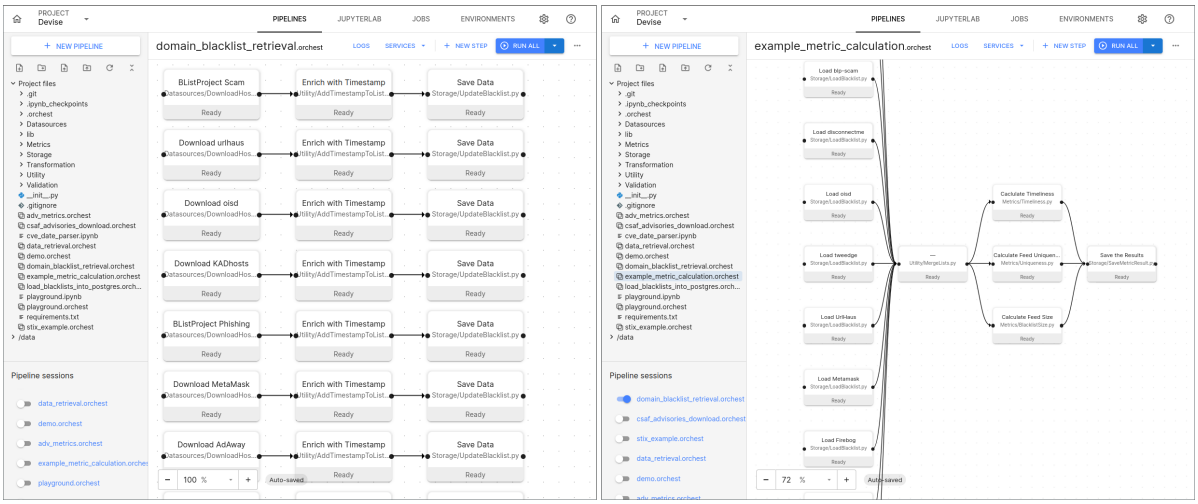


Abbildung 6: Pipelines zur Datensammlung (links) und Metrikberechnung (rechts)

Die beispielhafte Umsetzung in zwei Pipelines für die Sammlung von Daten und die Berechnung der Metriken ist in Abbildung 6 zu sehen. Die Aufteilung in zwei Pipelines ermöglicht eine größere Flexibilität in der Ausführung, so lässt sich etwa das Intervall zum Abruf der Daten kleiner wählen als das zur Berechnung der Metriken.

Grafana

Die von Orchest erzeugten Metriken werden aus der Datenbank gelesen und in vom Nutzer definierten Graphen und Dashboards dargestellt (vgl. Abbildung 7). Darüber hinaus kann ad hoc auf die Daten zugegriffen werden, um diese im Detail zu analysieren.

Abbildung 7: Visualisierte Metriken in der Grafana Oberfläche

Einstellung der Entwicklung von Orchest und Betrachtung der Alternative NiFi

Nachdem die Implementierung der Grundfunktionen von *FACQ* abgeschlossen war, wurde überraschenderweise und ohne langen Vorlauf die Entwicklung der *Orchest*-Plattform eingestellt. Dies beeinträchtigt die Funktionsweise des Demonstrators nicht, stellt aber einen produktiven Einsatz offensichtlich infrage. Eine eigenständige Weiterentwicklung von *Orchest* wäre für diesen Zweck möglich, da es sich entsprechend der aufgestellten Anforderungen an die Komponenten um ein Open-Source-Projekt handelt.

Durch den weiten Fortschritt des Projektes und Demonstrators, wurde von einem Ersatz von Orchest durch eine Alternative abgesehen. Trotzdem wurde eine Recherche nach möglichen Alternativen für vergleichbare Projekte durchgeführt.

Die passendste Alternative, die identifiziert wurde, ist Apache NiFi in Version 2. Mit dieser Version wird Python direkt zur Entwicklung von Processors unterstützt und damit die aufgestellten Anforderungen ebenfalls erfüllt:

- **Open Source:** NiFi wird als Open Source entwickelt.
- **Lokal:** NiFi kann auf eigenen Servern betrieben werden.
- **Automatisierung:** NiFi bietet ähnliche Scheduling-Optionen wie Orchest.
- **Metriken:** Metriken können als Python-Processor entwickelt werden, wenn auch nicht direkt in der Oberfläche.
- **Abstraktion:** Analog zu Orchest können als Bibliotheken von Processors bereitgestellte Quellen, Metriken und Transformationen in Pipelines nutzerfreundlich kombiniert werden.
- **Gestaltbare Pipelines:** NiFi bietet eine einfache Oberfläche, um Pipelines zu gestalten.

Zudem handelt es sich bei NiFi um ein etabliertes Projekt der Apache-Foundation, sodass eine Einstellung des Projekts unwahrscheinlich ist.

Die wichtigsten Positionen des zahlenmäßigen Nachweises

Der zahlenmäßige Nachweis setzt sich aus zwei Positionen zusammen:

- Personalkosten in Höhe von 240.052,89 €.

  Die Personalkosten wurden für die im Rahmen des Projektes erforderlichen Planungs- und Implementierungsaufgaben verwendet. So wurden die oben beschriebenen Arbeiten zur Instanziierung des Reifegradmodells für CTI-Daten durchgeführt und in einem Demonstrator beispielhaft implementiert.

- Reisekosten in Höhe von 4.153,81 €.

  Die nachgewiesenen Reisekosten wurden für die zur Koordinierung der Projektarbeiten erforderlichen Treffen der Verbundpartner, sowie die BMBF-Statustreffen verwendet.

Notwendigkeit und Angemessenheit der geleisteten Projektarbeit

Schon vor Beginn des Projektes war deutlich, dass die Forschung im Bereich von Reifegradmodellen bezüglich der Datenqualität von sicherheitsrelevanten Daten und im speziellen CTI deutliche Lücken aufweist. Dieser Eindruck wurde während der Projektlaufzeit erhärtet und noch wesentlich grundlegendere Mängel aufgezeigt als ursprünglich erwartet.

Somit bedurfte es grundlegenderer Arbeiten, um Dimensionen und Metriken für ein Reifegradmodell auswählen zu können. Diese grundlegenden Arbeiten wurden während der Projektzeit geleistet.

Es wurden Diskrepanzen und Lücken zwischen den in der Literatur betrachteten und in der Praxis für relevant befundenen Dimensionen aufgedeckt. Mit den entwickelten Methoden können diese Bereiche besser beleuchtet und mit entsprechenden Metriken versorgt werden, um Lücken zu schließen. Hierbei unterstützen die wesentlichen Arbeiten zur Strukturierung des Bereiches und Bewertung von vorliegenden Metriken.

Mit den Arbeiten wurde eine angemessene Basis geschaffen, um in konkreten Fällen ein Reifegradmodell zur Bewertung der Datenqualität von CTI zu instanziieren.

Die Umsetzung des DEVISE-Reifegradmodells wurde anhand einer möglichen Instanziierung im Bereich von CTI durchgeführt. Hierbei wurde auf die besondere Heterogenität des Bereiches geachtet und ein flexibles Modell geschaffen, das auf konkrete Fälle angepasst werden kann.

Um die Messung der Qualitätsmetriken von CTI zu ermöglichen ist technische Unterstützung notwendig. Für diesen Anwendungsfall gibt es keine Standardlösung, sodass innerhalb des Projektes ein System entworfen wurde, welches alle erforderlichen Schritte zur Berechnung der Datenqualität von CTI implementiert. In diesem Demonstrator wurden Schritte zur Unterstützung einer konkreten Instanziierung für jede Phase des Reifegradmodells integriert (Instanziierung, CTI-Datenbeschaffung, Transformation, Metrikberechnung, Darstellung).

Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Mit den aus dem Projekt gewonnen Erfahrungen bestehen konkrete Pläne ein System zur Evaluation von Metriken zu etablieren. Hierbei werden die Erkenntnisse des Demonstrators berücksichtigt und ähnliche Strukturen übernommen.

Die Ergebnisse zu Qualitätsdimensionen und Metriken bieten in diesem Zusammenhang wertvolle Grundlagen, um geeignete Qualitätsmetriken für die verarbeiteten Daten auszuwählen. In der Kombination verspricht dies eine Verbesserung bei der Einschätzung und fortlaufenden Kontrolle der Qualität verarbeiteter Daten, sowie weitere Erkenntnisse über den konkreten Inhalt der einzelnen Datenquellen hinaus.

Auch im wissenschaftlichen Sinne ergeben sich Anknüpfungspunkt für weitere Forschungen. Die noch unveröffentlichten Untersuchungen zu den Reifegraden existierender Metriken sowie der Anforderungen von Dimensionen an die zu messenden Daten bieten hierfür direkte Impulse. Sie zeigen vorhandene Lücken innerhalb des Forschungsgebietes auf, welche weitere Fortsetzungsmöglichkeiten bieten. So wurden verschiedene Dimensionen erkannt, welche zwar in der Praxis als relevant eingeschätzt werden, jedoch von Metriken nur unzureichend abgedeckt sind.

Bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen während der Durchführung des Vorhabens

Während der Projektlaufzeit sind einige Arbeiten erschienen, die dem Bereich des Vorhabens verwandte Themen behandeln.

So stellt Sakellariou, Fouliras, und Mavridis (2024) eine Vorgehensweise vor, um CTI-Qualitätsmetriken zu bewerten und zu entwickeln. Diese Arbeit ist parallel zu der hier erfolgten Ausarbeitung zur Reife von Metriken erschienen und betont die Wichtigkeit dieses Ansatzes. Während viele Gemeinsamkeiten vorhanden sind, ist der Ansatz jedoch ein anderer und es wird großes Augenmerk auf die Komplexitätsanalyse der Metriken gelegt. Es erfolgt keine Anwendung auf vorhandene Metriken wie sie in der vorgestellten Arbeit stattgefunden hat.

Weitere Arbeiten beschäftigen sich damit, wie Datenqualität im Bereich von CTI gemessen werden kann. Die Ergebnisse wurden in der Auswahl von Dimensionen und Metriken berücksichtigt (Zibak, Sauerwein, und Simpson (2022), Hofer (2021)). Genereller mit Datenqualität auseinandergesetzt hat sich Hassenstein und Vanella (2022), was in die allgemeine Betrachtung des Bereichs eingeflossen ist.

Technischere Entwicklungen und praxisnahe Betrachtungen gaben weiter hilfreiche Impulse. So bietet Ehrlinger und Wöß (2022) eine Übersicht über gängige allgemeine Arten Qualität zu messen und welche Tools benutzt werden.

Erfolgte und geplante Veröffentlichungen der Ergebnisse

Unter dem Titel “Understanding the Anatomy of Cybersecurity Advisory Feed Data” konnte ein Vortrag auf der internationalen Konferenz OCSC 2024 platziert werden, der zu wertvollem Feedback geführt und die Reichweite für den Forschungsbereich der Datenqualität im Bereich

von CTI erhöht hat. Dieser betrachtete, wie Qualität für komplexe CTI-Daten wie Cybersecurity-Advisories gemessen werden kann. Außerdem wurde ein Kurzvortrag zum Reifegradmodell für Metriken gehalten.

Zwei weitere Paper (“About the Maturity of Data Quality Measurement in Cyber Threat Intelligence” und “Identifying Common Data Requirements of Cyber Threat Intelligence”) sind entstanden, für die nach geeigneten Möglichkeiten zur Veröffentlichung gesucht wird. Diese vertiefen und erweitern die Inhalte aus den Abschnitten Bewertung von Metriken mittels eigenem Reifegradmodell und Identifikation von Anforderungen an Dateneigenschaften von Qualitätsdimensionen.

Literatur

Bruin, Tonia de, Michael Rosemann, Ron Freeze, und Uday Kulkarni. 2005. „Understanding the Main Phases of Developing a Maturity Assessment Model“. In *ACIS 2005 Proceedings - 16th Australasian Conference on Information Systems*. New Zealand: Australasian (ACIS). https://aisel.aisnet.org/acis2005/109.

CMMI Product Team. 2010. „CMMI for Development“. CMU/SEI-2010-TR-033. Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University. https://doi.org/10.1184/R1/6572342.v1.

Ehrlinger, Lisa, und Wolfram Wöß. 2022. „A Survey of Data Quality Measurement and Monitoring Tools“. *Frontiers in Big Data* 5 (März). https://doi.org/10.3389/fdata.2022.850611.

Hassenstein, Max J., und Patrizio Vanella. 2022. „Data Quality—Concepts and Problems“. *Encyclopedia* 2 (1): 498–510. https://doi.org/10.3390/encyclopedia2010032.

Hofer, Brigitte. 2021. „Master Thesis Assessing the Quality and Accuracy of Shared Cyber Security Information“. Mathesis, University of Innsbruck.

Sakellariou, Georgios, Panagiotis Fouliras, und Ioannis Mavridis. 2024. „A methodology for developing & assessing CTI quality metrics“. *IEEE Access*.

Zibak, Adam, Clemens Sauerwein, und Andrew C. Simpson. 2022. „Threat Intelligence Quality Dimensions for Research and Practice“. *Digital Threats: Research and Practice* 3 (4): 1–22. https://doi.org/10.1145/3484202.

Anhang A - Qualitätsdimensionen

Dieser Anhang enthält eine Liste von Dimensionen mit kurzer Beschreibung, Vertretern aus der Literatur und Anforderungen, die die Dimensionen an die verarbeiteten Daten stellt.

Accessibility Der Grad zu dem Informationen zugegriffen werden können.

Beschreibung Die Dimension Accessibility beschreibt, zu welchem Grad die Informationen zugreifbar sind. Es ist ein Indikator für die Kosten, die Daten zu beziehen, was die Geschwindigkeit und Leichtigkeit die Informationen zu finden und zu nutzen mit einschließt (Hofer 2021). Mehrere Eigenschaften und Operationen tragen zu dieser Dimension bei, unter anderem: Informationssammlung, Korrelation, Integration, Verbesserung & Kontextualisierung, Durchsuchen & Abfragen, Mustererkennung, Berichtsgenerierung, Verbreitung, Automatisierung und Integration mit existierenden Systemen (Rashid, Noor, und Altmann 2019). Teilweise wird Quality of Service (QoS) als Alias für die Dimension genutzt (Rashid, Noor, und Altmann 2019). Eine konkret vorgeschlagene Metrik zur Messung der Accessibility ist zum Beispiel die Web Accessibility Quality Metric (Freire u. a. 2008).

Anforderungen Die Dimension Accessibility hat keine Anforderungen an die Daten selbst, da gemessen wird, wie die Daten beschafft und verarbeitet werden. Um allerdings z. B. die Web Accessibility Metric anzuwenden, müssen `Metadaten` festgehalten werden über den Beschaffungs- und Verarbeitungsprozess. Dies beinhaltet die Fehlerrate, die benötigte Zeit pro Abfrage und wie viele manuelle Eingriffe erforderlich sind.

Synonyme Accessibility (Kahn, Strong, und Wang 2002; Wang und Strong 1996; Pipino, Lee, und Wang 2002; Hofer 2021; Black und van Nederpelt 2020), Quality of Service (QoS) (Rashid, Noor, und Altmann 2019), Availability (Zaveri u. a. 2016; Black und van Nederpelt 2020)

Accuracy Das Maß an Korrektheit der Daten.

Beschreibung Im Bereich CTI ist die Dimension Accuracy ist sehr überladen. Die meisten Arbeiten sind sich einig darüber, dass beschrieben wird wie nah die Daten an den korrekten Werten sind. Dies entspricht der Definition von Accuracy in der allgemeinen Datenqualitätsliteratur (Fox, Levitin, und Redman 1994; Talha, Abou El Kalam, und Elmarzouqi 2019). Hierbei handelt es sich somit um eine semantische Genauigkeit oder Semantic Accuracy. Allerdings existieren verschiedene Ansätze wie dies ausgedrückt wird. Zunächst gibt es die klassische Definition einer False-Positive-Rate (FPR), die in Statistik und anderen Forschungsfeldern weit verbreitet ist (Hofer 2021). Manche Arbeiten erweitern dies um den Prozentsatz von Duplikaten bzw. nicht zuordenbaren Objekten, sowie den Abstand zwischen bekannten Werten und vorgeschlagenen Werten (Batini und Scannapieco 2016). Andere hingegen schlagen vor, die True-Positive-Rate (TPR) zu nutzen, welche auch als Genauigkeit (Precision) bekannt ist (Grispos, Glisson, und Storer 2019; Bouwman u. a. 2020). Da häufig keine Daten als Ground-Truth vorhanden sind, greifen die meisten Ansätze auf vergleichende Analysen zwischen Datenfeeds zurück, indem die kombinierten Daten an Stelle der Ground-Truth genutzt werden (Bouwman u. a. 2020). Ein Ansatz erstellt zu diesem Zweck einen Korrelationsgraphen zwischen den Feeds (Meier u. a. 2018). Außerdem gibt es datenspezifische Ansätze, wie Tests gegen Sperrlisten mit sicher falschen Werten (z. B. für Blocklisten nicht routbare IPs oder IPs von Top-Alexa-Domains, sowie bekannten

Netzwerken zur Verbreitung von Inhalten (CDNs)), um eine Schätzung der FPR zu berechnen (Li u. a. 2019; Sinha, Bailey, und Jahanian 2008). Es gibt jedoch auch Arbeiten, die eine syntaktische Perspektive einnehmen und so eine Syntactic Accuracy definieren (Batini u. a. 2009; Schlette u. a. 2021; Behkamal u. a. 2014). Hierbei ist die Betrachtung ähnlich wie zuvor, nur dass nicht der Wert, sondern die Form der Daten analysiert wird.

Anforderungen Meistens wird Accuracy entweder über die FPR oder TPR berechnet. Die FPR benötigt dabei eine Definition, was als korrekt angesehen wird. Optimal dabei ist der Vergleich mit einer bekannten `Ground-Truth`, die jedoch für CTI-Daten häufig nicht verfügbar ist. Die beste Annäherung wäre hierfür eine `Feedbackschleife`, um im Nachgang die Daten zu kennzeichnen und so die FPR zu bestimmen. Eine sehr grobe Abschätzung der FPR kann durch Nutzung von Sperrlisten für offensichtlich falsche Werte gewonnen werden. Ein weiterer Ansatz bietet sich für `Verifizierbare Daten`, die mittels weiterer Quellen überprüft werden können. So lassen sich mögliche Anhaltspunkte in Logs, Netzwerkverkehr oder externen Diensten finden. Während die FPR nie über einen Vergleich mit anderen Quellen gewonnen werden kann, kann die TPR abgeschätzt werden, wenn Quellen eine erhebliche `Überschneidung` aufweisen. Hierbei liegt die Annahme zugrunde, dass Ereignisse, welche in mehr als einer unabhängigen Quelle vorkommen, mit hoher Wahrscheinlichkeit wahr sind. Für eine syntaktische Betrachtung muss es sich um `Strukturierte Daten` handeln.

Synonyme (Semantic Accuracy) Semantic Accuracy (Zaveri u. a. 2016; Behkamal u. a. 2014), Accuracy (Zibak, Sauerwein, und Simpson 2022; Umbrich, Neumaier, und Polleres 2015; Paweł Pawliński u. a. 2015; Wook u. a. 2021; Hofer 2021; Taleb u. a. 2016; Black und van Nederpelt 2020; Fox, Levitin, und Redman 1994; Grispos, Glisson, und Storer 2019), Naturalness (Black und van Nederpelt 2020), Plausibility (Black und van Nederpelt 2020), Precision (Fox, Levitin, und Redman 1994; Black und van Nederpelt 2020), Correctness (Al-Ibrahim u. a. 2017b), Free-of-Error (Kahn, Strong, und Wang 2002; Wang und Strong 1996; Pipino, Lee, und Wang 2002), Robustness (Canfora u. a. 2020)

Synonyme (Syntactic Accuracy) Syntactic Accuracy (Batini u. a. 2009; Schlette u. a. 2021; Behkamal u. a. 2014), Structural Consistency (Micic u. a. 2017), Consistency (Hofer 2021; Taleb u. a. 2016; Black und van Nederpelt 2020), Metadata compliance (Black und van Nederpelt 2020), Precision (Black und van Nederpelt 2020), Syntactic Validity (Zaveri u. a. 2016)

Appropriate Amount of Data Maß an (ausreichend) hilfreichen Informationen.

Beschreibung Die Dimension Appropriate Amount of Data beschreibt, ob ein Datenobjekt hinreichend hilfreiche Informationen enthält. Da diese Bewertung in der Regel nur aus dem Vergleich zwischen Daten und Realität folgen kann, muss hierfür etwa auf Heuristiken aus der Praxis zurückgegriffen werden, die die Abbildung der Realität anhand des Verlinkungsgrads und der Diversität der verlinkten Informationen beurteilt (Schlette u. a. 2021). Diese Metrik ist derzeit nur für das STIX-Format definiert, kann aber für weitere Formate adaptiert werden, solange diese miteinander verkettete Teilobjekte aufweisen.

Anforderungen Um abzuschätzen ob die Informationen hinreichend gut die Realität abbilden, benötigt die Metrik Appropriate Amount of Data `Verkettete Teilinformationen`.

Synonyme Appropriate Amount of Data (Kahn, Strong, und Wang 2002; Wang und Strong 1996; Schlette u. a. 2021; Pipino, Lee, und Wang 2002)

Believability Vertrauen in den Herausgeber.

Beschreibung Die Believability misst den Grad des Vertrauens in den Herausgeber oder die Einschätzung der Glaubwürdigkeit des Datensatzes. Ein verbreitetes Synonym dafür ist Reputation. Das Vertrauen selbst ist im Kern eine subjektive Bewertung durch den Nutzer der Daten. Die meisten vorgeschlagenen Metriken gehen von einer Einstufung des Vertrauens in den Hersteller oder den Datensatz aus und leiten daraus eine Qualitätsmetrik an (Hofer 2021; Schlette u. a. 2021; Sillaber, Mussmann, und Breu 2019; Schaberreiter u. a. 2019). Ein anderer Ansatz besteht darin, Daten zu markieren, wenn sie auch in anderen Quellen gesehen oder genutzt wurden. Dies ist z. B. in Plattformen wie MISP (C. Wagner u. a. 2016) umgesetzt.

Anforderungen Neben einer vertrauenswürdigen Einstufung der `Quellen` müssen die Daten eine Originalquelle enthalten, falls sie vom Feed-Anbieter abweicht. Alternativ kann auch ein Empfehlungssystem, wie z. B. Sichtungen in MISP, verwendet werden, was letztlich eine Art `Feedbackschleife` darstellt. Einige Feed-Anbieter integrieren direkt `Reputationswerte` in ihre Daten. Diese spiegeln jedoch nur die Meinung des Anbieters wider.

Synonyme Believability (Kahn, Strong, und Wang 2002; Wang und Strong 1996; Pipino, Lee, und Wang 2002; Wook u. a. 2021; Prat und Madnick 2008; Hofer 2021), Reliability Black und van Nederpelt (2020), Reputation (Kahn, Strong, und Wang 2002; Wang und Strong 1996; Pipino, Lee, und Wang 2002; Schlette u. a. 2021; Black und van Nederpelt 2020), Trustworthiness (Zaveri u. a. 2016)

Completeness Verhältnis von gegebenen zu erwarteten Informationen

Beschreibung Die Definitionen der Dimension Completeness haben gemeinsam, dass sie das Verhältnis von zur Verfügung gestellten Informationen zu erwarteten Informationen messen. Die Arbeiten unterscheiden sich in Details der Definitionen. Der erste Ansatz besteht darin, das Verhältnis der ausgefüllten Felder zu den verfügbaren Feldern zu messen (Fox, Levitin, und Redman 1994). Üblicherweise wird dabei zwischen obligatorischen und optionalen Feldern unterschieden. Bei den meisten Ansätzen wird gemessen, wie viele optionale Felder ausgefüllt sind (Hofer 2021; Grispos, Glisson, und Storer 2019; Schaberreiter u. a. 2019; Schlette u. a. 2021). Bei anderen Ansätzen ist Completeness definiert als das Maß, in dem alle notwendigen Werte gegeben sind (Talha, Abou El Kalam, und Elmarzouqi 2019) oder alle Informationen, die zur Durchführung von abgeleiteten Aktionen erforderlich sind (Paweł Pawliński u. a. 2014). Dies liegt nah an der Verwendung obligatorischer Felder, lässt aber Raum für Interpretationen. Auf Ebene von Elementen eines Feeds ist Completeness in CTI äquivalent zu Recall im Bereich Information Retrieval. Wie jedoch bereits für Accuracy erwähnt, gibt es meistens keine Ground-Truth mit der verglichen werden könnte. Folglich wird die Menge relevanter Informationen häufig durch die Kombination aller verfügbaren Feeds konstruiert (Li u. a. 2019; Meier u. a. 2018).

Anforderungen Die Anforderungen von Completeness variieren zwischen den unterschiedlichen Definitionen. Das grundlegende Verhältnis zwischen ausgefüllten und möglichen Attributen erfordert nur `Strukturierte Daten` mit entsprechend definierten Attributen. Im Hinblick auf ausgefüllte optionale Attribute müssen `Obligatorische Attribute` unterschieden werden. Dies gilt auch für die Messung des Vorhandenseins aller notwendigen Informationen, wenn diese als erforderlich definiert sind. Auf Feedebene kann die Abdeckung nur korrekt gegen eine `Ground-Truth` gemessen werden. Ist die Ground-Truth nicht

bekannt, kann die Coverage bei einer hinreichenden `Überschneidung` der Feeds abgeschätzt werden.

Synonyme - Elementebene Completeness (Micic u. a. 2017; Kahn, Strong, und Wang 2002; Umbrich, Neumaier, und Polleres 2015; Wang und Strong 1996; Pipino, Lee, und Wang 2002; Grispos, Glisson, und Storer 2019; Behkamal u. a. 2014; Wook u. a. 2021; Paweł Pawliński u. a. 2015; Zaveri u. a. 2016; Hofer 2021; Taleb u. a. 2016; Fox, Levitin, und Redman 1994; Batini u. a. 2009; Black und van Nederpelt 2020), Schema Completeness (Schlette u. a. 2021), Extensiveness (Schaberreiter u. a. 2019; Hofer 2021)

Synonyme - Feedebene Coverage Pinto (2018), TotalCoverage(Pitsillidis u. a. 2012), CoverageFalseNegatives (Bouwman u. a. 2020), CoveragePhishing (Sheng u. a. 2009)

Compliance Maß wie weit Daten Regularien entsprechen

Description Die Dimension Compliance erfasst inwieweit die Daten Gesetzen und Regularien entsprechen (Black und van Nederpelt 2020). Dies ist eine sehr offene Dimension, welche lokal sehr variieren kann, mit (Teil-)Ausprägungen wie z. B. Privacy (Mavzer u. a. 2021).

Anforderungen Die konkreten Anforderungen sind hierbei sehr domänenabhängig und nicht allgemein beschreibbar. Zur Messung müssen die Gesetze oder Regularien technisch überprüfbar sein.

Synonyme Compliance (Black und van Nederpelt 2020)

Concise Representation Aussagekraft von CTI und Redundanz in den Daten.

Description Concise Representation misst die Informationsdichte der Daten. Im Bezug auf CTI bedeutet dies in den meisten Fällen die Eindeutigkeit von Attributen (intensional) oder Objekten (extensional)(Schlette u. a. 2021).

Anforderungen Um Concise Representation zu messen, müssen Duplikate erkannt werden. Es gibt verschiedene Arten Duplikate zu erkennen und es hängt stark von den Daten selbst ab, welche vernünftig angewendet werden können. Generell kann dies sowohl auf der Ebene von Attributen als auch auf der Objektebene geschehen. Hierbei ist die einzige Anforderung, dass die Objekte `Identifizierbar` sind, entweder durch explizite Identität oder eine Duplikaterkennung.

Synonyme Concise Representation(Schlette u. a. 2021)

Consistency Drückt aus wie konsitent Werte in den Daten gegeben sind.

Beschreibung Die Dimension Consistency beschreibt die Konsistenz der Werte in den Daten. Es gibt Definitionen die ihren Fokus auf das Level einzelner Datenfelder legen – ob die Werte konsistent bezüglich eines Formates sind (Grispos, Glisson, und Storer 2019) –, auf das Level eines Berichts – ob die Quelle konform bezüglich eines Standards ist (Schlette u. a. 2021) – oder auf alle Aspekte der Daten (Hofer 2021; Talha, Abou El Kalam, und Elmarzouqi 2019). In anderen Arbeiten wird diese Dimension auch Compliance (Schaberreiter u. a. 2019) oder Correctness (Mohaisen u. a. 2017; Park u. a. 2018; Al-Ibrahim u. a. 2017a) genannt. Teilweise wird zwischen syntaktischer und repräsentativer Konsistenz unterschieden (Schlette u. a. 2021). Syntaktisch wird dabei bewertet, wie weit sich der Datenfeed an eine vordefinierte Syntax hält. Ohne eine definierte Syntax kann immer noch bestimmt werden, inwieweit die

Informationen in ähnlichen oder identischen Strukturen präsentiert werden, sowie ob die Informationen logisch konsistent über Zeit und Raum genutzt werden (Hofer 2021).

Anforderungen Die Konsistenz der Daten wird immer anhand von wohl definierten Regelsätzen gemessen. Somit benötigt diese Dimension eine wohl definierte Syntax oder einen `Standard` an dem sie sich orientiert. Dieser Standard kann sowohl die Struktur von Formats und Daten als auch der Werte der Attribute definieren. Solche Standards für die Daten können teilweise auch aus historischen Daten extrahiert werden.

Synonyme Consistency (Batini u. a. 2009; Hofer 2021; Black und van Nederpelt 2020; Grispos, Glisson, und Storer 2019), Validity (Black und van Nederpelt 2020), Representational Consistency (Schlette u. a. 2021), Consistent Representation (Kahn, Strong, und Wang 2002; Wang und Strong 1996; Pipino, Lee, und Wang 2002), Compliance (Schaberreiter u. a. 2019), Correctness (Mohaisen u. a. 2017; Park u. a. 2018; Al-Ibrahim u. a. 2017a)

Objectivity Das Maß wie wenig Emotionen und subjektive Meinung in den Informationen sind.

Description Mit der Dimension Objectivity wird ein Indikator gegeben, inwieweit die Informationen frei von Emotionen und Subjektivität sind. Dies kann durch die Betrachtung von Schlüsselwörtern angenähert werden, welche auf Unsicherheiten, Emotionen oder subjektive Meinung hinweisen (Schlette u. a. 2021).

Anforderungen Da Objectivity nur für natürlicher Sprache definiert ist, benötigen die Daten Attribute mit `Freitext`.

Synonyme Objectivity (Kahn, Strong, und Wang 2002; Wang und Strong 1996; Pipino, Lee, und Wang 2002; Schlette u. a. 2021; Black und van Nederpelt 2020)

Portability Der Grad mit dem die Daten verarbeitbar sind.

Description Die Dimension Portability beschreibt den Grad, mit dem die CTI-Daten in einem Format vorliegen, das die automatisierte Verarbeitung, den Import in Verwaltungssysteme, den Austausch mit Anderen und die Verschiebung an andere Speicherort ermöglicht, ohne die Qualität zu mindern (Paweł Pawliński u. a. 2014; Bouwman u. a. 2020). Ein weit verarbeiteter Ansatz ist die Nutzung von Validatoren, die Objekte auf Konformität zu ihrer Formatdefinition überprüfen (Hofer 2021; Schaberreiter u. a. 2019).

Anforderungen Für Ingestibility konnten keine datenspezifischen Anforderungen gefunden werden. Allerdings ist zur Messung eine Bewertung der verfügbaren Austauschformate erforderlich, die dann auf einen Feed angewendet werden kann. Diese Bewertung ist organisationsspezifisch und hängt stark von der Nutzung der Informationen ab.

Synonyme Portability (Hofer 2021; Black und van Nederpelt 2020), Ingestibility (Paweł Pawliński u. a. 2015), Interoperability (Zibak, Sauerwein, und Simpson 2022; Zaveri u. a. 2016)

Relevance Grad der Relevanz der Daten.

Beschreibung Relevanz ist eine der meistgenannten Dimensionen für CTI-Datenqualität. Alle Arbeiten stimmen darin überein, dass die Relevanz von CTI für eine bestimmte Organisation gemessen werden sollte. Die Relevanzmetrik kann definiert werden als das Verhältnis zwischen der Anzahl der kundenrelevanten Objekte und der Gesamtzahl der bereitgestellten Objekte (Schlette u. a. 2021). Die Herausforderung besteht darin, die Relevanz eines

Objekts oder eines Ereignisses zu bestimmen. Einige Arbeiten schlagen Gewichtskennzeichnungen vor (T. D. Wagner u. a. 2019; Mohaisen u. a. 2017; Park u. a. 2018), aber in der Regel ist es komplexer, da Relevanz sich aus unterschiedlichen Informationen zusammensetzt, wie z. B. Zielsysteme, Sektorfokus, geografischer Fokus oder Fokus auf bestimmte Interessengruppen (Bouwman u. a. 2020; Schlette u. a. 2021; T. D. Wagner u. a. 2019). Häufig wird auch vorgeschlagen, ein System von Rückmeldungen oder Empfehlungen einzubauen (Schlette u. a. 2021; Hofer 2021).

Anforderungen Eine Vielzahl von Attributen kann die Relevanz der Daten anzeigen. Dazu gehören Standortdaten (z. B. IPs sowie geografische Daten), verwendete Software oder Hardware (z. B. Common Platform Enumeration (CPE)) oder die Branchenzugehörigkeit von Organisationen. All diese Attribute haben gemeinsam, dass sie `Zielinformationen` liefern, mit denen sich die Wahrscheinlichkeit bestimmen lässt, ein Ziel der beschriebenen Bedrohung zu werden. In einigen Fällen sind die `Zielinformationen` nur zugänglich, wenn die CTI mit organisationsinternen Daten korreliert ist.

Synonyme Relevance (Zibak, Sauerwein, und Simpson 2022; Paweł Pawliński u. a. 2015; Hofer 2021; Al-Ibrahim u. a. 2017b; Black und van Nederpelt 2020), Credibility (Black und van Nederpelt 2020), Relevancy (Kahn, Strong, und Wang 2002; Wang und Strong 1996; Pipino, Lee, und Wang 2002; Zaveri u. a. 2016; Schlette u. a. 2021), Value-Added (Kahn, Strong, und Wang 2002; Wang und Strong 1996; Pipino, Lee, und Wang 2002)

Sensitivity Sensitivität.

Beschreibung Die Dimension Sensitivity beschreibt die Häufigkeit oder Intensität von böswilligem oder unerwünschtem Verhalten, die für die Aufnahme eines Datums in einen Feed erforderlich ist (H. Griffioen, Booij, und Doerr 2020). Es kann jedoch auch die Sensitiviät einzelner Feedeinträge gemessen werden: der Grad zu dem das Element, etwa eine Erkennungsregel, das gesuchte Verhalten detektieren oder beschreiben kann (Canfora u. a. 2020).

Anforderungen Die Daten unterliegen keinen Anforderungen, da auf Feedebene hier nicht die Daten, sondern der Feed oder Anbieter an sich gemessen wird. Auf der Ebene einzelner Feedeinträge ist die Generalisierbarkeit der Daten gemessen.

Synonyme Sensitivity (H. Griffioen, Booij, und Doerr 2020), Looseness (Canfora u. a. 2020)

Timeliness Aktualität.

Beschreibung Timeliness beschreibt den Grad der Aktualität von neu gewonnen Informationen. Wenngleich dies eine der Dimensionen mit den meisten Metriken ist, variiert die Umsetzung. Die meiste Zeit beschreibt sie die vergangene Zeit zwischen dem Sammeln und dem Verteilen von Informationen (Schaberreiter u. a. 2019). Ein großes Problem in der Praxis stellt das fehlende Wissen zum tatsächlichen Enstehungszeitpunkt einer Infomation dar. Deshalb wird häufig die relative Timeliness zwischen mehreren Feeds berechnet (Hofer 2021; Bouwman u. a. 2020; Schaberreiter u. a. 2019; Li u. a. 2019; Meier u. a. 2018). Einige nehmen auch die Tatsache, dass Informationen mit der Zeit an Bedeutung verlieren, in die Defintion mit auf (Sillaber, Mussmann, und Breu 2019), was unter anderem durch die Volatilität und das Alter der Information bestimmt werden kann (Schlette u. a. 2021; Schaberreiter u. a. 2019).

Anforderungen Um die Timeliness von Daten zu bestimmen, werden mindestens zwei Zeitstempel benötigt. Zum einen muss für jedes Datum bekannt sein, wann es importiert oder erzeugt wurde. Die Timeliness berechnet sich anschließend aus diesem `Zeitstempel` und einem `Referenzzeitstempel`. Idealerweise bezieht sich die Referenzzeit auf die tatsächliche Zeit, wann diese Information zum ersten Mal beobachtet wurde. In den meisten Fällen ist dieser Zeitpunkt allerdings schwierig, wenn nicht sogar unmöglich zu bestimmen. Deshalb werden in der Regel relative Zeiten genutzt, wie z. B. welcher Feed eine Information zuerst beinhaltet hat.

Synonyme Timeliness (Kahn, Strong, und Wang 2002; Wang und Strong 1996; Pipino, Lee, und Wang 2002; Grispos, Glisson, und Storer 2019; Paweł Pawliński u. a. 2015; Wook u. a. 2021; Zaveri u. a. 2016; Hofer 2021; Schlette u. a. 2021; H. J. Griffioen 2022; Black und van Nederpelt 2020), Currency (Black und van Nederpelt 2020), Latency (Black und van Nederpelt 2020), Punctuality (Black und van Nederpelt 2020), Currentness (Zibak, Sauerwein, und Simpson 2022; Fox, Levitin, und Redman 1994)

Uniqueness Der Anteil von exklusiven Informationen.

Beschreibung Die Dimension Uniqueness erfasst das Verhältnis von Informationen, die ein Feed exklusiv beisteuert. Somit werden immer mehrere Feeds miteinander verglichen und es wird die Menge an Beiträgen gemessen, die nur in diesem einen Feed enthalten sind (Li u. a. 2019; Mohaisen u. a. 2017; Park u. a. 2018). Dabei muss es nicht bei der reinen Zählung von Beiträgen bleiben, es lässt sich auch die (Un)Ähnlichkeit als Abstand zu gegebenen Indikatoren messen (Al-Ibrahim u. a. 2017a). Als Variation findet sich die Differential Contribution (Li u. a. 2019), die nur zwei Feeds auf ähnliche Weise miteinander vergleicht.

Anforderungen Um Uniqueness messen zu können, muss entscheidbar sein, ob zwei Objekte Duplikate sind. Dies kann über exakte Wertgleichheit, Ähnlichkeit oder Identitätswerte geschehen. Unabhängig von der Methode ist es also wesentlich, dass Objekte `identifizierbar` sind. Die Metrik ist offensichtlich nur aussagekräftig mit einem gewissen Grad an Überlappung, auch wenn die Berechnung auch ohne Überlappung möglich ist.

Synonyme Uniqueness (Micic u. a. 2017; Behkamal u. a. 2014; Black und van Nederpelt 2020; Al-Ibrahim u. a. 2017b; Kuehn u. a. 2021), Originality (H. J. Griffioen 2022), Conciseness (Zaveri u. a. 2016), Differential Contribution (Li u. a. 2019), Similarity (Azevedo, Medeiros, und Bessani 2019), DataPointDistance (Gong, Cho, und Lee 2018), FeedIndependence (Gong, Cho, und Lee 2018), Originality (H. Griffioen, Booij, und Doerr 2020), ExclusiveContribution (Li u. a. 2019), Contribution (Meier u. a. 2018)

Volume Die Menge der Daten in einem Feed.

Beschreibung Die Dimension Volume beschreibt, wie viele Daten ein Feed bereitstellt. Meistens wird diese gemessen, indem Datenpunkte innerhalb eines Intervalls gezählt werden (Li u. a. 2019).

Anforderungen Für diese Dimension werden keine besonderen Attribute vorausgesetzt, da eine Anzahl immer bestimmt werden kann.

Synonyme Volume (Li u. a. 2019), Size (Meier u. a. 2018)

Anhang B - Qualitätsmetriken und Reifegrade

Diese Liste zeigt die betrachteten und bewerteten Metriken. Die Metriken sind nach den vereinheitlichten Qualitätsdimensionen gruppiert und mittels des Metrik-Reifegradmodells (RG) eingestuft. Zusätzlich gibt es noch Ansätze die Gesamtqualität in einer Metrik zusammenzufassen. Diese wurden unter der Pseudodimension “Quality” eingeordnet.

Dimension	Metrik	RG
Accessibility		
	QoS (Rashid, Noor, und Altmann 2019)	0
	WAQM (Hofer 2021)	2
Accuracy - Semantic	Correctness (Al-Ibrahim u. a. 2017b)	3
	Accuracy (Bouwman u. a. 2020)	1
	Accuracy2 (Bouwman u. a. 2020)	1
	FeedError (Gong, Cho, und Lee 2018)	0
	Accuracy (Grispos, Glisson, und Storer 2019)	1
	Accuracy (Kuehn u. a. 2021)	3
	Accuracy (Li u. a. 2019)	3
	Accuracy (Meier u. a. 2018)	2
	AlertFalsePositive (Mu u. a. 2014)	1
	FalsePositiveRatio (Pawł Pawliński und Kompanek 2016)	0
	Purity (Pitsillidis u. a. 2012)	3
	VariationDistance (Pitsillidis u. a. 2012)	3
	KendallRankCorrelationCoefficient (Pitsillidis u. a. 2012)	3
	FalsePositives (Schaberreiter u. a. 2019)	3
	FPR (Sinha, Bailey, und Jahanian 2008)	3
	FNR (Sinha, Bailey, und Jahanian 2008)	3
	FalsePositiveRate (Hofer 2021)	3
	Robustness (Canfora u. a. 2020)	2
Accuracy - Syntactic		
	SyntacticAccuracy (Schlette u. a. 2021)	3
Appropriate Amount of Data	Volume (Li u. a. 2019)	2
	PartnerSharingActivity (Mavzer u. a. 2021)	0
	Volume (Pitsillidis u. a. 2012)	2
	AppropriateAmaountOfData (Schlette u. a. 2021)	3
Believability		
	Reliability (Gong, Cho, und Lee 2018)	0
	ReliabilityDataPooint (Gong, Cho, und Lee 2018)	0
	CertifiedCybersecurity (Mavzer u. a. 2021)	0
	PreviousTicketRatings (Mavzer u. a. 2021)	0
	PartnerSector (Mavzer u. a. 2021)	0
	Reputation (Mavzer u. a. 2021)	0
	TrustedCommunities (Rashid, Noor, und Altmann 2019)	0
	Verifiability (Schaberreiter u. a. 2019)	3

Dimension	Metrik	RG
	Trust (Schaberreiter u. a. 2019)	3
	Reputation (Schlette u. a. 2021)	3
	Reliability (Gong, Cho, und Lee 2018)	3
	AdmiraltyCode (Hofer 2021)	0
	Trustworthiness (Prat und Madnick 2008)	3
	Reasonableness (Prat und Madnick 2008)	3
	TemporalBelievability (Prat und Madnick 2008)	3
	Authenticity (Chandel u. a. 2019)	0
Completeness		
	Coverage (Bouwman u. a. 2020)	1
	CoverageFalseNegatives (Bouwman u. a. 2020)	1
	Integrity (Chandel u. a. 2019)	0
	Completeness (Grispos, Glisson, und Storer 2019)	1
	Completeness (Kuehn u. a. 2021)	2
	DBCompleteness (Kuehn u. a. 2021)	2
	Coverage (Li u. a. 2019)	3
	Completeness (Kührer, Rossow, und Holz 2014)	3
	Volume (Kührer, Rossow, und Holz 2014)	2
	Completeness (Mavzer u. a. 2021)	0
	Extensiveness (Mavzer u. a. 2021)	0
	Completeness (Meier u. a. 2018)	3
	Coverage (Pinto 2018)	2
	TotalCoverage (Pitsillidis u. a. 2012)	2
	ExclusiveContribution (Pitsillidis u. a. 2012)	2
	DifferentialContribution (Pitsillidis u. a. 2012)	2
	Completeness (Qiang u. a. 2018)	0
	Extensiveness (Schaberreiter u. a. 2019)	3
	Completeness (Schaberreiter u. a. 2019)	3
	SchemaCompleteness (Schlette u. a. 2021)	3
	Correctness (Yucel u. a. 2020)	2
	Overlap (Thomas u. a. 2016)	3
	DiamondModelOfIntrusionAlert (Hofer 2021)	3
	CoveragePhishing (Sheng u. a. 2009)	3
	Effectiveness (Chandel u. a. 2019)	0
Compliance		
	Privacy (Mavzer u. a. 2021)	0
Concise Representation		
	Concise Representation (Schlette u. a. 2021)	3
Consistency		
	Consistency (Grispos, Glisson, und Storer 2019)	1
	Representational Consistency (Schlette u. a. 2021)	3
	Compliance (Schaberreiter u. a. 2019)	3
Objectivity		
	Objectivity (Schlette u. a. 2021)	3
Portability		

Dimension	Metrik	RG
	Ingestibility (Bouwman u. a. 2020)	0
	Customizability (Qiang u. a. 2018)	0
	Interoperability (Schaberreiter u. a. 2019)	3
Quality		
	QoI (Al-Ibrahim u. a. 2017b)	3
	CompositionScore (Chandel u. a. 2019)	2
	CTIRating (Mavzer u. a. 2021)	0
	Trustworthiness (Mavzer u. a. 2021)	0
	Quality (Mavzer u. a. 2021)	0
	AlertRisk (Mu u. a. 2014)	1
	QoI (Rashid, Noor, und Altmann 2019)	0
	ObjectQuality (Schlette u. a. 2021)	3
	ReportQuality (Schlette u. a. 2021)	3
	ARIMA (Kohlrausch und Brin 2020)	2
Relevance		
	Relevance (Al-Ibrahim u. a. 2017b)	3
	Relevance (Bouwman u. a. 2020)	0
	FeedWeight (Gong, Cho, und Lee 2018)	0
	SectorOfActivity (Mavzer u. a. 2021)	0
	Relevance (Mavzer u. a. 2021)	0
	Following (Metcalf und Spring 2015)	2
	AlertRelevance (Mu u. a. 2014)	1
	AlertCorrelation (Mu u. a. 2014)	1
	Utility (Paweł Pawliński und Kompanek 2016)	0
	Utility (Al-Ibrahim u. a. 2017b)	2
	Fitness (Pinto 2018)	2
	Impact (Pinto 2018)	3
	Relevance (Qiang u. a. 2018)	0
	Relevancy (Schlette u. a. 2021)	3
	Relevance (T. D. Wagner u. a. 2017)	2
	RecommenderSystem (Hofer 2021)	3
	Impact (H. Griffioen, Booij, und Doerr 2020)	1
Sensitivity		
	Sensitivity (H. Griffioen, Booij, und Doerr 2020)	1
	Looseness (Canfora u. a. 2020)	1
Timeliness		
	Timeliness (Bouwman u. a. 2020)	1
	Timeliness (Chandel u. a. 2019)	0
	Timeliness (H. Griffioen, Booij, und Doerr 2020)	1
	Timeliness (Grispos, Glisson, und Storer 2019)	1
	Latency (Li u. a. 2019)	1
	ReactionTime (Kührer, Rossow, und Holz 2014)	2
	Freshness (Mavzer u. a. 2021)	0
	Timeliness (Mavzer u. a. 2021)	0
	Speed (Meier u. a. 2018)	2

Dimension	Metrik	RG
	Delay (Pawł Pawliński und Kompanek 2016)	0
	Latency (Pitsillidis u. a. 2012)	2
	Timeliness (Qiang u. a. 2018)	0
	Timeliness (Rashid, Noor, und Altmann 2019)	0
	Timeliness (Schaberreiter u. a. 2019)	3
	Timeliness (Schlette u. a. 2021)	3
	TimeStampsAndUpdateLogs (Hofer 2021)	2
	PlatformPublishingRate (Miranda u. a. 2021)	3
	RelativeFreshness (Miranda u. a. 2021)	3
	MeanTimeBetweenPlatforms (Miranda u. a. 2021)	3
	DecayOfAttributes (Mokaddem u. a. 2019)	3
	Novelty (Pinto 2018)	3
	Aging (Pinto 2018)	3
	Maintenance (Schaberreiter u. a. 2019)	3
Uniqueness		
	EntityUniqueness (Al-Ibrahim u. a. 2017b)	2
	SetUniqueness (Al-Ibrahim u. a. 2017b)	2
	Similarity (Azevedo, Medeiros, und Bessani 2019)	3
	ContainedSimilarity (Azevedo, Medeiros, und Bessani 2019)	3
	Uniqueness (Chandel u. a. 2019)	0
	DataPointDistance (Gong, Cho, und Lee 2018)	0
	FeedDistance (Gong, Cho, und Lee 2018)	0
	FeedIndependence (Gong, Cho, und Lee 2018)	0
	Originality (H. Griffioen, Booij, und Doerr 2020)	1
	Uniqueness (Kuehn u. a. 2021)	2
	ClusterUniqueness (Kuehn u. a. 2021)	2
	DifferentialContribution (Li u. a. 2019)	3
	Intersection (Li u. a. 2019)	3
	ExclusiveContribution (Li u. a. 2019)	2
	Contribution (Meier u. a. 2018)	3
	ReverseCounts (Metcalf und Spring 2015)	3
	ListCounts (Metcalf und Spring 2015)	2
	Intelligence (Schaberreiter u. a. 2019)	3
	Similarity (Schaberreiter u. a. 2019)	3

Anhang - Literatur

In den Anhängen verwendete Quellen.

Al-Ibrahim, Omar, Aziz Mohaisen, Charles Kamhoua, Kevin Kwiat, und Laurent Njilla. 2017a. „Beyond Free Riding: Quality of Indicators for Assessing Participation in Information Sharing for Threat Intelligence“. arXiv. https://doi.org/10.48550/ARXIV.1702.00552.

——. 2017b. „Beyond Free Riding: Quality of Indicators for Assessing Participation in Information Sharing for Threat Intelligence“. *CoRR*, Februar. https://arxiv.org/abs/1702.00552.

Azevedo, Rui, Ibéria Medeiros, und Alysson Bessani. 2019. „PURE: Generating Quality Threat Intelligence by Clustering and Correlating OSINT“. In *Proceedings - 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering, TrustCom/BigDataSE 2019*, 483–90. New York, NY, USA: IEEE. https://doi.org/10.1109/TrustCom/BigDataSE.2019.00071.

Batini, Carlo, Cinzia Cappiello, Chiara Francalanci, und Andrea Maurino. 2009. „Methodologies for Data Quality Assessment and Improvement“. *ACM Comput. Surv.* 41 (3). https://doi.org/10.1145/1541880.1541883.

Batini, Carlo, und Monica Scannapieco. 2016. *Data and Information Quality: Dimensions, Principles and Techniques*. Herausgegeben von M. J. Carey und S. Ceri. 1. Aufl. Data-Centric Systems and Applications. Springer Cham. https://doi.org/10.1007/978-3-319-24106-7.

Behkamal, Behshid, Mohsen Kahani, Ebrahim Bagheri, und Zoran Jeremic. 2014. „A Metrics-Driven Approach for Quality Assessment of Linked Open Data“. *Journal of Theoretical and Applied Electronic Commerce Research* 9 (2): 64–79. https://doi.org/10.4067/S0718-18762014000200006.

Black, Andrew, und Peter van Nederpelt. 2020. „Dimensions of Data Quality (DDQ)“. Research Paper. DAMA NL Foundation. https://www.dama-nl.org/wp-content/uploads/2020/09/DDQ-Dimensions-of-Data-Quality-Research-Paper-version-1.2-d.d.-3-Sept-2020.pdf.

Bouwman, Xander, Harm Griffioen, Jelle Egbers, Christian Doerr, Bram Klievink, und Michel van Eeten. 2020. „A different cup of TI? The added value of commercial threat intelligence“. In *29th USENIX Security Symposium (USENIX Security 20)*, 433–50. Berkeley, CA, USA: USENIX Association. https://www.usenix.org/conference/usenixsecurity20/presentation/bouwman.

Canfora, Gerardo, Mimmo Carapella, Andrea Del Vecchio, Laura Nardi, Antonio Pirozzi, und Corrado Aaron Visaggio. 2020. „About the Robustness and Looseness of Yara Rules“. In *Testing Software and Systems*, herausgegeben von Valentina Casola, Alessandra De Benedictis, und Massimiliano Rak, 12543:104–20. Lecture Notes in Computer Science. Cham: Springer International Publishing.

Chandel, Sonali, Mengdi Yan, Shaojun Chen, Huan Jiang, und Tian-Yi Ni. 2019. „Threat Intelligence Sharing Community: A Countermeasure Against Advanced Persistent Threat“. In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 353–59. New York, NY, USA: IEEE. https://doi.org/10.1109/MIPR.2019.00070.

Fox, Christopher, Anany Levitin, und Thomas Redman. 1994. „The notion of data and its quality dimensions“. *Information Processing & Management* 30 (1): 9–19. https://doi.org/10.1016/0306-4573(94)90020-5.

Freire, André P., Renata P. M. Fortes, Marcelo A. S. Turine, und Debora M. B. Paiva. 2008. „An Evaluation of Web Accessibility Metrics Based on Their Attributes“. In *Proceedings of the 26th Annual ACM International Conference on Design of Communication*, 73–80. SIGDOC '08.

Lisbon, Portugal: Association for Computing Machinery. https://doi.org/10.1145/1456536.1456551.

Gong, Seonghyeon, Jaeik Cho, und Changhoon Lee. 2018. „A Reliability Comparison Method for OSINT Validity Analysis“. *IEEE Transactions on Industrial Informatics* 14 (12): 5428–35. https://doi.org/10.1109/TII.2018.2857213.

Griffioen, H. J. 2022. „Cyber Threat Intelligence: Analysis of adversaries and their methods“. Phdthesis, Delft University of Technology. https://doi.org/10.4233/uuid:37f7367f-bc5e-4cde-a7fd-47d12621f853.

Griffioen, Harm, Tim Booij, und Christian Doerr. 2020. „Quality Evaluation of Cyber Threat Intelligence Feeds“. In *Applied Cryptography and Network Security*, herausgegeben von Mauro Conti, Jianying Zhou, und Emiliano Casalicchio, 12147:277–96. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-57878-7_14.

Grispos, George, William Bradley Glisson, und Tim Storer. 2019. „How Good is Your Data? Investigating the Quality of Data Generated During Security Incident Response Investigations“. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 7156–65. Honolulu, HI: University of Hawai'i at Mānoa. https://hdl.handle.net/10125/60152.

Kahn, Beverly K., Diane M. Strong, und Richard Y. Wang. 2002. „Information Quality Benchmarks: Product and Service Performance“. *Commun. ACM* 45 (4): 184–92. https://doi.org/10.1145/505248.506007.

Kohlrausch, Jan, und Eugene A. Brin. 2020. „ARIMA Supplemented Security Metrics for Quality Assurance and Situational Awareness“. *Digital Threats* 1 (1): 1–21. https://doi.org/10.1145/3376926.

Kuehn, Philipp, Markus Bayer, Marc Wendelborn, und Christian Reuter. 2021. „OVANA: An Approach to Analyze and Improve the Information Quality of Vulnerability Databases“. In *Proceedings of the 16th International Conference on Availability, Reliability and Security*. ARES '21. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3465481.3465744.

Kührer, Marc, Christian Rossow, und Thorsten Holz. 2014. „Paint It Black: Evaluating the Effectiveness of Malware Blacklists“. In *Research in Attacks, Intrusions and Defenses*, herausgegeben von Angelos Stavrou, Herbert Bos, und Georgios Portokalidis, 8688:1–21. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-11379-1_1.

Li, Vector Guo, Matthew Dunn, Paul Pearce, Damon McCoy, Geoffrey M. Voelker, und Stefan Savage. 2019. „Reading the Tea leaves: A Comparative Analysis of Threat Intelligence“. In *28th USENIX Security Symposium (USENIX Security 19)*, 851–67. Berkeley, CA, USA: USENIX Association. https://www.usenix.org/conference/usenixsecurity19/presentation/li.

Mavzer, Kadir Burak, Ewa Konieczna, Henrique Alves, Cagatay Yucel, Ioannis Chalkias, Dimitrios Mallis, Deniz Cetinkaya, und Luis Angel Galindo Sanchez. 2021. „Trust and Quality Computation for Cyber Threat Intelligence Sharing Platforms“. In *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*, 360–65. New York, NY, USA: IEEE. https://doi.org/10.1109/CSR51186.2021.9527975.

Meier, Roland, Cornelia Scherrer, David Gugelmann, Vincent Lenders, und Laurent Vanbever. 2018. „FeedRank: A tamper-resistant method for the ranking of cyber threat intelligence feeds“. In *International Conference on Cyber Conflict (ICCC)*, 321–44. New York, NY, USA: IEEE. https://doi.org/10.23919/CYCON.2018.8405024.

Metcalf, Leigh, und Jonathan M. Spring. 2015. „Blacklist Ecosystem Analysis: Spanning Jan 2012 to Jun 2014“. In *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security*, 13–22. WISCS '15. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2808128.2808129.

Micic, Natasha, Daniel Neagu, Felician Campean, und Esmaeil Habib Zadeh. 2017. „Towards a Data Quality Framework for Heterogeneous Data“. In *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 155–62. New York, NY, USA: IEEE. https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2017.28.

Miranda, Lucas, Daniel Vieira, Leandro Pfleger de Aguiar, Daniel Sadoc Menasché, Miguel Angelo Bicudo, Mateus Schulz Nogueira, Matheus Martins, Leonardo Ventura, Lucas Senos, und Enrico Lovat. 2021. „On the Flow of Software Security Advisories“. Herausgegeben von Hanan Lutfiyya. *IEEE Transactions on Network and Service Management* 18 (2): 1305–20. https://doi.org/10.1109/TNSM.2021.3078727.

Mohaisen, Aziz, Omar Al-Ibrahim, Charles Kamhoua, Kevin Kwiat, und Laurent Njilla. 2017. „Assessing Quality of Contribution in Information Sharing for Threat Intelligence“. In *2017 IEEE Symposium on Privacy-Aware Computing (PAC)*, 182–83. https://doi.org/10.1109/PAC.2017.39.

Mokaddem, Sami, Gérard Wagener, Alexandre Dulaunoy, und Andras Iklody. 2019. „Taxonomy driven indicator scoring in MISP threat intelligence platforms“. https://arxiv.org/abs/1902.03914.

Mu, Chengpo, Meng Yu, Yingjiu Li, und Wanyu Zang. 2014. „Risk balance defense approach against intrusions for network server“. *International Journal of Information Security* 13 (3): 255–69. https://doi.org/10.1007/s10207-013-0214-9.

Park, Jemon, Hisham Alasmary, Omar Al-Ibrahim, Charlies Kamhoua, Kevin Kwiat, Laurent Njilla, und Aziz Mohaisen. 2018. „QOI: Assessing Participation in Threat Information Sharing“. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6951–55. https://doi.org/10.1109/ICASSP.2018.8462036.

Pawliński, Paweł, Przemysław Jaroszewski, Piotr Kijewski, Łukasz Siewierski, Paweł Jacewicz, Przemysław Zielony, und Radosław Żuber. 2014. *Actionable Information for Security Incident Response*. LU: ENISA. https://data.europa.eu/doi/10.2824/38111.

——. 2015. „Actionable Information for Security Incident Response“. Technical Report. European Union Agency for Network; Information Security. https://doi.org/10.2824/38111.

Pawliński, Pawł, und Andrew Kompanek. 2016. „Evaluating Threat Intelligence Feeds“. Presented at the FIRST Technical Colloquium for Threat Intelligence, Munich, DEU. https://www.first.org/resources/papers/munich2016/kompanek-pawlinski-evaluating-threat-ntelligence-feeds.pdf.

Pinto, Alex. 2018. „Determining the Fit and Impact of CTI Indicators on your Monitoring Pipeline (TIQ-Test 2.0)“. Presented at the FIRST 2018 conference, Kuala Lumpur, MYS. https://www.first.org/conference/2018/program#pdetermining-the-fit-and-impact-of-cti-indicators-on-your-monitoring-pipeline-tiq-test-2-0.

Pipino, Leo L., Yang W. Lee, und Richard Y. Wang. 2002. „Data Quality Assessment“. *Communications of the ACM* 45 (4): 211–18. https://doi.org/10.1145/505248.506010.

Pitsillidis, Andreas, Chris Kanich, Geoffrey M. Voelker, Kirill Levchenko, und Stefan Savage. 2012. „Taster's Choice: A Comparative Analysis of Spam Feeds“. In *Proceedings of the 2012*

Internet Measurement Conference, 427–40. IMC '12. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2398776.2398821.

Prat, Nicolas, und Stuart Madnick. 2008. „Measuring Data Believability: A Provenance Approach". In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, 393–93. https://doi.org/10.1109/HICSS.2008.243.

Qiang, Li, Jiang Zhengwei, Yang Zeming, Liu Baoxu, Wang Xin, und Zhang Yunan. 2018. „A Quality Evaluation Method of Cyber Threat Intelligence in User Perspective". In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, 269–76. New York, NY, USA: IEEE. https://doi.org/10.1109/TrustCom/BigDataSE.2018.00049.

Rashid, Zahid, Umara Noor, und Jörn Altmann. 2019. „Network Externalities in Cybersecurity Information Sharing Ecosystems". In *Economics of Grids, Clouds, Systems, and Services. GECON 2018*, herausgegeben von Massimo Coppola, Emanuele Carlini, Daniele D'Agostino, Jörn Altmann, und José Ángel Bañares, 11113:116–25. LNCCN. Cham: Springer. https://doi.org/10.1007/978-3-030-13342-9_10.

Schaberreiter, Thomas, Veronika Kupfersberger, Konstantinos Rantos, Arnolnt Spyros, Alexandros Papanikolaou, Christos Ilioudis, und Gerald Quirchmayr. 2019. „A Quantitative Evaluation of Trust in the Quality of Cyber Threat Intelligence Sources". In *Proceedings of the 14th International Conference on Availability, Reliability and Security*. ARES '19. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3339252.3342112.

Schlette, Daniel, Fabian Böhm, Marco Caselli, und Günther Pernul. 2021. „Measuring and visualizing cyber threat intelligence quality". *International Journal of Information Security* 20 (März): 21–38. https://doi.org/10.1007/s10207-020-00490-y.

Sheng, Steve, Brad Wardman, Gary Warner, Lorrie Faith Cranor, Jason Hong, und Zhang Chengshan. 2009. „An Empirical Analysis of Phishing Blacklists". In *Proceedings of Sixth Conference on Email and Anti-Spam*. https://doi.org/https://doi.org/10.1184/R1/6469805.v1.

Sillaber, Christian, Andrea Mussmann, und Ruth Breu. 2019. „Experience: Data and Information Quality Challenges in Governance, Risk, and Compliance Management". *Journal of Data and Information Quality* 11 (2): 1–11. https://doi.org/10.1145/3297721.

Sinha, Sushant, Michael Bailey, und Farnam Jahanian. 2008. „Shades of grey: On the effectiveness of reputation-based ‚blacklists'". In *2008 3rd International Conference on Malicious and Unwanted Software (MALWARE)*, 57–64. New York, NY, USA: IEEE. https://doi.org/10.1109/MALWARE.2008.4690858.

Taleb, Ikbal, Hadeel T. El Kassabi, Mohamed Adel Serhani, Rachida Dssouli, und Chafik Bouhaddioui. 2016. „Big Data Quality: A Quality Dimensions Evaluation". In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, 759–65. New York, NY, USA: IEEE. https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0122.

Talha, M., A. Abou El Kalam, und N. Elmarzouqi. 2019. „Big Data: Trade-off between Data Quality and Data Security". *Procedia Computer Science* 151: 916–22. https://doi.org/https://doi.org/10.1016/j.procs.2019.04.127.

Thomas, Kurt, Rony Amira, Adi Ben-Yoash, Ori Folger, Amir Hardon, Ari Berger, Elie Bursztein, und Michael Bailey. 2016. „The Abuse Sharing Economy: Understanding the Limits of Threat Exchanges". In *Research in Attacks, Intrusions, and Defenses*, herausgegeben von Fabian

Monrose, Marc Dacier, Gregory Blanc, und Joaquin Garcia-Alfaro, 9854:143–64. LNSC. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-45719-2_7.

Umbrich, Jürgen, Sebastian Neumaier, und Axel Polleres. 2015. „Quality Assessment and Evolution of Open Data Portals“. In *2015 3rd International Conference on Future Internet of Things and Cloud*, 404–11. New York, NY, USA: IEEE. https://doi.org/10.1109/FiCloud.2015.82.

Wagner, Cynthia, Alexandre Dulaunoy, Gérard Wagener, und Andras Iklody. 2016. „MISP: The Design and Implementation of a Collaborative Threat Intelligence Sharing Platform“. https://doi.org/10.1145/2994539.2994542.

Wagner, Thomas D., Khaled Mahbub, Esther Palomar, und Ali E. Abdallah. 2019. „Cyber threat intelligence sharing: Survey and research directions“. *Computers & Security* 87 (November). https://doi.org/10.1016/j.cose.2019.101589.

Wagner, Thomas D., Esther Palomar, Khaled Mahbub, und Ali E. Abdallah. 2017. „Relevance Filtering for Shared Cyber Threat Intelligence (Short Paper)“. In *Information Security Practice and Experience*, herausgegeben von Joseph K. Liu und Pierangela Samarati, 10701:576–86. LNSC. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-72359-4_35.

Wang, Richard Y., und Diane M. Strong. 1996. „Beyond Accuracy: What Data Quality Means to Data Consumers“. *Journal of Management Information Systems* 12 (4): 5–33. https://doi.org/10.1080/07421222.1996.11518099.

Wook, Muslihah, Nor Asiakin Hasbullah, Norulzahrah Mohd Zainudin, Zam Zarina Abdul Jabar, Suzaimah Ramli, Noor Afiza Mat Razali, und Nurhafizah Moziyana Mohd Yusop. 2021. „Exploring big data traits and data quality dimensions for big data analytics application using partial least squares structural equation modelling“. *Journal of Big Data* 8 (März). https://doi.org/10.1186/s40537-021-00439-5.

Yucel, Cagatay, Ioannis Chalkias, Dimitrios Mallis, Evangelos Karagiannis, Deniz Cetinkaya, und Vasilios Katos. 2020. „On the Assessment of Completeness and Timeliness of Actionable Cyber Threat Intelligence Artefacts“. In *Multimedia Communications, Services and Security*, herausgegeben von Andrzej Dziech, Wim Mees, und Andrzej Czyżewski, 1284:51–66. CCIS. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-59000-0_5.

Zaveri, Amrapali, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, und Soeren Auer. 2016. „Quality assessment for linked data: A survey“. *Semantic Web* 7 (1): 63–93. https://doi.org/10.3233/SW-150175.