Kategorisierung und Visualisierung von Datenschutzaspekten in Geschäftsprozessmodellen

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Melanie Windrich

geboren in
Duisburg

Eingereicht im im Mai 2023

1. Gutachter:	Prof. Dr. Andreas Speck
2. Gutachter:	Prof. Dr. Nils Gruschka

Datum der mündlichen Prüfung: 30.05.2023

Kurzfassung

In Zeiten zunehmender Digitalisierung gewinnt das Geschäftsprozessmanagement an Bedeutung. Hierbei werden relevante Geschäftsprozesse erhoben, analysiert, optimiert und umgesetzt. Ein wichtiger Bestandteil des Geschäftsprozessmanagements ist die grafische Modellierung der Prozesse. Die entstehenden Geschäftsprozessmodelle können unter anderem auch zur Wahrung der Compliance im Unternehmen beitragen. Ein aktuell viel diskutierter Bereich ist hier die Einhaltung des Datenschutzrechts, insbesondere der Europäischen Datenschutzgrundverordnung (DSGVO). Geschäftsprozessmodelle sind hier einerseits Unterstützung, um die dort verankerten Dokumentationspflichten zu erfüllen. Andererseits können die Prozessmodelle aber auch bei der Identifizierung und Optimierung von datenschutzrelevanten Vorgängen unterstützen. Allerdings lassen sich derartige Informationen in Prozessmodellen der verbreiteten Modellierungsnotationen nicht besonders leicht erkennen, da der Datenschutz nicht im Fokus der Entwickler steht und die Modellierungsnotationen keine speziellen Werkzeuge für die Repräsentation von Datenschutzaspekten enthalten.

Diese Arbeit beschreibt daher ein Konzept, Datenschutz in Prozessmodellen möglichst einfach nachvollziehbar darzustellen, indem einzelne Elemente der Geschäftsprozessmodelle in verschiedene Kategorien der Datenschutzrelevanz einsortiert und entsprechend in einer Ampelnotation eingefärbt werden. Die Grundlage hierfür bildet eine komplexe Systematik zur Kategorisierung der Modellelemente auf Basis verschiedener aus der DSGVO abgeleiteter Kriterien. Zu dem Konzept werden noch einige mögliche Ergänzungen vorgestellt. Darüber hinaus wird ein Prototyp gezeigt, der die Kategorisierung teilweise automatisiert. Hierfür wird unter Anderem auf Technologien aus dem Bereich des Maschinellen Lernens zurückgegriffen. Sowohl das Konzept selbst als auch der Prototyp werden positiv evaluiert.

Die Arbeit bietet einen Ansatz, das Bewusstsein für den Datenschutz bei allen beteiligten Personengruppen zu erhöhen. Das fördert die Einhaltung der entsprechenden Regularien und schützt betroffene Personen. Außerdem erleichtert das Konzept auch die Arbeit von Datenschutzbeauftragten und Aufsichtsbehörden.

Abstract

In times of ongoing digitization and digitalization, business process management is gaining more and more importance. Business process management means the discovery, analysis, optimization and technical implementation of relevant business processes. An important part of business process management is the graphical modeling of those processes, which can be used for different purposes like maintaining compliance in accordance to data protection laws, such as the European Data Protection Regulation (GDPR). In this context, business process models can be useful for fulfilling the documentation requirements. At the same time, process models can also help to identify critical parts of the business process and improve it with regard to data protection. However, one of the issues in identifying these critical parts is that the common modeling notations do not have any special utilities for integrating data protection aspects in process models. Because of that, the relevant information is difficult to identify.

This thesis describes a concept to represent data protection aspects in business process models as simple as possible. This is achieved by categorizing the relevant elements of a process model in three classes, indicating the criticality of the element regarding data protection. The elements are then colored according to their category. green is used for elements that are not relevant regarding data protection, yellow is used for elements that are relevant and red is used for critical elements. The foundation for the categorization is a complex system based on several criteria derived from the GDPR. Supplementary to this concept, some extensions are explained. Additionally a functioning prototype is presented that implements the concept and partially automatizes the categorization based on multiple approaches, including machine learning. The concept as well as the prototype are evaluated in the thesis.

In summary, this thesis provides an approach to increase data protection awareness for all involved groups of people. This promotes the compliance and protects the rights of the data subjects. Additionally the concept and the prototype presented supports the work of data protection officers and increases the efficiency of supervisory authorities.

Vorwort

An dieser Stelle möchte ich einigen wichtigen Menschen danken, die mich bei dieser Arbeit begleitet haben.

Allen voran möchte ich Prof. Dr. Andreas Speck für die großartige Betreuung und seine Unterstützung danken. Prof. Dr. Nils Gruschka danke ich von Herzen für seine Bereitschaft zur Begutachtung und für die allzeit gute Zusammenarbeit. Timo Wilgen gebührt größter Dank für das Korrekturlesen, außerdem für ein immer offenes Ohr in den vergangenen Jahren. Aljoscha Jagenow, Christian Mahrt, Mira Radonjic-Simic und Sven Niemand danke ich für viele produktive Gespräche und gute Ratschläge. Und auch allen anderen Menschen, die mich in den letzten Jahren begleitet und auf verschiedene Arten unterstützt haben, möchte ich meinen Dank aussprechen.

Besonders hervorheben möchte ich noch Julian für die Hilfe bei allen Grafik- und LaTeX-Problemen, aber auch für alles drum herum. Und zu guter Letzt danke ich dem kleinen Prinzen dafür, dass er immer da ist.

In memoriam Joachim Kurt Roering

Inhaltsverzeichnis

1. Einleitung

Die Digitalisierung ist ein zentrales Thema unserer Zeit. Über nahezu jede Branche finden sich aktuelle Berichte, welche die Relevanz dieses Thema hervorheben (siehe etwa [BMS18; GS23; HS20; Re20; Ri20; Sc20; Se21; SS21]). Gleichzeitig werden aber auch regelmäßig die Schwächen deutscher Unternehmen, insbesondere kleiner und mittelständischer (KMU)[Le20], aber auch der öffentlichen Verwaltung [Me22], dargelegt.

Auch Giesbert Rühl beschreibt in [Rü21] die enorme Relevanz der Digitalisierung, gerade auch im Kontext der Covid-19-Pandemie, die global zu einem sprunghaften Anstieg in der Nutzung aller Arten digitaler Dienste geführt hat. Er vertritt die Auffassung, dass die Digitalisierung in Europa und insbesondere auch in Deutschland einen viel zu negativen Ruf besitzt, sie tatsächlich aber gerade hier die Wirtschaft deutlich fördern könnte. Außerdem habe die Europäische Union im Vergleich zu anderen Großmächten, wie etwa den USA oder China strukturelle Nachteile. Deren Überwindung benötige zwangsweise eine ausgeprägte länderübergreifende europäische Zusammenarbeit.

Selbst der Europäische Rat und der Rat der Europäische Union (EU) haben erkannt, wie wichtig die Digitalisierung für die EU ist: „Der digitale Wandel ist ein Schlüsselelement für die wirtschaftliche Entwicklung und strategische Autonomie der EU“ [RE22]. Hierfür wurde eine umfangreiche Digitalstrategie ausgearbeitet (siehe [RE22]). Ein Element hiervon ist die sogenannte *Europäische Datenstrategie*, welche unter anderem ein angemessen hohes Maß an Datenschutz zum Ziel hat. Hierfür wurde 2016 die Datenschutzgrundverordnung (DSGVO) veröffentlicht.

Ein genereller Ansatz zur Digitalisierung ist das Geschäftsprozessmanagement[DP21; ST21], insbesondere das agile Geschäftsprozessmanagement [Ba19; ZL21].

In [Fl18] schreiben Fleischmann et al.: „Je klarer ein Unternehmen seine Geschäftsprozesse definiert und je konsequenter es diese im täglichen Geschehen umsetzt, umso leistungsfähiger ist es“. Für die möglichst klare Definition der Geschäftsprozesse, ist die Geschäftsprozessmodellierung ein sinnvolles Mittel[Fl18]. Allerdings können Geschäftsprozessmodelle immer nur einen Teil der Realität abbilden, wobei dieser Teil vom Fokus der modellierenden Personen abhängt. Ein Aspekt, der hier häufig weitestgehend außer Acht gelassen wird, ist die Compliance[GMS06]. Dabei bezeichnet Compliance die „Einhaltung von Gesetzen, Regeln und Normen“[He18]. Die Vernachlässigung ist oft

ungünstig, da die Compliance von Prozessen eine zentrale Rolle im Prozessmanagement spielen sollte[Ni22; SGN07].

Einen Teil der gesetzlichen Regeln, die eine Organisation einhalten muss, betreffen den Datenschutz. Nach den obigen Ausführungen, ist es also durchaus sinnvoll, dieses Thema in Geschäftsprozessmodellen darzustellen. Hierbei stellt sich allerdings die Frage, in welcher Form dies umgesetzt werden kann. Erste Ansätze hierfür finden sich in [Ag19; BCM19; BF20; BMS15; PMB17; Pu19a]. Diese werden näher in Kapitel 6 beschrieben. Kurz zusammengefasst finden sich hier zwei Grundprinzipien:

Ein Teil der Arbeiten nutzt Design Pattern, um Sachverhalte aus dem Datenschutz in einem Geschäftsprozessmodell einzubringen. Hier werden also gewissermaßen kleine Teilprozessmodelle entwickelt, die einen bestimmten Sachverhalt, beispielsweise das Einholen einer Einwilligung, abbilden. Diese können dann in komplexeren Geschäftsprozessmodellen weiterverwendet werden. Der andere Teil der Arbeiten führt zusätzliche oder angepasste Modellelemente für eine bestimmte Modellierungsnotation ein, die Hinweise auf datenschutzrelevante Vorgänge geben sollen. Beide Ansätze haben Stärken und Schwächen, wobei diese auch von der jeweiligen Zielsetzung des Anwenders abhängen.

Wenn das Ziel die Compliance des Modells, also die Korrektheit in Bezug auf das Datenschutzrecht, ist, bieten sich die Design Pattern an. Hier können klar definierte gesetzliche Vorgaben vergleichsweise einfach eingebaut werden. Allerdings sind nicht alle Vorgaben des Gesetzes so klar definiert, dass sie in einem allgemeingültigen Pattern abgebildet werden können.

Die Erweiterung der Modellierungsnotation eignet sich besser, um allgemein auf das Thema Datenschutz aufmerksam zu machen. Auch hiermit kann in einem gewissen Rahmen die Compliance des Prozesses verbessert werden, beispielsweise indem Probleme, die dank der Notationserweiterung auffallen, in der Optimierung des Prozesses behoben werden. Außerdem hilft dieser Ansatz unter Umständen bei der Dokumentation der Regeleinhaltung, welche im Datenschutzrecht in verschiedenen Formen gefordert wird. Die bisherigen Ansätze in diese Richtung sind allerdings allesamt relativ komplex und erfordern einen vergleichsweise hohen Einarbeitungsaufwand. Dies ist gerade deshalb nachteilig, weil Prozessmodelle häufig auch als Hilfsmittel zur Kommunikation genutzt werden [CKO92]. Hier könnte der zusätzliche Informationsgehalt grundsätzlich große Vorteile z.B. in der Kommunikation zwischen Prozessmanagement, Datenschutzbeauftragten und Führung bieten. Allerdings müssen hierfür natürlich alle Beteiligten zumindest ein rudimentäres Verständnis der Notationselemente haben. Es kann aber nicht von allen Beteiligten eine entsprechende Einarbeitung verlangt werden. Außerdem werden sowohl die Geschäftsprozessmodellierung als auch der Datenschutz häufig eher als Belastung wahrgenommen, da sie einen relativ hohen Aufwand aber vermeintlich keinen unmittelbaren Nutzen mit sich bringen[Al09; In09; Ne22; SW20]. Wenn der

Aufwand hier durch die Einarbeitung in eine solche Notationserweiterung noch höher wird, reduziert dies die Motivation zusätzlich.

1.1. Ziele der Arbeit

Diese Arbeit setzt sich daher das Ziel, ein möglichst einfaches und intuitives Konzept zu entwickeln und anschließend zu evaluieren um Datenschutzaspekte in Geschäftsprozessmodellen darzustellen.

Hierfür wird die Business Process Model and Notation (BPMN) als beispielhafte Prozessmodellierungsnotation verwendet, da diese sowohl im wissenschaftlichen Umfeld als auch in der Praxis sehr verbreitet ist. Die einzelnen Modellelemente werden dann bzgl. ihrer Datenschutzrelevanz in drei Kategorien eingeteilt. Für die Visualisierung dieser drei Kategorien werden in erster Linie farbliche Markierungen, genauer in den Farben Rot, Gelb und Grün, verwendet, da diese nachweislich intuitiv verständlich sind und unmittelbar die gewünschten Emotionen im Betrachter auslösen (siehe hierfür Abschnitt 4.1).

In einem weiteren Schritt wird ein Prototyp entwickelt, der auch die teilautomatisierte Kategorisierung der Prozesselemente unterstützt und zur Evaluation des Konzepts verwendet werden kann. Auch hier bietet sich BPMN als Basis an, da diese die Möglichkeit der Erweiterung mit bringt, was das Vorhaben unterstützt.

Der beschriebene Ansatz kann, ähnlich wie die oben bereits kurz eingeführten verwandten Arbeiten, für verschiedene Zwecke und Zielgruppen eingesetzt werden. Einige potentiell profitierende Nutzergruppen nennen beispielsweise Bartolini et al., die in [BMS15] und [BCM19] einen Ansatz verfolgen, in dem Prozessmodelle in Hinblick auf Datenschutzanforderungen mit Icons annotiert werden sollen. Als potentiell profitierende Nutzer nennt [BMS15] zunächst die Verantwortlichen[1], welchen durch die Annotation ihre Pflichten bzgl. des Datenschutzes klar dargelegt werden. Es werden aber auch Auditoren und Datenschutzaufsichtsbehörden genannt. Diesen soll der Ansatz einerseits ermöglichen, auf den ersten Blick die Compliance mit den entsprechenden Datenschutzregularien erkennen, andererseits aber auch einen strukturierten Weg bieten, mögliche Datenschutzverletzungen aufzudecken.

Selbige Ziele können auch mit dem in dieser Arbeit präsentierten Konzept erreicht werden. Es sind allerdings noch Weitere denkbar:

Zunächst sollte hier die verbesserte Kommunikation zwischen unterschiedlichen – direkt oder indirekt – Beteiligten eines Prozesses genannt werden. Insbesondere ein Datenschutzbeauftragter kann mit Hilfe der eingefärbten Prozessmodelle einfacher mit anderen Beteiligten, die weniger Kenntnisse in diesem Bereich besitzen, über mögliche

[1] Im Sinne von Art 4 Nr. 7

Datenschutzprobleme kommunizieren. Dies wiederum kann einige Vorteile mit sich bringen und fördert auch die oben genannten Ziele wie die klare Darlegung der Pflichten des Verantwortlichen.

Außerdem kann das Verständnis und Bewusstsein für Datenschutzprobleme auch ohne Einbeziehung des Datenschutzbeauftragten in den Fachabteilungen steigen und damit das Datenschutzniveau im Unternehmen insgesamt ansteigen, bei gleichzeitiger Entlastung des Datenschutzbeauftragten.

Generell – sei es durch die verbesserte Kommunikation oder auch durch das bessere Verständnis aller Beteiligten – kann das Konzept auch dazu führen, dass Fehler im Prozess frühzeitig gefunden werden. Aus dem Bereich der Qualitätssicherung ist bekannt, dass früh gefundene Fehler deutlich günstiger zu beheben sind, als solche, die in späteren Entwicklungsphasen gefunden werden. Dies lässt sich direkt auch auf den Datenschutz in internen Prozessen übertragen. Schließlich sind einerseits weniger Änderungen in schon bestehenden Prozessen nötig, wenn Fehler schon bei der Konzeption eines neues Prozesses gefunden werden. Andererseits werden so aber auch direkte Kosten durch Bußgelder oder auch zivilrechtliche Zahlungen wegen Datenschutzverstößen verhindert.

Unabdingbar für die Erreichung dieser Ziele ist eine möglichst hohe Akzeptanz bei allen Nutzergruppen. Hierfür ist, wie oben bereits erläutert, ein möglichst intuitives Verständnis und eine geringe Einarbeitungszeit, sowie möglichst wenig zusätzlicher Aufwand für alle Beteiligten, enorm wichtig.

In dieser Arbeit wird im wesentlichen die Situation in Deutschland betrachtet. Dies ist insbesondere wegen der rechtlichen Vorgaben, aber unter Umständen auch wegen kultureller Gegebenheiten relevant. Die Ergebnisse lassen sich zumindest in Teilen wahrscheinlich auch auf andere Nationen, insbesondere der EU übertragen. Ein erster Abgleich ist mit norwegischen Partnern erfolgt (siehe Abschnitt 10.2.1).

1.2. Methodik

Die Arbeit ist im Bereich Design Science (siehe z.B. [BHM20]) anzusiedeln. Design Science stellt eine Zwischenstufe zwischen der erklärenden Wissenschaft auf der einen Seite und der Praxis auf der anderen Seite dar. Hierbei wird im Gegensatz zur erklärenden Wissenschaft nicht nach der Ursache für ein Problem gesucht, sondern mit Hilfe des erklärenden Wissens versucht, eine neuartige Lösung zu entwickeln, welche dann in der Praxis umgesetzt werden kann[Bu85; SMB21]. Ziel ist die Entwicklung eines möglichst nützlichen Artefakts[Pe07]. Grundsätzlich sind dabei alle drei Disziplinen von einander abhängig, wie Abbildung 1.1 zeigt.

Die Problemstellung, die mit dem Design Science Ansatz bearbeitet wird, kommt dabei in der Regel direkt aus der Praxis. So ist es im Grunde auch im vorliegenden

Abbildung 1.1.: Abgrenzung Design Science (nach [SMB21])

Fall: Wie oben dargelegt, müssen alle Unternehmen die einschlägigen Datenschutzregeln einhalten. Dafür müssen die datenschutzrelevanten Abläufe aber zwingend auch dokumentiert werden, um deren Einhaltung sicherzustellen[2]. Viele, gerade größere, Unternehmen nutzen ohnehin Geschäftsprozessmodelle um die internen Abläufe zu strukturieren und zu dokumentieren. Hier bietet sich eine Kombination beider Themen an.

In [He04] werden sieben Richtlinien vorgeschlagen, die im Design Science befolgt werden sollten:

1. **„Design as an Artifact“:** Die Forschung sollte ein Artefakt, wie etwa ein Modell oder eine Methode erstellen.
2. **„Problem Relevance:** Ziel ist ein technische Lösung zu einem relevanten Problem aus der Praxis.
3. **„Design Evaluation“:** Das erstellte Artefakt muss ordentlich evaluiert werden.
4. **„Research Contributions“:** Die Forschung sollte nützliche und überprüfbare Wissenschaftliche Beiträge liefern.
5. **„Research Rigor“:** Sowohl in der Entwicklung als auch der Evaluation müssen strenge Methoden angewandt werden.
6. **„Design as a Search Process“:** Design Science ist ein iterativer Prozess, der eine möglichst gute Lösung für ein Problem sucht.
7. **„Communication of Research“:** Die Forschungsergebnisse sollten Fachpublikum sowohl aus dem technischen als auch dem Managementbereich präsentiert werden.

Diese werden in den einzelnen Arbeitsschritten dieser Arbeit umgesetzt, die in Abbildung Abbildung 1.2 abgebildet werden.

[2] Aber auch, weil die DSGVO einige Dokumentationspflichten vorschreibt

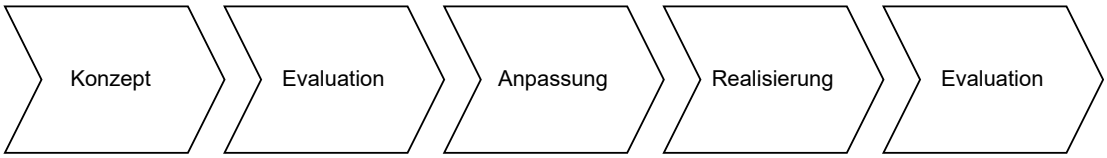


Abbildung 1.2.: Vorgehen der Arbeit

Zunächst wird ein Konzept entwickelt, wie mit Hilfe von Farben Datenschutzaspekte in Geschäftsprozessmodellen visualisiert werden können und nach welchen Kriterien die einzelnen Modellelemente hierfür kategorisiert werden müssen. Anschließend wird dieses Konzept umfassend durch Experteninterviews und eine Vergleichsstudie evaluiert. Auf Basis der Ergebnisse wird das Konzept um einige zusätzliche Elemente erweitert. Es folgt die Realisierung des Konzepts durch die Entwicklung eines Prototyps, welcher abschließend wieder evaluiert wird.

Die Design Science Richtlinien werden dabei folgendermaßen umgesetzt:

1. **„Design as an Artifact“:** Als Artefakt kann definitiv der Prototyp, aber auch das Konzept als solches im Sinne einer Methode gewertet werden.

2. **„Problem Relevance:** Die Praxisrelevanz wurde oben bereits ausgeführt.

3. **„Design Evaluation“:** Sowohl die Methode als auch der Prototyp werden evaluiert.

4. **„Research Contributions“:** Auf Basis der Evaluation zeigt sich ein nützlicher Beitrag für die Praxis.

5. **„Research Rigor“:** Im ganzen Designprozess wird mit strengen wissenschaftlichen Methoden gearbeitet. Es besteht allerdings das Problem, dass die Datenschutzgesetzgebung vergleichsweise vage formuliert ist. Daher sind an vielen Stellen keine strengen mathematischen Definitionen möglich. Durch das entwickelte Konzept wird diese unschärfe aber möglichst weit abstrahiert und in klare Kategorien sortiert.

6. **„Design as a Search Process“:** Der iterative Prozess zeigt sich in Abbildung 1.2.

7. **„Communication of Research“:** Der Kern des Konzepts wurde bereits in [WSG21] veröffentlicht. Hinzu kommt diese Dissertation und weitere geplante Ausarbeitungen.

1.3. Aufbau der Arbeit

Diese Arbeit besteht neben dieser Einleitung aus drei Hauptteilen Teilen.

Zunächst folgen in Teil I die wichtigsten Grundlagen, die für das weitere Verständnis dieser Arbeit notwendig sind. Diese betreffen die Kernthemen dieser Arbeit, nämliche den Datenschutz, die Geschäftsprozessmodellierung und die Visualisierung, aber auch technische Hintergründe wie ausgewählte Bereiche der Künstlichen Intelligenz und Ontologien.

Anschließend wird in Teil II das entwickelte Konzept präsentiert. Hierfür wird zunächst ein Überblick über den Stand der Forschung in verwandten Entwicklungen gegeben. Daraufhin werden drei Beispiele für datenschutzkritische Geschäftsprozesse eingeführt die im Folgenden zur Veranschaulichung dienen. Nach der Darstellung des eigentlichen Konzepts zur Visualisierung von Datenschutzaspekten in Geschäftsprozessmodellen, wird dieses auf die drei Beispielprozesse angewandt. Anschließend folgt eine erste Evaluation des vorgestellten Konzepts. Der Teil endet im nächsten Kapitel mit einer theoretischen Aufarbeitung der verschiedenen Möglichkeiten zur Automatisierung. wird.

In Teil III wird dann die Entwicklung eines Prototyps, als Proof of Concept des zuvor dargestellten Ansatzes, beschrieben. Hierfür erfolgt zunächst eine Anforderungsanalyse, bevor auf einige Details der Realisierung und letztlich deren Evaluation eingegangen wird.

Der letzte Teil IV fasst die Arbeit zusammen und liefert einen Überblick über zukünftige Arbeiten im Kontext dieser Arbeit und weitere Fragestellungen in diesem Zusammenhang.

In dieser Arbeit wird aus Gründen der besseren Lesbarkeit das generische Maskulinum verwendet. Andere Geschlechteridentitäten werden dabei ausdrücklich inkludiert.

Teil I.

Grundlagen

Dieser Teil beschreibt die wichtigsten Grundlagen, die für das Verständnis der folgenden Ausführungen essenziell sind. Dies sind zunächst eine Grundlagen zum Datenschutz, insbesondere den Vorgaben der Europäischen Datenschutzgrundverordnung. Anschließend wird das Thema Geschäftsprozessmodellierung am Beispiel der Business Process Model and Notation erläutert. Daraufhin folgen grundlegende Informationen zum Thema Design mit Fokus auf Modellierungsnotationen. Es wird auf die Fragestellung eingegangen, welche Formen der Visualisierung nach dem Stand der Forschung für welche Zwecke am besten geeignet sind. Ein besonderer Fokus liegt hier auf der Visualisierung durch Farbe, weshalb auch einige Studien zur Farbtheorie dargelegt werden. Abschließend folgt noch ein Kapitel zum Thema Künstliche Intelligenz. Im Speziellen wird hier auf die Methode des Maschinellen Lernen und das Anwendungsfeld Natural Language Processing eingegangen.

2. Datenschutz

Die Grundlage des Datenschutzes bildet in Deutschland das Recht auf informationelle Selbstbestimmung. Seinen Ursprung hat dieses 1983 im sogenannten Volkszählungsurteil des Bundesvefassungsgerichts [Bu83], welches das Recht auf informationelle Selbstbestimmung aus Art. 2 i. V. m. Art. 1 des Grundgesetzes ableitet [Al17].

Mit der Anwendung der Europäischen Datenschutz-Grundverordnung (DSGVO) [DA18] im Mai 2018 wurde der Datenschutz im ganzen Europäischen Wirtschaftsraum (EWR), also der EU, Norwegen, Island und Liechtenstein weitestgehend vereinheitlicht und nationale Regelungen größtenteils abgelöst. In diesem Kapitel werden daher zunächst die wichtigsten Begrifflichkeiten dieser Verordnung erläutert, die im folgenden verwendet werden. Anschließend folgt ein Abschnitt zum Standard-Datenschutzmodell, welches eine Hilfestellung für deren Umsetzung bieten soll (siehe Abschnitt 2.2). Abschließend werden noch kurz einige weitere relevante Gesetze eingeführt, die sich mit dem Datenschutz befassen und im Verlauf dieser Arbeit verwendet werden.

2.1. Europäische Datenschutzgrundverordnung

Die DSGVO regelt im Wesentlichen die Verarbeitung personenbezogener Daten. In diesem Kapitel wird zunächst erläutert, was überhaupt der Anwendungsbereich der DSGVO ist. Anschließend wird erläutert was nach Definition der DSGVO personenbezogene Daten sind (siehe Abschnitt 2.1.2). In Abschnitt 2.1.3 wird dann dargelegt, was unter dem Begriff „*Verarbeitung*“ zu verstehen ist, woraufhin in Abschnitt 2.1.5 dann auf die verschiedenen beteiligten Personengruppen eingegangen wird.

2.1.1. Anwendungsbereich

Der Anwendungsbereich der DSGVO wird in den Artikeln 2 und 3 definiert. Dort wird zwischen dem sachlichen und dem räumlichen Anwendungsbereich unterschieden.

Sachlicher Anwendungsbereich

Artikel 2 der DSGVO besagt zunächst, dass die Verordnung immer dann gilt, wenn personenbezogene Daten (teil-)automatisiert verarbeitet werden oder aber bei nicht-

automatisierter Verarbeitung in einem Dateisystem gespeichert werden. Was genau ein Dateisystem ist, wird an dieser Stelle nicht näher erläutert. In Erwägungsgrund 15 der DSGVO findet sich hierzu aber der Satz: „Akten oder Aktensammlungen sowie ihre Deckblätter, die nicht nach bestimmten Kriterien geordnet sind, sollten nicht in den Anwendungsbereich dieser Verordnung fallen“. Dies bedeutet im Umkehrschluss, dass auch Papierakten sehr wohl in den Anwendungsbereich fallen, wenn diese „nach bestimmten Kriterien geordnet“ werden.

In Art. 2 werden dann auch noch einige Ausnahmen der obigen Regel aufgeführt. Dies sind einerseits einige spezielle EU-Regularien auf die an dieser Stelle nicht näher eingegangen werden soll. Andererseits werden dort aber auch Datenverarbeitungen zu persönlichen oder familiären Zwecken, sowie für die Strafverfolgung und dem Schutz der öffentlichen Sicherheit, ausgenommen.

Räumlicher Anwendungsbereich

Zum räumlichen Anwendungsbereich wird in Art. 3 der DSGVO im Wesentlichen festgelegt, dass selbige immer dann Anwendung findet, wenn:

1. Der für die Datenverarbeitung Verantwortliche (oder ein Auftragsverarbeiter) seinen Sitz in der EU hat;

2. Der von der Datenverarbeitung Betroffene sich in der EU befindet; oder

3. Wenn die Datenverarbeitung an einem Ort stattfindet, der „aufgrund Völkerrechts dem Recht eines Mitgliedstaats unterliegt“.

Insgesamt fallen also eine Vielzahl von Datenverarbeitungen unter die Vorschriften der DSGVO, nämlich all jene, bei denen entweder der Verarbeitende oder der Betroffene in der EU ansässig ist.

2.1.2. Personenbezogene Daten

Die DSGVO regelt ausschließlich den Umgang mit *personenbezogenen Daten* Art. 1 Abs. 1 DSGVO. Diese werden in Art. 4 Nr. 1 definiert als „[...] Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen [...]“. Das bedeutet insbesondere, dass keine Unternehmensdaten und Geschäftsgeheimnisse unter dem Schutz der DSGVO (und auch der sonstigen Datenschutzgesetzgebung) stehen.

Unter besonderem Schutz stehen die sogenannten *besonderen Kategorien personenbezogener Daten*, wie sie in Art. 9 DSGVO definiert werden. Dies betrifft:

„Die Verarbeitung personenbezogener Daten, aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person[...]."

2.1.3. Verarbeitung personenbezogener Daten

Nachdem nun geklärt ist, was überhaupt personenbezogene Daten sind, stellt sich die Frage, was genau unter der Verarbeitung solcher Daten zu verstehen ist und welche Vorschriften hierfür gelten.

Arten von Verarbeitungstätigkeiten

Eine Liste mit verschiedenen Arten von Tätigkeiten, die als Verarbeitung personenbezogener Daten gewertet werden können, findet sich in Art. 4 DSGVO:

> Im Sinne dieser Verordnung bezeichnet der Ausdruck: [...]
>
> 2. „Verarbeitung" jeden mit oder ohne Hilfe automatisierter Verfahren ausgeführten Vorgang oder jede solche Vorgangsreihe im Zusammenhang mit personenbezogenen Daten wie das Erheben, das Erfassen, die Organisation, das Ordnen, die Speicherung, die Anpassung oder Veränderung, das Auslesen, das Abfragen, die Verwendung, die Offenlegung durch Übermittlung, Verbreitung oder eine andere Form der Bereitstellung, den Abgleich oder die Verknüpfung, die Einschränkung, das Löschen oder die Vernichtung;

Im weiteren Verlauf werden einige Tätigkeiten noch genauer definiert:

> 3. „Einschränkung der Verarbeitung" [bezeichnet] die Markierung gespeicherter personenbezogener Daten mit dem Ziel, ihre künftige Verarbeitung einzuschränken;
>
> 4. „Profiling" [bezeichnet] jede Art der automatisierten Verarbeitung personenbezogener Daten, die darin besteht, dass diese personenbezogenen Daten verwendet werden, um bestimmte persönliche Aspekte, die sich auf eine natürliche Person beziehen, zu bewerten, insbesondere um Aspekte bezüglich Arbeitsleistung, wirtschaftliche Lage,

Gesundheit, persönliche Vorlieben, Interessen, Zuverlässigkeit, Verhalten, Aufenthaltsort oder Ortswechsel dieser natürlichen Person zu analysieren oder vorherzusagen;

5. „Pseudonymisierung“ [bezeichnet] die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden; [...]

Grundsätze für die Verarbeitung

Grundsätze für die Verarbeitung personenbezogener Daten werden in Art. 5 Abs. 1 DSGVO geregelt. Zusammengefasst werden können diese mit den folgenden Stichworten:

- Rechtmäßigkeit, Verarbeitung nach Treu und Glauben, Transparenz
- Zweckbindung
- Datenminimierung
- Richtigkeit
- Speicherbegrenzung
- Integrität und Vertraulichkeit

Nähere Erläuterungen finden sich beispielsweise in Erwägungsgrund 39 der DSGVO.

Besonders hervorzuheben ist hier der erste Aspekt, die Rechtmäßigkeit. Hierzu werden in Art. 6 Abs. 1 sechs verschiedene Sachverhalte aufgeführt, die eine Datenverarbeitung rechtfertigen:

> „Die Verarbeitung ist nur rechtmäßig, wenn mindestens eine der nachstehenden Bedingungen erfüllt ist:
>
> a) Die betroffene Person hat ihre Einwilligung zu der Verarbeitung der sie betreffenden personenbezogenen Daten für einen oder mehrere bestimmte Zwecke gegeben;

b) die Verarbeitung ist für die Erfüllung eines Vertrags, dessen Vertragspartei die betroffene Person ist, oder zur Durchführung vorvertraglicher Maßnahmen erforderlich, die auf Anfrage der betroffenen Person erfolgen;

c) die Verarbeitung ist zur Erfüllung einer rechtlichen Verpflichtung erforderlich, der der Verantwortliche unterliegt;

d) die Verarbeitung ist erforderlich, um lebenswichtige Interessen der betroffenen Person oder einer anderen natürlichen Person zu schützen;

e) die Verarbeitung ist für die Wahrnehmung einer Aufgabe erforderlich, die im öffentlichen Interesse liegt oder in Ausübung öffentlicher Gewalt erfolgt, die dem Verantwortlichen übertragen wurde;

f) die Verarbeitung ist zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich, sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz personenbezogener Daten erfordern, überwiegen, insbesondere dann, wenn es sich bei der betroffenen Person um ein Kind handelt.

Unterabsatz 1 Buchstabe f gilt nicht für die von Behörden in Erfüllung ihrer Aufgaben vorgenommene Verarbeitung.“

Hierbei wiederum bedarf Buchstabe a, in dem die Einwilligung als rechtmäßiger Verarbeitungsgrund genannt wird, näherer Betrachtung.

2.1.4. Einwilligung

Art. 7 DSGVO führt nämlich einige Bedingungen für diese Einwilligung auf:

(1) Beruht die Verarbeitung auf einer Einwilligung, muss der Verantwortliche nachweisen können, dass die betroffene Person in die Verarbeitung ihrer personenbezogenen Daten eingewilligt hat.

(2) Erfolgt die Einwilligung der betroffenen Person durch eine schriftliche Erklärung, die noch andere Sachverhalte betrifft, so muss das Ersuchen um Einwilligung in verständlicher und leicht zugänglicher Form in einer klaren und einfachen Sprache so erfolgen, dass es von den anderen Sachverhalten klar zu unterscheiden ist. Teile der Erklärung sind dann nicht verbindlich, wenn sie einen Verstoß gegen diese Verordnung darstellen.

(3) Die betroffene Person hat das Recht, ihre Einwilligung jederzeit zu widerrufen. Durch den Widerruf der Einwilligung wird die Rechtmäßigkeit der aufgrund der Einwilligung bis zum Widerruf erfolgten Verarbeitung nicht berührt. Die betroffene Person wird vor Abgabe der Einwilligung hiervon in Kenntnis gesetzt. Der Widerruf der Einwilligung muss so einfach wie die Erteilung der Einwilligung sein.

(4) Bei der Beurteilung, ob die Einwilligung freiwillig erteilt wurde, muss dem Umstand in größtmöglichem Umfang Rechnung getragen werden, ob unter anderem die Erfüllung eines Vertrags, einschließlich der Erbringung einer Dienstleistung, von der Einwilligung zu einer Verarbeitung von personenbezogenen Daten abhängig ist, die für die Erfüllung des Vertrags nicht erforderlich sind.

Die Einwilligung muss also insbesondere freiwillig und widerrufbar sein und nachgewiesen werden können. Weitere Bedingungen liegen für Kinder vor (siehe Art. 8 DSGVO).

Allein diese Anforderungen machen die Einwilligung zu einem eher ungünstigen Verarbeitungsgrund, weil schon die Einholung mit recht hohem Aufwand einhergeht. Zusätzlich es auch im Nachgang Probleme mit sich bringen, wenn eine Einwilligung als einzige Verarbeitungsgrundlage vorliegt, da der Verantwortliche sie beispielsweise nachweisen können muss und der Betroffene, wie oben schon erwähnt, jederzeit seine Einwilligung widerrufen kann, womit die Rechtsgrundlage für die Datenverarbeitung wegfällt, was einen recht hohen Aufwand mit sich bringen kann.

Für besondere Kategorien personenbezogener Daten gilt nach Art. 9 Abs 2 lit. a DSGVO im Übrigen zusätzlich du den allgemeinen Anforderungen an eine Einwilligung, dass diese „ausdrücklich“ erteilt werden muss. Das heißt, eine konkludente Einwilligung, also eine solche, die sich aus dem Verhalten der betroffenen Person ableiten lässt, aber nicht explizit ausgesprochen wird, ist nicht zulässig [Ve21].

2.1.5. Beteiligte Rechtssubjekte

Neben den personenbezogenen Daten und der Verarbeitung selbiger werden in Art. 4 der DSGVO die einige Arten von Rechtssubjekten, bzw. deren Rollen in Bezug auf die Verarbeitung personenbezogener Daten definiert:

„Im Sinne dieser Verordnung bezeichnet der Ausdruck:[...]

7. „Verantwortlicher“ die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen

Daten entscheidet; sind die Zwecke und Mittel dieser Verarbeitung durch das Unionsrecht oder das Recht der Mitgliedstaaten vorgegeben, so kann der Verantwortliche beziehungsweise können die bestimmten Kriterien seiner Benennung nach dem Unionsrecht oder dem Recht der Mitgliedstaaten vorgesehen werden;

8. „Auftragsverarbeiter“ eine natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die personenbezogene Daten im Auftrag des Verantwortlichen verarbeitet;

9. „Empfänger“ eine natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, der personenbezogene Daten offengelegt werden, unabhängig davon, ob es sich bei ihr um einen Dritten handelt oder nicht. Behörden, die im Rahmen eines bestimmten Untersuchungsauftrags nach dem Unionsrecht oder dem Recht der Mitgliedstaaten möglicherweise personenbezogene Daten erhalten, gelten jedoch nicht als Empfänger; die Verarbeitung dieser Daten durch die genannten Behörden erfolgt im Einklang mit den geltenden Datenschutzvorschriften gemäß den Zwecken der Verarbeitung;

10. „Dritter“ eine natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, außer der betroffenen Person, dem Verantwortlichen, dem Auftragsverarbeiter und den Personen, die unter der unmittelbaren Verantwortung des Verantwortlichen oder des Auftragsverarbeiters befugt sind, die personenbezogenen Daten zu verarbeiten;[...]

16. „Hauptniederlassung“

    a) im Falle eines Verantwortlichen mit Niederlassungen in mehr als einem Mitgliedstaat den Ort seiner Hauptverwaltung in der Union, es sei denn, die Entscheidungen hinsichtlich der Zwecke und Mittel der Verarbeitung personenbezogener Daten werden in einer anderen Niederlassung des Verantwortlichen in der Union getroffen und diese Niederlassung ist befugt, diese Entscheidungen umsetzen zu lassen; in diesem Fall gilt die Niederlassung, die derartige Entscheidungen trifft, als Hauptniederlassung;

    b) im Falle eines Auftragsverarbeiters mit Niederlassungen in mehr als einem Mitgliedstaat den Ort seiner Hauptverwaltung in der Union oder, sofern der Auftragsverarbeiter keine Hauptverwaltung in der Union hat, die Niederlassung des Auftragsverarbeiters in der Union, in der die Verarbeitungstätigkeiten im Rahmen der Tätigkeiten einer Niederlassung eines Auftragsverarbeiters haupt-

sächlich stattfinden, soweit der Auftragsverarbeiter spezifischen Pflichten aus dieser Verordnung unterliegt;

17. „Vertreter“ eine in der Union niedergelassene natürliche oder juristische Person, die von dem Verantwortlichen oder Auftragsverarbeiter schriftlich gemäß Artikel 27 bestellt wurde und den Verantwortlichen oder Auftragsverarbeiter in Bezug auf die ihnen jeweils nach dieser Verordnung obliegenden Pflichten vertritt;
18. „Unternehmen“ eine natürliche oder juristische Person, die eine wirtschaftliche Tätigkeit ausübt, unabhängig von ihrer Rechtsform, einschließlich Personengesellschaften oder Vereinigungen, die regelmäßig einer wirtschaftlichen Tätigkeit nachgehen;
19. „Unternehmensgruppe“ eine Gruppe, die aus einem herrschenden Unternehmen und den von diesem abhängigen Unternehmen besteht; [...]
21. „Aufsichtsbehörde“ eine von einem Mitgliedstaat gemäß Artikel 51 eingerichtete unabhängige staatliche Stelle;
22. „betroffene Aufsichtsbehörde“ eine Aufsichtsbehörde, die von der Verarbeitung personenbezogener Daten betroffen ist, weil
    a) der Verantwortliche oder der Auftragsverarbeiter im Hoheitsgebiet des Mitgliedstaats dieser Aufsichtsbehörde niedergelassen ist,
    b) diese Verarbeitung erhebliche Auswirkungen auf betroffene Personen mit Wohnsitz im Mitgliedstaat dieser Aufsichtsbehörde hat oder haben kann oder
    c) eine Beschwerde bei dieser Aufsichtsbehörde eingereicht wurde;
26. „internationale Organisation“ eine völkerrechtliche Organisation und ihre nachgeordneten Stellen oder jede sonstige Einrichtung, die durch eine zwischen zwei oder mehr Ländern geschlossene Übereinkunft oder auf der Grundlage einer solchen Übereinkunft geschaffen wurde.“

Relevant für diese Arbeit sind im Wesentlichen fünf der oben definierten Begrifflichkeiten. Allen voran ist hier die *betroffene Person* zu nennen, die als „identifizierte oder identifizierbare natürliche Person“ Art. 4 Nr. 1 DSGVO definiert wird, deren (personenbezogene) Daten verarbeitet werden (vgl. Abschnitt 2.1.2).

Der *Verantwortliche* ist im einfachsten Fall die Person, welche die Daten einer betroffenen Person verarbeitet. In der Regel finden Datenverarbeitungen, die von der

DSGVO abgedeckt sind, aber eher im Namen eines Unternehmens statt. Dann ist nicht der verarbeitende Mitarbeiter, sondern das Unternehmen an sich, vertreten durch z.B. den Geschäftsführer, verantwortlich. Der Verantwortliche ist z.B. für die Einhaltung der in Abschnitt 2.1.3 erläuterten Verarbeitungsgrundsätze verantwortlich (siehe Art. 5 (2) DSGVO).

Eine wichtige Ausnahme bildet die sogenannte Auftragsverarbeitung. Ein *Auftragsverarbeiter* ist hierbei eine (natürliche oder juristische) Person, die unmittelbar im Auftrag einer anderen (natürlichen oder juristischen) Person, Daten verarbeitet. Das können beispielsweise IT-Dienstleister sein, wenn die Speicherung von Daten in externe Rechenzentren ausgelagert wird.

Ein *Dritter* ist demnach jeder, der im jeweiligen Kontext nicht Betroffener, Verantwortlicher oder Auftragsverarbeiter ist.

Auftragsverarbeitung, aber auch andere Gründe erfordern regelmäßig eine Weitergabe von personenbezogenen Daten. Wenn Daten weitergegeben werden, wird das Ziel der Datenweitergabe immer als *Empfänger* bezeichnet. Ein Empfänger kann also durchaus gleichzeitig noch eine weitere Rolle, wie beispielsweise Auftragsverarbeiter, einnehmen.

Bei der Weitergabe personenbezogener Daten ist zu beachten, dass betroffene Personen Art. 13 DSGVO immer über die Weitergabe ihrer Daten und die entsprechenden Empfänger zu informieren sind.

Neben diesen Rollen, ist noch eine Weitere im Verlauf der Arbeit relevant: Der *Datenschutzbeauftragte*. Einen solchen müssen alle öffentlichen Stellen und auch viele nicht-öffentliche Stellen, insbesondere Unternehmen, benennen. Bei nicht-öffentlichen Stellen gilt die Maßgabe, dass ein Datenschutzbeauftragter benannt werden muss, wenn mindestens 20 Personen regelmäßig automatisiert personenbezogene Daten verarbeiten, oder die Datenverarbeitung die Kerntätigkeit des Unternehmens darstellt und die betroffenen dabei systematisch Überwacht werden oder besondere Kategorien personenbezogener Daten verwendet werden. Es kann aber auch freiwillig ein Datenschutzbeauftragter benannt werden. Grundsätzlich können interne oder externe Datenschutzbeauftragte benannt werden [De22b].

Wenn ein Datenschutzbeauftragter benannt wird, fungiert dieser als genereller Ansprechpartner in allen Datenschutzbelangen. Er berät und schult die Leitung und andere Beschäftigte seiner Organisation. Außerdem überwacht er die Einhaltung der entsprechenden Vorschriften und ist Ansprechpartner für Betroffene und Aufsichtsbehörden. Zu beachten ist, dass Datenschutzbeauftragte unabhängig in ihrer fachlichen Entscheidung sein sollen und Führungskräften nicht weisungsgebunden sind [De22b].

Die entsprechenden Regelungen zu Datenschutzbeauftragten finden sich in Art. 37–39 DSGVO, §§ 5–7, 38 BDSG (siehe auch Abschnitt 2.3).

2.1.6. Verzeichnis von Verarbeitungstätigkeiten

Unter gewissen Umständen haben Verantwortliche laut Art. 30 DSGVO ein Verzeichnis aller Verarbeitungstätigkeiten in seiner Verantwortung zu führen. Dieses muss für jede Verarbeitungstätigkeit insbesondere die entsprechenden Zwecke, die Kategorien Betroffener Personen und verarbeitetet Daten, aber auch mögliche Empfänger beinhalten. Außerdem sollen die jeweiligen Löschfristen und eine Beschreibung der eingehaltenen technischen und organisatorischen Maßnahmen (siehe Abschnitt 2.1.8) enthalten sein. Ähnliches gilt für Auftragsverarbeiter (siehe Art. 30 Abs. 2 DSGVO).

2.1.7. Datenschutzfolgeabschätzung

Nach Art. 35 DSGVO muss für Datenverarbeitungen, die „voraussichtlich ein hohes Risiko für die Rechte und Freiheiten natürlicher Personen zur Folge“ haben, vorab eine sogenannte Datenschutzfolgeabschätzung durchgeführt werden. Die Schleswig-Holsteinische Datenschutzbeauftragte führt in [Di18] hierfür in Bezugnahme auf [Da17] die folgenden Fälle auf:

- „Bewerten oder Einstufen (Scoring)“
- „Automatisierte Entscheidungsfindung mit Rechtswirkung oder ähnlich bedeutsamer Wirkung“
- „Systematische Überwachung“
- „Vertrauliche oder höchstpersönliche Daten“
- “Datenverarbeitung in großem Umfang“
- „Abgleichen oder Zusammenführen von Datensätzen“
- „Daten zu schutzbedürftigen betroffenen Personen“
- „Innovative Nutzung oder Anwendung neuer technologischer oder organisatorischer Lösungen“
- „Betroffene Personen werden an der Ausübung eines Rechts oder der Nutzung einer Dienstleistung bzw. Durchführung eines Vertrags gehindert“

Die Datenschutzfolgeabschätzung muss neben einer Beschreibung der Datenverarbeitung und deren Zwecken insbesondere eine Bewertung bzgl. Notwendigkeit und Verhältnismäßigkeit in Bezug auf die Risiken für die Betroffenen, sowie geplante Sicherheitsvorkehrungen enthalten (siehe Art. 35 Abs. 7).

2.1.8. Technische und Organisatorische Maßnahmen (TOM)

In Abschnitt 2.1.6 werden bereits technische und organisatorische Maßnahmen (TOM) erwähnt. Konkret werden hierfür in Art. 32 Abs. 1 DSGVO folgende Maßnahmen erwähnt:

„a) die Pseudonymisierung und Verschlüsselung personenbezogener Daten;

b) die Fähigkeit, die Vertraulichkeit, Integrität, Verfügbarkeit und Belastbarkeit der Systeme und Dienste im Zusammenhang mit der Verarbeitung auf Dauer sicherzustellen;

c) die Fähigkeit, die Verfügbarkeit der personenbezogenen Daten und den Zugang zu ihnen bei einem physischen oder technischen Zwischenfall rasch wiederherzustellen; [d)] ein Verfahren zur regelmäßigen Überprüfung, Bewertung und Evaluierung der Wirksamkeit der technischen und organisatorischen Maßnahmen zur Gewährleistung der Sicherheit der Verarbeitung.“

Allgemein wird ebenda aber auch deutlich gemacht, dass auch andere Maßnahmen ergriffen werden können, um die Rechte und Freiheiten betroffener Personen angemessen zu schützen.

2.2. Standard-Datenschutzmodell (SDM)

Die Einhaltung der Datenschutzgesetzgebung wird in Deutschland von den Datenschutzaufsichtsbehörden der Länder und des Bundes kontrolliert. Die jeweiligen Landesdatenschutzbeauftragten und der Bundesdatenschutzbeauftragte bilden zusammen die Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder (DSK) [DS22]. Die DSK erstellt unter anderem Orientierungshilfen für Unternehmen, Behörden und andere Organisationen für den korrekten Umgang mit dem Datenschutz. Ein zentrales Werk stellt das sogenannte Standard-Datenschutzmodell (SDM) dar.

Das SDM soll eine Hilfestellung zur Festlegung von TOM für die Einhaltung der Vorschriften aus der DSGVO bieten. Zum SDM existiert dafür die Hauptveröffentlichung [AK22], in der die verschiedenen Vorschriften der DSGVO zunächst erläutert und dann in TOM überführt werden. Die TOM werden näher in einem Referenzmaßnahmenkatalog beschrieben, wobei die Referenzmaßnahmen thematisch in sogenannte Bausteine zusammengefasst sind, die jeweils eine typische Verarbeitungssituation abbilden. Bisher wurde allerdings nur ein geringer Anteil der Bausteine auch tatsächlich veröffentlicht wurden.

Die Referenzmaßnahmen bauen auf sogenannten Gewährleistungszielen auf, welche die rechtlichen Anforderungen bündeln sollen und sind entsprechend strukturiert. Die definierten Gewährleistungsziele sind:

- Datenminimierung
- Verfügbarkeit
- Integrität
- Vertraulichkeit
- Nichtverkettung
- Transparenz
- Intervenierbarkeit

Für jedes Gewährleistungsziel wird im SDM eine Reihe von Maßnahmen definiert, die dieses Ziel stützen. Darüber hinaus werden die Gewährleistungsziele aber auch als „Design-Strategie“ verstanden. Die DSGVO beschreibt in Art. 25 die Prinzipien *Datenschutz durch Technikgestaltung (Data Protection by Design)* und *datenschutzfreundliche Voreinstellungen (Data Protection by Default)*. Demnach soll grundsätzlich schon bei der Planung einer Datenverarbeitung der Datenschutz eingezogen werden. Hierbei können die oben aufgeführten Gewährleistungsziele Richtschnur darstellen.

2.3. Weitere Gesetze

Die DSGVO gilt als EU-Verordnung zwar unmittelbar in allen Mitgliedsstaaten, beinhaltet aber einige Öffnungs- und Spezifizierungsklauseln, die es den Mitgliedsstaaten erlauben, einzelne Aspekte auf nationaler Ebene zu regeln [De18; Ro18]. In Deutschland ist dies im Wesentlichen im neuen **Bundesdatenschutzgesetz (BDSG-neu)** geschehen. Darüber hinaus existiert jeweils ein **Landesdatenschutzgesetz (LDSG)** für jedes Bundesland, in welchen aber hauptsächlich die Vorgaben für öffentliche Stellen dargestellt werden [In23b]. Außerdem existieren noch einige Regularien, die sich im Kern nicht mit dem Datenschutz befassen, am Rande aber einige Vorschriften diesbezüglich beinhalten. Dies sind etwa die Sozialgesetzbücher, die Abgabenordnung oder das Gesetz über das Bundesamt für Sicherheit in der Informationstechnik, aber auch noch viele weitere [De22b].

Die DSGVO ist allerdings immer vorrangig zu betrachten, da die EU Mitgliedsstaaten keine Regelungen treffen dürfen, die dieser widersprechen und auch keine Regelungen

im nationalen Recht wiederholen dürfen. Das nationale Recht dient lediglich der Ergänzung und Verdeutlichung unklarer Sachverhalte. Unter allen deutschen Regelungen ist zu beachten, dass hier grundsätzlich spezielles Recht vor allgemeinem Recht gilt. Ist ein gewisser Sachverhalt etwa in der Abgabenordnung geregelt, so darf eine Steuerbehörde sich nicht auf eine eventuell anders lautende Vorschrift des BDSG berufen, da die Abgabenordnung hier in aller Regel bindend sein wird [De22b].

Im weiteren Verlauf der Arbeit werden darüber hinaus noch einzelne Inhalte des Vierten Buchs des Sozialgesetzbuchs (SGB IV), sowie des Einkommensteuergesetzes (EStG) verwendet. Diese beziehen sich zwar nicht unmittelbar auf den Datenschutz, werden aber in Abschnitt 9.1 als Beispiele für eine Verarbeitung zur Erfüllung einer rechtlichen Verpflichtung nach Art. 6 Abs. 1 lit. c DSGVO verwendet.

Auf genaue Regelungen der genannten Gesetze wird an dieser Stelle verzichtet, da diese keine entscheidende Rolle für die weiteren Ausführungen spielen.

3. Geschäftsprozessmodellierung

„Geschäftsprozessmanagement (engl.: business process management (BPM)) ist die Kunst und Wissenschaft, die Arbeit in einer Organisation so zu gestalten, dass konsistente Ergebnisse sichergestellt und Verbesserungspotenziale genutzt werden“ [Du21, S. 1]. Damit stellt es einen relevanten Faktor für jedes Unternehmen dar.

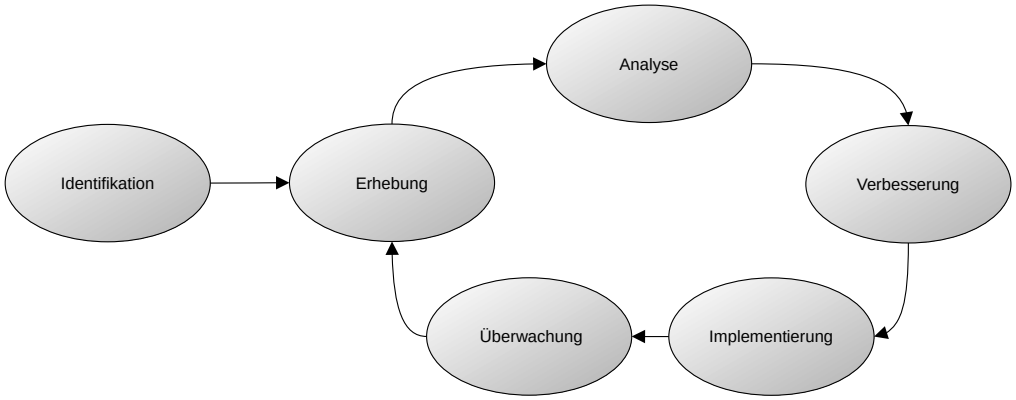


Abbildung 3.1.: BPM-Lebenszyklus (nach [Du21])

Die üblichen Abläufe im BPM werden als „BPM-Lebenszyklus“ bezeichnet (siehe Abbildung 3.1). Zunächst hierfür alle Prozesse, die für ein Geschäftsproblem eines Unternehmens (oder einer sonstigen Organisation) relevant sind, identifiziert werden. Anschließend werden diese, meist in Form eines Prozessmodells, erhoben. In der folgenden Analysephase werden Probleme der Prozesse analysiert und dokumentiert um anschließend in der „Prozessverbesserung“ optimiert zu werden. Die optimierten Prozesse werden dann implementiert, also in die Praxis umgesetzt. Das kann einerseits über organisationsbezogene Änderungen, aber auch durch Prozessautomatisierung geschehen. Alle eingeführten Prozesse müssen abschließend dauerhaft überwacht werden um

neue Fehler zu erkennen. Sollten weitere Probleme gefunden werden, beginnt ein neuer Zyklus [Du21].

Diese Dissertation beschäftigt sich überwiegend mit den Phasen „Prozesserhebung“ und „Prozessanalyse“. Im Folgenden werden daher die zentralen Elemente der Geschäftsprozessmodellierungsnotation BPMN als eine Möglichkeit der Prozesserhebung beschrieben. Dann wird noch kurz die darauf aufbauende PICTURE-Methode, die häufig in der öffentlichen Verwaltung zum Einsatz kommt, erläutert. Die Prozessanalyse wird in Teil II betrachtet.

3.1. BPMN

Als verbreitetste Notation zur Modellierung und Automatisierung von Geschäftsprozessmodellen kann seit einiger Zeit die Business Process Model and Notation (BPMN) der Object Management Group (OMG) angesehen werden [FR19b; PS16]. Die aktuelle Version 2.0 des Standards wird in [Ob13] beschrieben.

Im BPMN-Standard werden mehrere Diagrammtypen unterschieden: Prozessdiagramme, die Abläufe innerhalb einer Organisation beschreiben; Kollaborationsdiagramme, die die Wechselwirkungen mit anderen Organisationen innerhalb eines Prozesses beschreiben; Choreographien, die im Wesentlichen Nachrichtenflüsse zwischen unterschiedlichen Teilnehmern darstellen; sowie Konversationsdiagramme, welche eine logische Relation der Nachrichtenflüsse ohne zeitlichen Ablauf darstellen. In dieser Arbeit werden im wesentlichen Prozess- und Kollaborationsdiagramme betrachtet, die im Folgenden auch nur als Prozessmodelle bezeichnet werden.

BPMN Prozessmodelle bestehen aus Notationselementen fünf verschiedener Kategorien:

- Flussobjekte
- Daten
- Konnektoren
- Teilnehmer
- Artefakte

Die einzelnen Notationselemente werden im Allgemeinen mit schwarzen Linien und weißer bzw. klarer Füllung dargestellt. Der Standard erlaubt aber explizit auch andere Füll- und Linienfarben [Ob13, S. 41, f.].

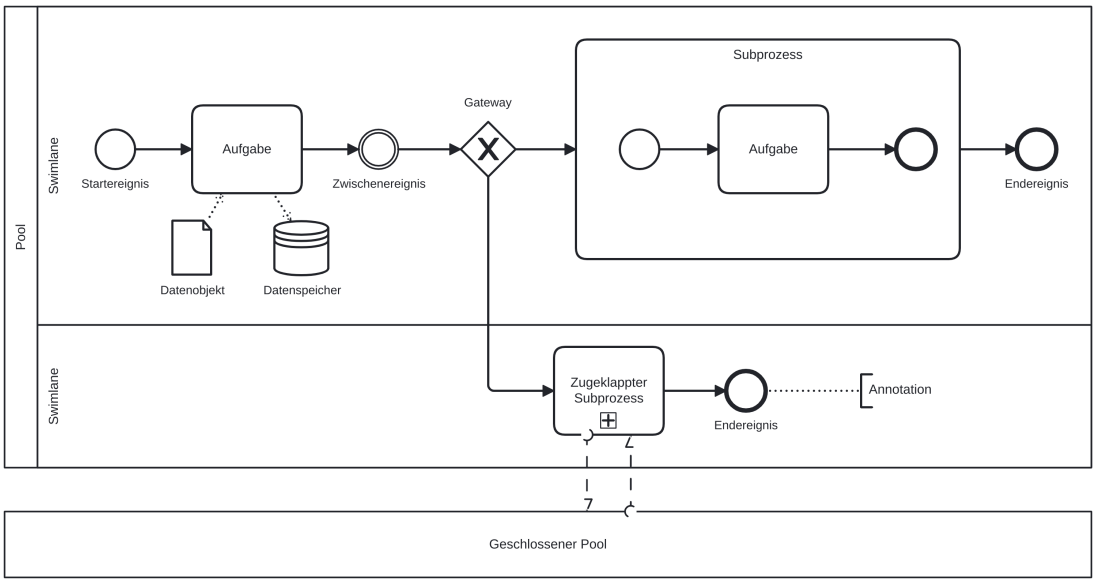


Abbildung 3.2.: Beispiel Prozessmodell zur Veranschaulichung der BPMN Notationselemente

3.1.1. Flussobjekte

Flussobjekte beschreiben das Verhalten eines Prozesses. Es gibt drei Unterkategorien von Flussobjekten, die dann wiederum verschiedene konkrete Notationselemente enthalten:

- Aktivitäten
- Ereignisse
- Gateways

Aktivitäten

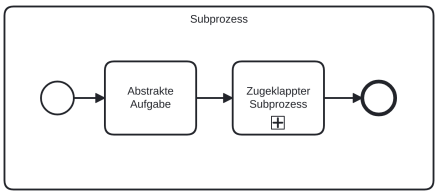


Abbildung 3.3.: Notationselemente für abstrakte Aufgaben und Subprozesse

Aktivitäten bilden die Basis eines BPMN-Prozesses und stellen die zu erledigenden Arbeit dar. Sie werden immer durch ein Rechteck mit abgerundeten Ecken dargestellt. Je nach konkreter Ausprägung kann dieses durch weitere Symbole ergänzt werden.

Aktivitäten haben zwei grundlegende Ausprägungen: Aufgaben (Tasks) und Subprozesse (auch Teilprozesse [FR19a] oder Unterprozesse [GL13] genannt). Die entsprechenden Notationen sind in Abbildung 3.3 gezeigt.

Aufgaben sind hier atomare Tätigkeiten [Ob13], die nach dem Muster „[Objekt] + [Verb]“ ([FR19a, S. 33]) benannt werden sollten und sowohl manuelle als auch automatisierte Tätigkeiten abbilden können.

Subprozesse hingegen sind eigene kleine Prozesse, die für eine bessere Übersicht zusammengefasst werden. Sie können direkt im Prozess abgebildet werden oder nur als zugeklappter Subprozess eingefügt werden, dessen Inhalt an dieser Stelle nicht ersichtlich ist.

Die Grundlage bildet eine sogenannte abstrakte [Ob13] oder unspezifizierte [GL13] Aufgabe, wie oben abgebildet. Hierbei ist formal nicht näher definiert, wie die Aufgabe ausgeführt wird oder worin sie besteht. Jegliche Informationen darüber können nur aus der Bezeichnung entnommen werden. Dieser Aufgabentyp wird häufig in Prozessen verwendet, die nicht automatisiert über eine Workflowengine ausgeführt werden sollen, sondern lediglich der Veranschaulichung dienen und wenn Details über die Art der Aufgabe entweder nicht relevant oder (noch) nicht bekannt sind.

Neben den oben erläuterten abstrakten Aufgaben, sieht der BPMN-Standard auch noch eine Reihe von konkreteren Ausprägungen einer Aufgabe vor.

Abbildung 3.4.: Notationselemente für Nachrichten sendende und empfangende Aufgaben

Für diese Arbeit sind im wesentlichen Nachrichten sendende Aufgaben und Nachrichten empfangende Aufgaben, wie in Abbildung 3.4 gezeigt, relevant. Entsprechend ihrer Bezeichnungen, werden

Sende-Aufgaben (send tasks) verwendet, wenn in einer Aufgabe eine Nachricht an einen anderen Teilnehmer (einen anderen Pool) versendet werden soll.

Empfangs-Aufgaben (receive tasks) hingegen werden erst dann ausgeführt, wenn eine entsprechende Nachricht von einem anderen Teilnehmer ankommt und verarbeiten diese dann weiter.

Da es durchaus naheliegend ist, dass die Nachrichten personenbezogene Daten enthalten, werden diese Arten von Aufgaben in Kapitel 8 und Kapitel 12 gesondert betrachtet.

Darüber hinaus sind im BPMN-Standard auch noch weitere Arten von Aufgaben vorgesehen. Diese sind inhaltlich für die Arbeit zwar nicht sonderlich relevant, werden aber an einzelnen Stellen zur Veranschaulichung verwendet und daher an dieser Stelle auch kurz eingeführt.

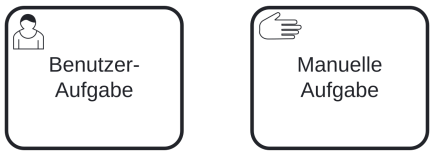


Abbildung 3.5.: Notationselemente für manuelle und Benutzer-Aufgaben

So existieren etwa zwei Aufgabentypen, die explizit eine menschliche Bearbeitung erfordern (siehe Abbildung 3.5).

Benutzer-Aufgaben (user tasks) werden dann verwendet, wenn die Aufgabe mit Hilfe einer Software, insbesondere einer Workflowengine, bearbeitet werden soll. Der Nutzer gibt hier beispielsweise Daten in ein System ein und bestätigt dies anschließend.

Manuelle Aufgaben (manual tasks) hingegen werden komplett unabhängig von Software bearbeitet. Ein Beispiel hierfür wäre das Vernichten von Papierakten.

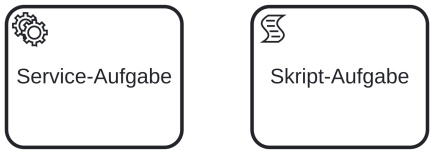


Abbildung 3.6.: Notationselemente für Service- und Skript-Aufgaben

Außerdem sind zwei eher technische Aufgabentypen vorgesehen, die in Abbildung 3.6 dargestellt werden.

Service-Aufgaben werden durch eine externe Software, beispielsweise einen Webservice, bearbeitet.

Skript-Aufgaben hingegen starten direkt in der Workflowengine ein zuvor definiertes Skript, welches in der Regel eine eher simple Aufgabe wie einfache Berechnungen oder Ähnliches ausführt.

Abbildung 3.7.: Notationselement für Geschäftsregel-Aufgaben

Geschäftsregel-Aufgaben (business rule tasks, siehe Abbildung 3.7) sind relativ neu im BPMN-Standard. Hier können schon bei der Modellierung einfache „wenn-dann"-Regeln definiert werden, welche dann zur Laufzeit des Prozesses automatisch geprüft werden. Beispielsweise könnte in einem Versandhandel eine Regel definiert werden, welche Versandkosten anhand des Warenwerts und oder -gewichts bestimmt.

Abbildung 3.8.: Notationselement für Aufruf-Aufgaben

Aufrufaktivitäten (call activities) unterscheiden sich von anderen Aktivitäten dadurch, dass sie global definiert sind und aus beliebigen Prozessen aufgerufen werden können. Sie werden durch einen dickeren Rand gekennzeichnet und können sowohl die Form einer einfachen Aufgabe, als auch eines Subprozesses (siehe Abbildung 3.8) haben.

[FR19a; GL13; Ob13]

Ereignisse

Abbildung 3.9.: Notationselemente für die verschiedenen Kategorien von Ereignissen

Ereignisse stellen alles dar, was im Laufe des Prozesses „Betrachtenswertes [sic!] passiert" ([FR19a, S. 33]).

Grundsätzlich kann zwischen drei Arten von Ereignissen unterschieden werden, die in Abbildung 3.9 dargestellt werden.

Startereignisse stehen am Anfang eines jeden Prozesses und sind dessen Auslöser. Sollte der Prozess aus mehreren Pools bestehen, so muss auch in jedem Pool ein Startereignis vorhanden sein. Ein Startereignis hat nie einen eingehenden Kontrollfluss.

Zwischenereignisse können an beliebigen Stellen und in beliebiger Anzahl im Prozess modelliert werden (außer am Anfang und am Ende).

Endereignisse beenden einen Prozess. Daher muss jeder Prozess mindestens ein Endereignis haben. Es können allerdings auch mehrere Endereignisse auftreten, wenn der Prozess in mehre Pfade aufgeteilt wird. Allerdings kann ein Endereignis nie einen ausgehenden Kontrollfluss haben.

Außerdem gibt es analog zu den Aufgaben unterschiedliche Typen von Ereignissen, die grundlegend beschreiben, welcher Art das Ereignis ist. Beispielsweise existieren spezielle Ereignistypen für Nachrichten und zeitgesteuerte Ereignisse.

Relevant ist hierbei eine Unterscheidung zwischen eingetretenen und ausgelösten Ereignissen [FR19a].

Eingetretene Ereignisse (catching events) werden durch einen externen Umstand ausgelöst und bewirken etwas im Prozess. Beispielsweise sind Startereignisse immer eingetretene Ereignisse.

Ausgelöste Ereignisse hingegen werden vom Prozess selbst bzw. dem ausführenden Teilnehmer des Prozesses aktiv ausgelöst. Dies kann auf Zwischenereignisse und auf Endereignisse zutreffen.

Abbildung 3.10.: Notationselemente für eintretende und ausgelöste Nachrichten-Ereignisse

In Abbildung 3.10 sind beispielsweise zwei Nachrichten-Ereignisse zu sehen. Der nicht ausgefüllte Briefumschlag repräsentiert eine eingehende Nachricht, also ein eingetretenes Ereignis. Der ausgefüllte Briefumschlag hingegen repräsentiert eine Nachricht, die vom Prozess versendet wird, also ein ausgelöstes Ereignis.

Darüber hinaus existieren noch einige weitere Ereignistypen, die in dieser Arbeit allerdings nicht verwendet werden und auch keine Relevanz im Datenschutzkontext haben und daher hier nicht näher erläutert werden.

Gateways

Die dritte Art von Flussobjekten neben Aktivitäten und Ereignissen, stellen Gateways dar. Sie werden verwendet, um den Kontrollfluss des Prozesses zu verzweigen und wieder zusammenzuführen. Hierfür sieht der BPMN-Standard fünf verschiedene Arten von Gateways vor, die Abbildung 3.11 entnommen werden können.

Abbildung 3.11.: Notationselemente für die verschiedenen Arten von Gateways

Exklusive Gateways (exclusive) werden im Falle einer Verzweigung verwendet, wenn nur genau einer der ausgehenden Pfade ausgewählt werden kann. Es handelt sich also um ein ausschließendes Oder (XOR). Teilweise wird das exklusive Gateway, anders als in Abbildung 3.11, als einfache Raute ohne weiters Symbol dargestellt [GL13].

Inklusive Gateways (inclusive) lassen es zu, dass beliebig viele ausgehende Kontrollflüsse (mindestens aber einer) der Verzweigung gewählt werden können. Sie repräsentieren also ein logisches Oder (OR) [GL13].

Parallele Gateways repräsentieren eine Und-Verknüpfung. Nach einer solchen Verknüpfung werden alle ausgehenden Kontrollflüsse bearbeitet. Bei einer entsprechenden Zusammenführung, läuft der Prozess nur dann weiter, wenn alle eingehenden Kontrollflüsse fertig bearbeitet wurden [GL13].

Komplexe Gateways werden überwiegend in Form einer Zusammenführung und nur sehr selten als Verzweigung verwendet [FR19a]. Die komplexe Zusammenführung wird in der Regel mit einer Regel versehen, die näheres darüber aussagt, wann der Prozess weitergeführt werden soll. So kann beispielsweise eine komplexe Zusammenführung in einem Prozess hinter einer parallelen Verzweigung mit fünf ausgehenden Kontrollflüssen liegen. Die Zusammenführung kann dann so definiert sein, dass der Prozess schon dann weiter läuft, wenn z.B. drei der fünf eingehenden Kontrollflüsse abgearbeitet wurden[FR19a; GL13].

Ereignisbasierte Gateways werden nur als Verzweigungen verwendet und verhalten sich im Wesentlichen wie exklusive Verzweigungen. Allerdings werden hier keine Bedingungen angeben, die bestimmen, welcher Kontrollfluss ausgewählt wird. Stattdessen folgt auf das Gateway auf jedem Pfad ein unterschiedliches Ereignis. Sobald ein Ereignis eintritt, wird der entsprechende Pfad gewählt. Rein theoretisch könnte ein Ereignisbasiertes Gateway auch als inklusive Verzweigung fungieren, falls zwei oder mehr Ereignisse exakt gleichzeitig eintreten [GL13].

3.1.2. Daten

Die Kategorie „Daten“ fasst *Datenobjekte* und *Datenspeicher* zusammen. Insgesamt beinhaltet diese Kategorie sieben Notationselemente, die in Abbildung 3.12 dargestellt sind.

Abbildung 3.12.: Notationselemente für Daten

Datenobjekte beinhalten einzelne oder auch mehrere Informationen, die im Prozessfluss von den Aktivitäten erstellt oder verwendet werden.

Datenobjekte, welche in einem Prozess erstellt und in einem anderen Prozess verwendet werden, werden im erstellenden Prozess mit einem ausgefüllten Blockpfeil versehen und als *Datenoutput* bezeichnet. Im verwendenden Prozess werden die Datenobjekte mit einem nicht ausgefüllten Blockpfeil versehen und *Dateninput* genannt. Datenobjekte können generell sowohl in digitaler als auch in Papierform vorliegen [GL13].

Datenspeicher repräsentieren im Modell beispielsweise eine Datenbank oder einen Aktenschrank, in die Daten abgelegt oder aus welchen Daten abgerufen werden.

Ob eine Aktivität lesend oder schreibend auf ein Datenobjekt bzw. Datenspeicher zugreift, ist an den entsprechenden Datenflüssen (siehe 3.1.3) abzulesen.

Nachrichten sind im Grunde genommen Datenobjekte, die zwischen zwei Teilnehmern eines Prozesses ausgetauscht werden.

Unter dem Piktogramm wird jeweils die Bezeichnung des Elements angegeben. Hierbei ist es auch möglich, den Status eines Dokuments anzugeben. Hierfür wird der Status in eckigen Klammern an die Bezeichnung angehängt. Beispielsweise könnte aus „Formular [leer]“ im Verlauf eines Prozesses „Formular [ausgefüllt]“ werden. Selbiges gilt analog für Datenspeicher.

3.1.3. Konnektoren

Flussobjekte und Daten müssen natürlich mit einander Verknüpft werden. Hierfür kennt die BPMN drei verschiedene Arten von Konnektoren. Diese sind in Abbildung 3.13 dargestellt. Sequenzflüsse verbinden alle Flussobjekte miteinander. Datenassoziationen

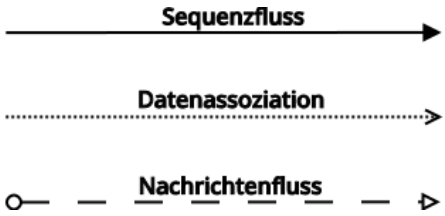


Abbildung 3.13.: Arten von Konnektoren

verbinden jeweils ein Datenobjekt oder Datenspeicher mit einem Flussobjekt, wie etwa einer Aufgabe oder einem Ereignis. Nachrichtenflüsse werden immer dann verwendet, wenn Nachrichten zwischen zwei Teilnehmern versendet werden. Hierfür reicht ein Nachrichtenfluss zwischen zwei entsprechenden Aufgaben oder Ereignissen. Zur besseren Veranschaulichung kann zusätzlich noch ein Nachrichtenobjekt hinzugefügt werden.

3.1.4. Teilnehmer

Ein wichtiger Aspekt bei der Prozessmodellierung ist die Darstellung der teilnehmenden Personen bzw. Einheiten. Hierfür nutzt BPMN das Konzept der Pools und Lanes.

Ein Pool kann dabei je nach Kontext des Prozesses beispielsweise ein Unternehmen oder auch eine einzelne natürliche Person, wie etwa einen Kunden, darstellen. Er wird

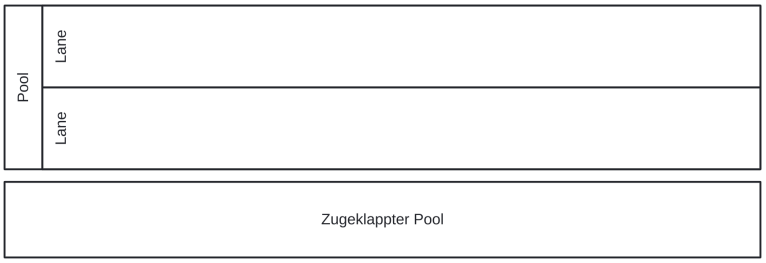


Abbildung 3.14.: Notationselemente für die Darstellung der Prozessteilnehmer

in der Regel durch ein Rechteck dargestellt, in welchem der eigentliche Prozessfluss abgebildet wird.

Ein Prozess kann natürlich mehrere Teilnehmer, also auch mehrere Pools haben, zwischen welchen Nachrichten ausgetauscht werden können. Es ist hierbei möglich, Teilnehmer als zugeklappten Pool, also Black-Box, darzustellen. In einem zugeklappten Pool befinden sich keine weiteren Modellelemente. Es können aber weiterhin Nachrichten ausgetauscht werden. Diese Darstellung macht immer dann Sinn, wenn die konkreten Abläufe eines Teilnehmers unbekannt oder schlicht irrelevant für den Prozess sind.

Pools können beliebig unterteilt werden in sogenannte Lanes (teilweise auch Schwimmbahnen genannt). Dies bietet sich beispielsweise für die Abbildung verschiedener Abteilungen eines Unternehmens an. Lanes wiederum können weiter in Sub-Lanes unterteilt werden. Kontrollflüsse können auch zwischen mehreren Lanes eines Pools verlaufen.

Wichtig bei der Unterscheidung zwischen Pools und Lanes ist der Umstand, dass Pools immer selbstständige Teilnehmer abbilden und Lanes untergeordnete Einheiten. Kunde und Lieferant sollten also beispielsweise in aller Regel nicht durch zwei Lanes des gleichen Pools abgebildet werden [GL13].

3.1.5. XML-Repräsentation

Für die Speicherung und Weitergabe von BPMN-Diagrammen soll nach dem Standard ein spezielles Format der Extensible Markup Language (XML) genutzt werden, welches in mehreren XML Schemata[1] definiert wird:

DC.xsd definiert grafische Eigenschaften(*Font, Point, Bounds*)

DI.xsd definiert allgemeine Elemente wie *Diagram*, *Node*, oder *Edge*

BPMNDI.xsd kombiniert die bereits genannten Schemata miteinander zu grafischen Elementen

[1] Grundlagen zu XML Schema finden sich unter https://www.w3.org/TR/xmlschema-0/

Semantic.xsd definiert die tatsächlichen Modellelemente wie etwa *Activity* und *DataObject* und stellt semantische Zusammenhänge her

BPMN20.xsd ist ein abstraktes Oberschema

Alle Schemata können auf der Website der OMG[2] heruntergeladen werden.

3.1.6. Erweiterung

Der BPMN-Standard erlaubt explizit die Erweiterung der Notation mit zusätzlichen Elementen, sowie das Erweitern der Standardelemente mit zusätzlichen Attributen. Hierbei darf allerdings die Semantik der einzelnen Elemente nicht verändert werden und das „basic look-and-feel“ muss erhalten bleiben [Ob13, S. 42].

Hierfür existieren zwei Methoden, deren Grundlagen im Folgenden beschrieben werden:

- die Erweiterung des Metamodells mit der *Meta Object Facility (MOF)*, sowie
- die Erweiterung des BPMN XML Schemas.

Von diesen Erweiterungsmechanismen wird in verschiedensten Domänen Gebrauch gemacht [Za19]. Einige beispielhafte Erweiterungen werden auch in Kapitel 6 diskutiert. Weitere Details zum Erweiterungsmechanismus finden sich im BPMN-Standard [Ob13], sowie in [SCV11] und [Za19].

MOF Erweiterung

Abbildung 3.15 zeigt das einen Ausschnitt des BPMN-Metamodells, der die Standard-Modellelemente darstellt. Das Metamodell wurde mit der von der OMG bereitgestellten MOF entwickelt und kann damit auch erweitert werden [SCV11].

Abgeleitet werden alle Erweiterungen von einem `BaseElement`. Dies können im Wesentlichen die Klassen des Metamodells sein. Eine `Extension` hat dabei ein Attribut, welches angibt, ob die Erweiterung verstanden werden muss, um das Modell korrekt zu verarbeiten. Außerdem hat jede Extension eine `ExtensionDefinition`, welche einen Namen vergibt und außerdem eine Menge von Attributen (`ExtensionAttributeDefinition`) definiert, welche wiederum eine Menge von Werten haben [Za19].

Das Klassendiagramm für den MOF-Erweiterungsmechanismus ist in Abbildung 3.16 dargestellt.

[2] https://www.omg.org/spec/BPMN/2.0.2/

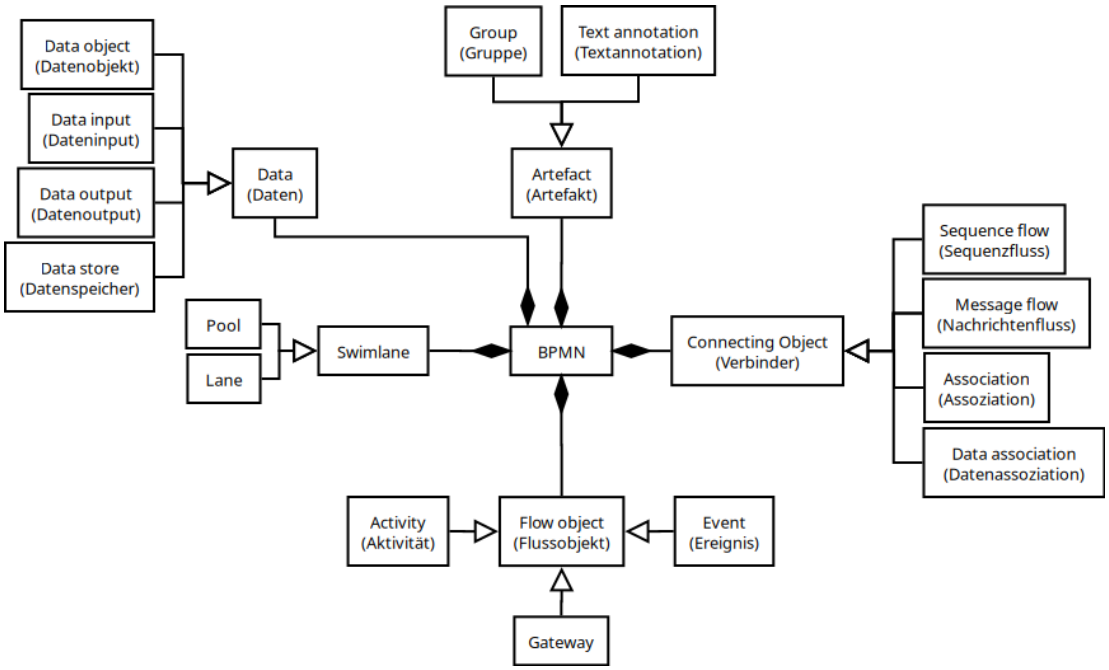


Abbildung 3.15.: BPMN-Metamodell (nach [Za19])

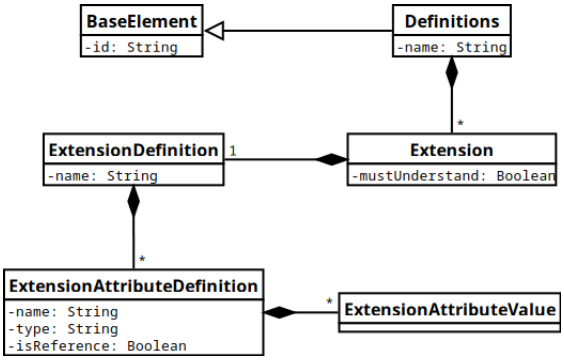


Abbildung 3.16.: Klassendiagramm für die BPMN-Erweiterung mit MOF (nach [SCV11; Za19])

XML Erweiterung

Außerdem kann eine Erweiterung auch über ein XML Schema definiert werden. Ein Klassendiagramm, die diesen Erweiterungsmechanismus beschreibt, findet sich in Abbildung 3.17. Die XML-Repräsentation des Erweiterungsschemas nach [Ob13, S. 58] findet sich in Listing 3.1.

Für die Erweiterung über XML Schema, muss für eine separate Schema-Datei erstellt werden [Ob13; SCV11]. Diese muss vor den eigentlichen Definitionen der Erweiterung

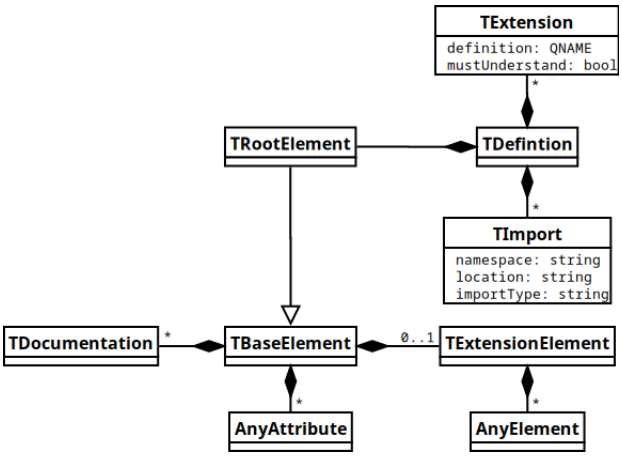


Abbildung 3.17.: Klassendiagramm für die BPMN-Erweiterung mit XML (nach [SCV11])

```
1 <xsd:element name="extension" type="tExtension"/>
2 <xsd:complexType name="tExtension">
3     <xsd:sequence>
4         <xsd:element ref="documentation" minOccurs="0" maxOccurs=
              "unbounded"/>
5     </xsd:sequence>
6     <xsd:attribute name="definition" type="xsd:QName"/>
7     <xsd:attribute name="mustUnderstand" type="xsd:boolean" use="
          optional"/>
8 </xsd:complexType>
```

Listing 3.1: XML-Schema für Erweiterungen

ein Element `<xsd:group>` enthalten, in welchem alle Erweiterungen zusammengefasst werden.

3.2. Picture

Die *PICTURE-Methode* ist ein Verfahren zum Prozessmanagement in der öffentlichen Verwaltung. Sie bietet ursprünglich neben einem Vorgehensmodell, eine eigene Modellierungssprache und ein entsprechendes Werkzeug. Die Modellierungssprache basiert dabei auf einer Reihe von Prozessbausteinen, die jeweils einen üblichen Verwaltungsvorgang, wie etwa *Inhaltliche Prüfung vornehmen* oder *Rückfrage durchführen* abbilden. Die Prozessbausteine sind in die fünf Kategorien *Inhaltliche Verwaltungsarbeit*, *Verschriftlichung und Dokumentation*, *Informationsbeschaffung und Koordination*, *Informationsflüsse und Beteiligungen* und *Medienwechsel* gruppiert. Die Prozessbausteine können mit zusätzlichen Attributen näher spezifiziert werden [Be07b]. Auf das

Vorgehensmodell soll an dieser Stelle nicht näher eingegangen werden. Details können in [Be07b] nachgelesen werden.

In neueren Versionen des Modellierungswerkzeugs (*PICTURE Prozessplattform*) können Prozesse auch mit BPMN oder dem sogenannten *PICTURE-BPMN* modelliert werden. Picture-BPMN ist im Wesentlichen eine Kombination aus BPMN und der klassischen Picture-Notation. Die Prozessmodelle entsprechen im Wesentlichen ganz normalen BPMN-Prozessen, es können aber spezielle Typen von Aufgaben verwendet werden, die den Picture-Prozessbausteinen entsprechen.

4. Design von Notationen

In der konzeptionellen Modellierung wird häufig angenommen, dass die grafischen Details der Notation vergleichsweise unwichtig sind, solange die Notation klar definiert ist. Viele Designentscheidungen werden daher auch nicht näher begründet. Hitchman beklagt diesen Umstand in [Hi02] und argumentiert, dass die Notation sehr wohl enorm wichtig ist. Auch in [Mo10] wird dargelegt, dass grafische Notationen im Bereich der Informationssystementwicklung häufig auf Basis des persönlichen Geschmacks designt werden, obwohl es durchaus wissenschaftliche Erkenntnisse gibt, die betrachtet werden sollten.

Jef Raskin, Benutzerschnittstellenentwickler, unter anderem für den Macintosh, befasst sich in [Ra94] mit dem Begriff der Intuition im Kontext der „intuitiven“ Bedienung von Computersystemen. Er kommt zu dem Schluss, dass intuitiv hier im Wesentlichen mit vertraut gleichgesetzt werden kann („In short, 'intuitive' in this context is an almost exact synonym of 'familiar'“ [Ra94, S. 2]). Er verwendet unter anderem das Beispiel einer Computermaus, deren Nutzung heute wohl die meisten Menschen als völlig intuitiv bezeichnen würden. In der Anfangszeit der PC-Ära war dies aber keineswegs der Fall. Viele Erstanwender kamen gar nicht allein auf die Idee der korrekten Nutzung. Nur durch den inzwischen enorm hohen Bekanntheitsgrad und die damit einhergehende Vertrautheit, stellt sich das Gefühl der intuitiven Nutzung ein. Selbiges gilt auch für das allgemeine Design von Benutzerschnittstellen. Hier ist es auch so, dass ein bekanntes Design in einem neuen Kontext, also beispielsweise einer neuen Software, ebenfalls diese Vertrautheit auslöst. Zwar bewertet Raskin dies eher kritisch und sieht die Anforderung intuitive Benutzerschnittstellen zu entwickeln als innovationsfeindlich, aber sie bietet dennoch einige Vorteile. So geht mit einer intuitiven – also vertrauten – Benutzerschnittstelle meist eine geringe Einarbeitungszeit und somit eine hohe anfängliche Produktivität einher. Dies ist gerade für Sachverhalte, in denen der Nutzer sich gar nicht wirklich mit der Materie beschäftigen möchte ein entscheidender Vorteil.

Um zu einem intuitiven Design zu gelangen, muss zunächst überlegt werden, welche Aspekte hierfür relevant sind. Bertin definiert in [Be83], was als Standardwerk des Grafikdesigns gilt, sechs visuelle Variablen, die bei der Entwicklung einer grafischen Notation betrachtet werden müssen: Die Form, die Größe, die Orientierung, die Farbe, die Helligkeit und die Textur[Mo10]. Im Folgenden werden nur die Farbe und Form

näher betrachtet, da die anderen Variablen letztlich als Teilaspekte hiervon verstanden werden können, welche Form und Farbe nur näher beschreiben.

4.1. Farben

Ein wichtiger Aspekt im Design grafischer Notationen ist die Farbe. Ein Grund hierfür ist, dass der Mensch Unterschiede zwischen Farben deutlich schneller wahrnehmen kann, als beispielsweise zwischen Formen [Mo10].

Dass Farbe einen psychologischen Einfluss auf den Menschen (und auch auf Tiere) hat, ist seit Jahrzehnten gut untersucht. So beschreibt Goldstein z.B. schon 1942 von einem neurologischen Standpunkt die Wirkung verschiedener Farben auf seine Patienten und die Probanden mehrerer Studien. Insbesondere wird in der Arbeit der Unterschied in der Wirkung der Farben Rot und Gelb auf der einen Seite und Grün und Blau auf der anderen Seite erörtert. Rot bringt den Menschen Goldsteins Ansicht nach aus dem Gleichgewicht, während Grün dieses Gleichgewicht erhält oder sogar wieder herstellt. Des weiteren schreibt Goldstein, Rot „quäle den Organismus, rege ihn auf und erlaube ihm nicht, ruhig zu sein“. Außerdem sei rot unter anderem „sehr unangenehm“ und „aggressiv“ (frei übersetzt nach [Go42]). Gelb hat laut der Arbeit eine ähnliche, aber schwächere Wirkung. Grün hingegen wird als angenehm, weich und beruhigend beschrieben. Hier wird Blau als das entsprechend schwächere Pendant genannt[Go42].

Auch in [JH74] werden Untersuchungen bzgl. dieser Farben geschildert. Hier wird ein Erregungslevel bei der Betrachtung der Farben gemessen. Rot hat hier mit Abstand den höchsten Wert. Aber auch jüngere Studien legen ähnliche Zusammenhänge nah. In [El07; EM07] wird festgestellt, dass eine rote Umgebung im Vergleich zu einer grünen zwar eine Leistungssteigerung bei Test zu Folge hat, gleichzeitig von den Probanden aber eher vermieden wird.

Aber auch in der Praxis haben die Farben, insbesondere Rot, Gelb und Grün eine hohe Relevanz, da sie auf Grund ihrer weltweiten Verbreitung als Ampelfarben allgemein bekannt und daher intuitiv verständlich sind [Mc99]. Weitere Anwendungen der beschrieben Farben finden sich in Abschnitt 6.3.

Die Visualisierung durch Farben hat den Nachteil, dass vergleichsweise wenig unterschiedliche Werte sinnvoll unterschieden werden können. Zwar gibt es theoretisch sehr viele Farbschattierungen, diese sind für den ungeübten Betrachter, gerade unabhängig von einander, allerdings recht schwer zu unterscheiden. [Mo10] gibt beispielsweise eine „Kapazität“ der Dimension Farbe von sieben bis zehn verschiedenen Werten an. Das deckt sich mit den von Goethe in [Go10] genannten acht verschiedene Farben (Gelb, Rotgelb, Gelbrot, Blau, Rotblau, Blaurot, Rot (Purpur) und Grün).

4.2. Formen

Die einzige Dimension, die nach [Mo10] keine limitierte Kapazität hat, ist die Form. Somit können nur mit Hilfe von verschiedenen Formen theoretisch unbegrenzt viele Sachverhalte ausgedrückt werden.

Unterschieden werden kann hier zwischen einfachen geometrischen Formen, wie etwa Kreisen und Rechtecken auf der einen Seite, und komplexeren Piktogrammen oder Icons auf der anderen Seite. Beide Varianten haben Vor- und Nachteile. Piktogramme können tendenziell als intuitiver betrachtet werden, da hier – im besten Falle – mit allgemein bekannten Symbolen gearbeitet wird, deren Bedeutung für die meisten Menschen die gleiche sein sollte. Die meisten Menschen werden etwa bei einem Vorhängeschloss relativ schnell verstehen, dass etwas gesperrt ist (oder sein sollte). Eine einfache geometrische Form hingegen wird diese Reaktion wohl eher nicht hervorrufen. Ausnahmen gibt es allerdings auch hier, wobei die Interpretation auch abhängig vom Kontext ist. Im Bereich des Straßenverkehrs beispielsweise werden einige Formen schnell mit Verkehrszeichen assoziiert. Außerdem sind geometrische Formen aber in der Regel leichter zu Zeichnen und auch auf den ersten Blick zu erkennen – gerade bei kleinen Abbildungen.

4.3. Text

Neben der grafischen Darstellung spielt aber auch Text eine wichtige Rolle im Bereich Notationsdesign. Stellt man sich beispielsweise die komplette Menüführung einer modernen Software komplett grafisch vor, wirkt dies doch sehr ungewohnt und verwirrend, also unintuitiv.

Zwar sollten grafische Notationen für eine maximale Ausdrucksstärke grundsätzlich vorgezogen werden [Mo10], doch Text kann durchaus eine sinnvolle Ergänzung darstellen. Nach der „dual coding theory“ [Pa86] ist die Kombination von Text und grafischen Elementen besser geeignet um Informationen darzustellen, als beide Varianten für sich allein. Die textuellen Elemente können hierbei gut zur weiteren Verdeutlichung der Bedeutung grafischer Elemente verwendet werden. Generell gibt es hier zwei verschiedene Ansätze. Einerseits können Annotationen, also Kommentare direkt im Modell verwendet werden. Andererseits können aber auch die verwendeten Symbole selbst mit Text versehen sein [Mo10]. In der Prozessmodellierung finden sich beide Varianten. BPMN hat beispielsweise ein eigenes Notationselement für Annotationen und auch die meisten anderen Elemente (z.B. Aufgaben) werden zur Verdeutlichung mit Text versehen. Insgesamt gilt es, so viel Text wie nötig, aber auch so wenig wie möglich zu verwenden, um alle nötigen Informationen unterzubringen aber gleichzeitig die Darstellung nicht zu überfrachten.

4.4. Barrierefreiheit

Interessant ist in diesem Kontext auch die Barrierefreiheit, die beim Design einer Notation berücksichtigt werden sollte. Grafische Darstellungen haben gegenüber textuellen Beschreiben hier generell den Vorteil, dass sie für Menschen mit einer Lese-Schwäche einfacher verständlich sind. Andererseits benötigen Menschen mit einer Sehbehinderung teilweise eine Unterstützung durch Sprachausgabe. Hier sind Texte tendenziell einfacher zu verarbeiten. Bei der Verwendung von Farbe muss außerdem eine mögliche Farbenfehlsichtigkeit beachtet werden. Hier können je nach Ausprägung nur einzelne Farben oder – im Extremfall – gar keine Farben wahrgenommen werden. Hilfreich sind für die Barrierefreiheit generell redundante Notationen, in denen ein Zusammenhang auf verschiedene Arten, also etwa durch Farbe und Text vermittelt wird.

5. Künstliche Intelligenz

Nach der Definition des Gabler Wirtschaftslexikon beschäftigt sich Künstliche Intelligenz (KI), „mit Methoden, die es einem Computer ermöglichen, solche Aufgaben zu lösen, die, wenn sie vom Menschen gelöst werden, Intelligenz erfordern“ [LS18].

Das Forschungsfeld kann dabei in zwei Dimensionen betrachtet werden: Den Methoden und den potentiellen Anwendungsgebieten.

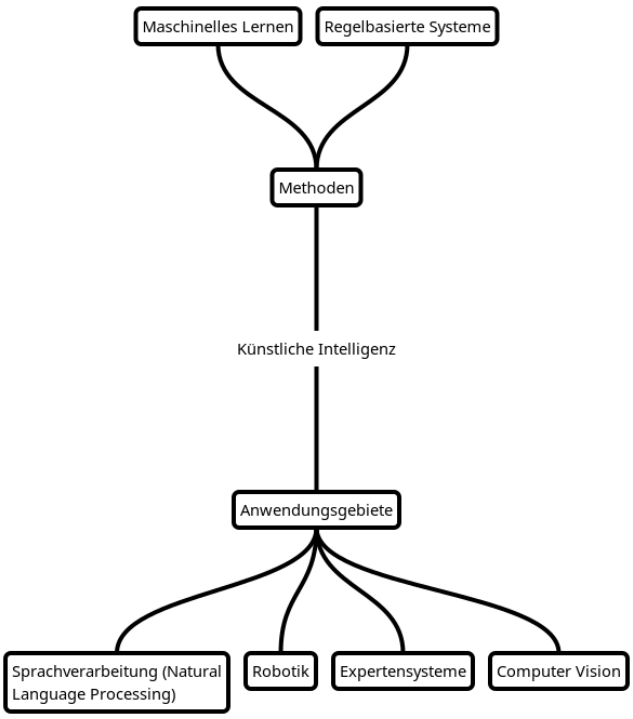


Abbildung 5.1.: Dimensionen Künstlicher Intelligenz (in Anlehnung an [LS18; Ma21; We20])

Abbildung 5.1 bietet eine Übersicht über einige Beispiele beider Dimensionen, wobei sowohl die Einteilung, als auch die einzelnen Definitionen nicht unbedingt als eindeutig angesehen werden können [We20]. In dieser Arbeit werden im Wesentlichen Methoden aus dem Bereich *Maschinelles Lernen* im Anwendungsgebiet *Natural Language Processing* verwendet, weshalb auch nur diese beiden Themen im Folgenden näher erläutert werden.

5.1. Maschinelles Lernen

Maschinelles Lernen (ML)[1] versucht anhand der automatisierten Ableitung von Regeln aus großen Mengen von Trainingsdaten, Problemlösungsstrategien zu entwickeln und zu optimieren [Ma21; PH20]. Wichtig dabei ist, dass die Algorithmen hierfür nicht spezifisch für einen Anwendungsfall sind, da kein Vorwissen über die verwendeten Daten vorausgesetzt wird [Ma21]. Grundlage hierfür bieten Statistische Methoden. [GBC16] beschreibt den Zusammenhang folgendermaßen:

> „Machine learning is essentially a form of applied statistics with increased emphasis on the use of computers to statistically estimate complicated functions and a decreased emphasis on proving confidence intervals around these functions."

5.1.1. Lernmodelle

Grundsätzlich gibt es hierfür drei verschiedene Lernmodelle, die in Abbildung 5.2 dargestellt und im Folgenden näher erläutert werden: Das *überwachte Lernen*, das *unüberwachte Lernen* und das *bestärkende Lernen*.

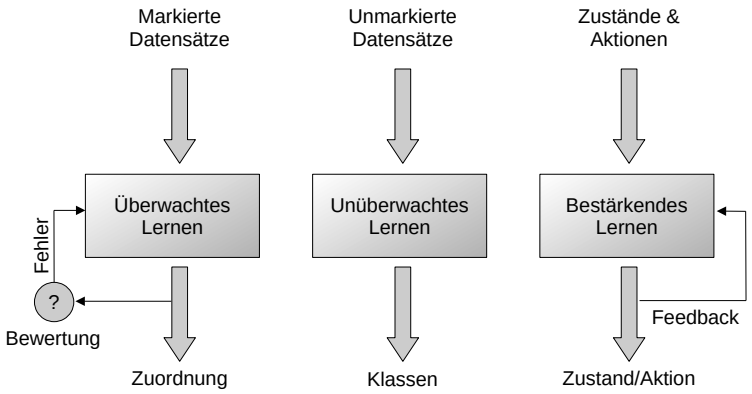


Abbildung 5.2.: Verschiedene Arten des Maschinellen Lernens (nach [We20, S. 39])

Überwachtes Lernen

Beim überwachten Lernen (Supervised Learning) erhält der Algorithmus eine Rückmeldung, die besagt, ob seine Vorhersage korrekt ist. Hierfür wird der Trainingsdatensatz

[1] Auch im Deutschen wird häufig die englische Bezeichnung „Maschine Learning" verwendet

mit den entsprechend korrekten Ausgaben angereichert. Aus der Abweichung zwischen Vorhersage und vorgegebenem Wert lernt der Algorithmus [We20]. Das überwachte Lernen lässt sich generell gut für zwei Kategorien von Anwendungsfällen verwenden, wobei über die Art der erwünschten Ausgabe abgegrenzt wird [Ma21; We20]:

Klassifikation wird für diskrete Ausgabewerte verwendet, also eine vergleichsweise geringe, abzählbare Menge möglicher Klassen.

Regression hingegen wird für kontinuierliche Ausgabewerte verwendet, also z.B. Gleitkommazahlen.

Unüberwachtes Lernen

Beim unüberwachten Lernen (Unsupervised Learning) enthält der Trainingsdatensatz – im Gegensatz zum überwachten Lernen – keine gewünschten Ausgaben. Statt über die Abweichung von Vorhersage und vorgegebenem Wert, arbeitet der Algorithmus hier durch Segmentierung des Datensatzes. Dabei werden Elemente des Datensatzes, die gleiche Merkmale aufweisen, in gemeinsame Klassen aufgeteilt.

Bestärkendes Lernen

Bestärkendes Lernen (Reinforcement Learning), auch Verstärkungslernen genannt, erfordert – ähnlich wie das überwachte Lernen – eine Rückmeldung. Diese ist allerdings nicht für jedes einzelne Element des Trainingsdatensatzes notwendig. Stattdessen erhält der Algorithmus nach einer Reihe von Aktionen eine „Belohnung“, wenn diese zu einem bestimmten Ziel geführt haben.

Das bestärkende Lernen bietet sich daher bei Problemen an, für die es wenig Trainingsdaten gibt, oder wenn das gewünschte Verhalten (noch) nicht eindeutig festgelegt werden kann [We20].

In der Regel wird dieses Lernmodell auf Zustandsbasierte Probleme, wie beispielsweise Brettspiele oder Robotersteuerungen angewandt. Im Falle des Brettspiels reagiert der Algorithmus z.B. jeweils auf die Aktionen seines Gegners. Häufig kann hier nicht unmittelbar bestimmt werden, ob eine Aktion „richtig“ oder „falsch“ ist. An einem gewissen Punkt erreicht der Algorithmus aber ein gewisses Ziel, in dem er beispielsweise Punkte im Spiel generiert oder gar gewinnt. Hierfür werden dann Belohnungen vergeben. Ziel des Algorithmus ist es, zu lernen, auf beliebige Zustände so zu reagieren, dass er möglichst viele Belohnungen erhält [PH20].

5.2. Natural Language Processing

Felix Weber übersetzt Natural Language Processing (NLP) als „Verarbeitung natürlicher Sprache“ und beschreibt es als „Teilgebiet der Informatik und Künstlichen Intelligenz, das sich mit den Wechselwirkungen zwischen Computern und menschlichen (natürlichen) Sprachen befasst, insbesondere mit der Programmierung von Computern zur Verarbeitung und Analyse großer Mengen an natürlichen Sprachdaten.“ [We20, S. 37]. Grundsätzlich muss dabei zwischen der Verarbeitung von gesprochener und geschriebener Sprache unterschieden werden. An dieser Stelle wird aber nur letzteres näher betrachtet, da auch nur dies eine Relevanz für diese Arbeit hat.

(Geschriebene) Texte werden im Bereich des NLP immer als Sequenzen aufgefasst. Eine Sequenz kann hierbei z.B. ein Wort sein. Wörter werden jeweils durch einen Vektor reeller Zahlen, den sogenannten *Embedding*, dargestellt. Wörter mit einer ähnlichen Bedeutung sollen dabei ähnliche Embeddings besitzen, d.h. die Vektoren sollen einen geringen Abstand haben.

5.2.1. Levenshtein-Distanz

Ein Problem im Bereich der Verarbeitung natürlicher Sprache sind Schreibfehler und generell unterschiedliche Schreibweisen. So werden etwa beim Tippen schnell unbemerkt eine falsche Taste angeschlagen. Bei einigen Komposita sind Schreibweisen mit und ohne Bindestrich geläufig. Und letztlich beherrschen auch viele Leute nicht alle Details der Rechtschreibung optimal, sodass schnell Fehler entstehen. Meist ist für den (menschlichen) Leser trotzdem klar, welches Wort gemeint ist. Die Maschine würde bei einem einfachen Vergleich aber keine Übereinstimmung feststellen. Hierfür gibt es einige Algorithmen, die dieses Problem zu lösen versuchen. Einer davon ist die sogenannte Levenshtein-Distanz.

Die Levenshtein-Distanz ist dabei definiert als die kleinste Anzahl an Operationen die nötig sind um von einer Zeichenkette zu einer anderen zu gelangen. Mögliche Operationen sind hier das Einfügen, Löschen und Ersetzen von Zeichen [Na01].

An Hand der Levenshtein-Distanz kann eine prozentuale Übereinstimmung von zwei Zeichenketten definiert werden. Hierfür wird die Levenshtein-Distanz zweier Zeichenketten durch die Länge der längsten der beiden Zeichenketten geteilt und das Ergebnis von 1 subtrahiert.

5.2.2. NLP mit überwachtem Lernen

Im weiteren Verlauf dieser Arbeit wird ein NLP-Problem, genauer ein Klassifizierungsproblem, mit einem Verfahren aus dem überwachten Lernen bearbeitet. Darum werden

die nötigen Schritte an dieser Stelle konkret in diesem Kontext basierend auf [Pe22] beschrieben.

Erstellen eines Datensatzes

Zunächst muss ein sogenanntes Dataframe erstellt werden mit dem ein Modell angelernt und anschließend getestet werden kann. Hierfür wird eine möglichst große Menge an Daten, beispielsweise Texten, benötigt. Je nach Anwendungsfall müssen hier vorab schon einige Operationen stattfinden um beispielsweise relevante Teile aus einem Gesamtwerk zu extrahieren oder Ähnliches.

Klassifizierung

Anschließend müssen die Datenpunkte manuell klassifiziert werden. Das, was später das Modell erledigen soll, muss hier zunächst eigenhändig erledigt werden. Es könnte etwa zu jedem vorhandenen Text die Sprache oder der Autor aufgeführt werden.

Bereinigen des Datensatzes

Um einen sinnvoll nutzbares Dataframe für das ML zu erlangen, müssen die Texte anschließend noch überarbeitet werden. Hierfür werden mehrere Schritte ausgeführt:

1. Transformation in Kleinschreibung
2. Entfernen von Punktzeichen (?.:,;!)
3. Lemmatisierung oder Stemming (siehe unten)
4. Entfernen von Stoppwörtern (Artikel, Konjunktionen, Präpositionen usw.)

Lemmatisierung und Stemming Lemmatisierung und Stemming sind verschiedene Verfahren des NLP um zusammengehörige Wörter zu gruppieren um sie anschließend besser klassifizieren zu können [KB21].

Beim Stemming werden hierfür Worte auf ein Stammwort reduziert, welches für sich betrachtet keinen Sinn ergeben muss, aber die Grundlage für verschiedene Ableitungen bildet.

Bei der Lemmatisierung basiert auf einem ähnlichen Verfahren, produziert aber Wurzelwörter statt Stammwörtern. Diese vermitteln bereits die Bedeutung des lemmatisierten Wortes.

Für beide Verfahren gibt es verschiedene Algorithmen, die an dieser Stelle nicht näher betrachtet werden.

Codierung

Die Texte und auch die Klassen müssen anschließend noch numerisch codiert werden. Bei den Klassen ist dies in der Regel recht einfach. Hier können einfach fortlaufende Ganzzahlen verwendet werden.

Die Codierung der Texte ist komplexer. Diese werden in der Regel in Vektoren überführt. Auch hierfür gibt es verschiedene Algorithmen. Im Verlauf dieser Arbeit wird die Methode *Term Frequency - Inverse Document Frequency* (TF-IDF) verwendet. Hierbei gibt die *Term Frequency* an, wie oft ein Wort bzw. Term in einem Dokument vorkommt. *Die Inverse Document Frequency* vergibt auf Basis der Term Frequency Gewichte an die Worte. Insgesamt wird so die Bedeutung eines Begriffs angegeben [Ko19].

5.2.3. Aufteilung des Datensatzes

Vor dem Training des Modells muss der Datensatz nun noch in einen Trainingsdatensatz, mit dem das Modell angelernt wird und einen Testdatensatz, mit dem die Leistungsfähigkeit des Modells geprüft wird, aufgeteilt werden.

5.2.4. Auswahl eines Algorithmus und Festlegen der Hyperparameter

Es gibt verschiedenste Algorithmen um das Modell anzulernen. Auf diese soll an dieser Stelle nicht näher eingegangen werden. Es macht aber unter Umständen Sinn, mehrere Algorithmen für einen konkreten Anwendungsfall zu testen. Jeder Algorithmus verwendet verschiedene Parameter, die vorab festgelegt werden müssen. Sie werden *Hyperparameter* genannt. In der Praxis ist es sinnvoll, verschiedene Parameter zu testen um ein bestmögliches Ergebnis zu erzielen.

Teil II.

Konzeption

Dieser Teil beschreibt die Konzeption des Visualisierungsansatzes.

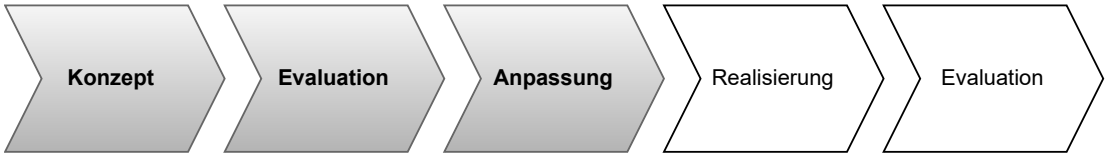


Abbildung 5.3.: Im Teil „Konzeption“ betrachtete Schritte des Vorgehens der Arbeit

In Bezug auf das Vorgehen der Arbeit, welches in Abschnitt 1.2 beschrieben wird, behandelt dieser Teil die ersten drei großen Schritte, wie in Abbildung 5.3 abgebildet: Die eigentliche Konzeption, eine Evaluation der Konzeption und schließlich eine Anpassung des Konzepts.

Hierfür werden als Ausgangspunkt in Kapitel 6 zunächst einige bestehende Ansätze aus verwandten Gebieten beschrieben. Dies betrifft einerseits die Visualisierung von Datenschutz, aber auch allgemeinere Themen, wie die Erweiterung der BPMN oder die Visualisierung durch Farbe.

Danach werden in Kapitel 7 drei exemplarische Geschäftsprozesse aus den Bereichen Personalwesen, Gesundheitswesen und Verwaltung eingeführt, in welchen die Verarbeitung personenbezogener Daten und somit auch der Datenschutz typischerweise eine zentrale Rolle spielen. Die Prozessmodelle dienen später der Veranschaulichung des in dieser Arbeit entwickelten Konzepts.

Der Kern dieses Konzepts wird im Anschluss in Kapitel 8 vorgestellt. Es wird diskutiert, welche Elemente eines Geschäftsprozesses besonders relevant sind und welche Rolle sie bei der Betrachtung des Datenschutzes spielen. Außerdem wird ein Ansatz zur Visualisierung der Datenschutzrelevanz auf Basis von farblichen Markierungen dargestellt.

Nach der Anwendung des Konzepts auf die zuvor eingeführten Geschäftsprozesse folgt in Kapitel 10 eine zweistufige Evaluierung des entwickelten Konzepts. Hierfür wird zunächst das Ergebnis einer Befragung von Experten aus dem Bereich Datenschutz erläutert. Daraufhin wird ein Vergleichsexperiment mit einer deutlich größeren Personengruppe beschrieben. Die Ergebnisse beider Methoden sprechen für die Validität des Konzepts. Es werden aber auch einige Erweiterungsmöglichkeiten genannt, die im folgenden Kapitel 11 aufgegriffen werden.

Den Übergang zur Realisierung bildet in Kapitel 9 die Generalisierung der Anwendung des Konzepts mit theoretischen Ausführungen zur automatisierten Umsetzung.

6. Verwandte Arbeiten

Im Bereich der Datenschutzvisualisierung in Prozessmodellen gibt es bisher vergleichsweise wenige Ansätze. Allerdings gibt es einige Randbereiche, die es zu betrachten lohnt. Hierzu zählen die folgenden Themengebiete:

- Datenschutzaspekte in Prozessmodellen,
- (sonstige) Erweiterung der BPMN,
- Visualisierung durch Farbe,

wobei diese sich natürlich an einigen Stellen überschneiden. Im Folgenden werden jeweils einige Arbeiten aus den einzelnen Bereichen kurz dargestellt, wobei jeweils nur solche Arbeiten ausgewählt werden, die auch wirklich einen klaren Bezug zu dieser Arbeit haben.

6.1. Datenschutz in Geschäftsprozessmodellen

Ein besonders nahes Themenfeld ist der generelle Umgang mit Datenschutzaspekten in Geschäftsprozessmodellen.

Agostinelli et al. präsentieren in [Ag19] eine Sammlung von Design Pattern, die genutzt werden können, um Sachverhalte aus dem Datenschutzrecht in BPMN-Prozessmodellen abzubilden und die Compliance der Prozesse sicherzustellen. Es werden Pattern vorgestellt für den Umgang mit:

- einem Datenleck („Data Breach“),
- dem Einholen einer Einwilligung („Consent to Use the Data“),
- dem Auskunftsrecht („Right to Access“),
- dem Recht auf Datenübertragbarkeit („Right of Portability“),
- dem Recht auf Widerruf einer Einwilligung („Right to Withdraw“),
- dem Recht auf Berichtigung („Right to Rectify“),

- dem Recht auf Vergessenwerden („Right to be Forgotten“).

In [BF20] nutzen Besik und Freytag ebenfalls Design Pattern für die Abbildung der Einwillung in BPMN-Geschäftsprozessmodellen. Es werden Pattern für die Einholung der Einwilligung und deren Widerruf vorgestellt. Neben den Pattern wird auch noch ein Formular eingeführt, welches die Minimalanforderungen an eine Einwilligung abbildet.

In [GCC17] wird ein Modell beschrieben, um Datenschutzrisiken in Prozessmodellen technisch abzubilden.

6.2. Erweiterungen der BPMN

Ein anderes interessantes Feld sind Erweiterungen der BPMN. Generell existiert eine Vielzahl solcher BPMN-Erweiterungen. Ein Überblick findet sich beispielsweise in [Za19]. An dieser Stelle sollen allerdings insbesondere solche, die im Kontext Datenschutz stehen, betrachtet werden. Genau zu diesem Thema gibt es allerdings bisher recht wenig Arbeiten, weshalb auch einige Ansätze aus dem Bereich Sicherheit und Risikomanagement betrachtet werden, da hier einige Anknüpfungspunkte bestehen.

Yulia Cherdantseva beschreibt in ihrer Dissertation [Ch14] unter der Bezeichnung „Secure*BPMN“ einen umfangreichen Ansatz zur Darstellung von Sicherheitsaspekten in Prozessmodellen. Sie nutzt hierfür einerseits selbst entwickelte Piktogramme, andererseits aber auch Farben. Beispielsweise führt sie ein Symbol für die „Security Goal Criticality“, ein, welches je nach Kritikalität in den Schattierungen weiß (niedrig), grau (mittel) und schwarz (hoch) verwendet werden soll. Außerdem wird ein Symbol eingeführt, dass für Datenobjekte, Nachrichten und Datenspeicher deren Sensibilität darstellen soll. Hierfür wird ein Fünfeck gewählt, welches je nach Sensibilität mehr oder weniger Ausrufezeichen enthält und außerdem verschieden eingefärbt wird. Als Grundlage für diese Darstellung dienen die Ausführungen in [Je09] (siehe auch Abschnitt 6.3).

Unter der Bezeichnung „Privacy-Enhanced Business Process Model and Notation“ stellen Pullonen et al. in [PMB17; Pu19a] eine BPMN-Erweiterung vor, die die Nachverfolgung von Datenflüssen und die Kommunikation bzgl. Datenschutzanforderungen erleichtern soll. Hierfür wird eine Menge von Technologien, die den Datenschutz verbessern können, und darauf basierende Stereotypen beschrieben. Zu jedem Stereotyp werden die möglichen ein- und ausgehenden Objekte definiert. Im Prozessmodell werden diese Stereotypen durch dunkelblaue Bezeichnungen unter den entsprechenden Aufgaben repräsentiert. Außerdem werden die Modellelemente eingefärbt: Aufgaben, die im Zusammenhang mit Verschlüsselung stehen, werden rot eingefärbt, benötigte Datenobjekte, die etwa private Schlüssel darstellen, werden grün eingefärbt.

Bartolini et al. schlagen in [BMS15] und [BCM19] eine BPMN-Erweiterung vor, die Datenschutzanforderungen in Prozessmodellen mit Hilfe einer Ontologie (siehe auch [BM15]) abbilden. In der Benutzerschnittstelle führen sie hierfür eine besondere Art von Aufgabe, den „Data Protection Task“, ein. Dieser ist mit einem roten Icon versehen, welches an einen Datenspeicher mit einem Schutzschild erinnert.

In [AMA13] wird eine recht komplexe Erweiterung der BPMN mit der Bezeichnung „Security Risk-Aware BPMN“ vorgeschlagen. Es werden die folgenden Arten der Darstellung genutzt:

- Farbliche Markierung für verschiedene Konstrukte (schwarz = Allgemeines, rot = Risiko und blau = Risikobehandlung).
- Piktogramme, die an Modellelemente angeheftet werden.
- Annotationen.

Brucker et al. präsentieren in [Br13] eine Erweiterung des BPMN-Metamodells um einige sicherheitsspezifische Aspekte, wie etwa Zugriffskontolle, mit der Bezeichnung „SecureBPMN“. In dem Papier wird auch ein erster Versuch einer grafischen Umsetzung durch Piktogramme in Modellelementen gezeigt. Allerdings berichten die Autoren von einer größeren Fallstudie, in welcher der Ansatz zu einem sehr unübersichtlichen und daher wenig hilfreichen Modell geführt hat. Daher wurde der Ansatz verworfen.

Ähnlich ist auch der in [SDG17] präsentierte Ansatz. Interessant ist hier, dass die neu eingeführten Piktogramme alle in einem orangefarbenen Kreis eingebettet sind. Dies lenkt den Fokus bei der Betrachtung klar in deren Richtung.

Auch in [RFP07] wird eine vergleichbare Erweiterung vorgestellt. Die Autoren nutzen hier Piktogramme in Form von Vorhängeschlössern, die teilweise noch einzelne Buchstaben oder ähnliches enthalten, welche an die verschiedenen BPMN-Elemente angehängt werden können, um gewisse Sicherheitsanforderungen darzustellen.

Zareen stellt in [ZAA20] ebenfalls eine BPMN-Erweiterung vor, welche insbesondere zusätzliche Icons für die Darstellung sicherheitsbezogener Aspekte nutzt. Ziel der Arbeit ist ein Framework für das „Security Requirements Engineering“, also die Anforderungsanlyse mit besonderem Fokus auf Sicherheitsaspekte.

In [MSB11] wird ebenfalls eine sicherheitsspezifische Erweiterung der BPMN vorgeschlagen. Hier steht jedoch die Ausführung der Prozesse im Vordergrund. Eine grafische Erweiterung wird daher nicht eingeführt. Stattdessen sollen aber Annotationen nach einem bestimmten Schema verwendet werden, um die relevanten Sachverhalte zu hinterlegen.

Sang und Zhou stellen in [SZ15] eine BPMN-Erweiterung vor, die insbesondere neue, sicherheitsspezifische Ereignis-Typen einführt. Als Anwendungfall verwenden sie

einen Prozess aus dem Gesundheitswesen, der Ansatz sollte sich aber auch auf andere Bereiche übertragen lassen.

Als allgemeiner Ansatz sei an dieser Stelle noch PictureBPMN erwähnt, was bereits in Abschnitt 3.2 eingeführt wird.

6.3. Visualisierung von Risiken durch Farbe

Kummer und Mendling präsentieren in [KM21] einen Ansatz, in dem durch rote bzw. grüne Einfärbung von BPMN Aufgaben dargestellt wird, ob die aufgeführten Aktionen ein Risiko darstellen, oder aber eine Kontrollmöglichkeit bieten, um ein Risiko zu verringern. In der Arbeit wird auch eine durchgeführte Befragung erläutert und ausgewertet, die einen positiven Effekt der Einfärbung für das Verständnis von Risiken nahelegt.

In [FI22] wird das sogenannte Traffic Light Protocol (TLP) beschrieben, welches in verschiedenen Bereichen Verwendung findet. Ursprüngliches Ziel des Protokolls war eine Verbesserung kollaborativen Arbeitens durch sinnvolleres Teilen von Informationen. Hierfür werden vier Klassen verwendet, die durch Farben (bzw. den entsprechenden Bezeichnungen der Farben) gekennzeichnet werden:

TLP:RED	Private Informationen für einzelne Empfänger
TLP:AMBER	Eingeschränkte Verbreitung, wenn notwendig an Personen der eigenen Organisation oder Klienten
TLP:GREEN	Eingeschränkte Verbreitung in der eigenen Community
TLP:CLEAR	Komplett freie Veröffentlichung zulässig

Das Jericho Forum erweitert das TLP in [Je09] beispielsweise, um die Sicherheit in einem bestimmten Kontext zu erhöhen. Hierfür werden eingefärbte Kreise in den Farben des TLP verwendet, in denen je nach Klasse unterschiedlich viele Ausrufezeichen enthalten sind:

Public: Weiß, kein Ausrufezeichen,

Proprietary: Grün, ein Ausrufezeichen,

Restricted Sharing: Bernstein, zwei Ausrufezeichen,

Confidential: Rot, drei Ausrufezeichen.

Die Berliner Datenschutzbeauftragte hat im Rahmen der Corona Pandemie Anbieter von Videokonferenzdiensten in Bezug auf den Datenschutz bewertet. Auch sie verwendet hierfür in [Be20] eine Ampelnotation mit den Farben rot, gelb und grün, wobei letzteres bedeutet, dass keine Mängel gefunden wurden. Gelb bedeutet, dass Mängel gefunden wurden und somit nur unter bestimmten Rahmenbedingungen zu nutzen sind, die

Mängel aber voraussichtlich vergleichsweise problemlos behebbar sind. Rot letztlich meint, dass gravierende Mängel vorhanden sind, die umfassendere Maßnahmen zur Behebung erfordern.

6.4. Zusammenfassung

Die Recherche nach verwandten Arbeiten hat ergeben, dass einige Ideen sehr häufig – teilweise in verschiedenen Kontexten – verwendet werden. Als wichtigste Methoden haben sich die folgenden herauskristallisiert:

- Verwendung von Farbe, um Aspekte hervorzuheben oder zur Klassifizierung;
- Icons bzw. Piktogramme, die an Modellelemente angehängt werden, um Zusatzinformationen zu vermitteln;
- Annotationen an verschiedenen Modellelementen für Zusatzinformationen (teilweise für die Automatisierung);
- Design Pattern in Form von Prozesssequenzen, die in Prozessmodellen wiederverwendet werden können.

Die Verwendung von Design Pattern ist hier von den anderen Ansätzen abzugrenzen, da sie ein anderes Ziel verfolgen. Bei den anderen Ansätzen geht es vorrangig darum, (Datenschutz-)Aspekte im Prozessmodell hervorzuheben und so die nötige Awareness zu schaffen. Die Design Pattern hingegen verfolgen hier das Ziel, rechtssichere Prozesse zu erstellen. Die Anwendung von Design Pattern erfordert aber Fachwissen der Nutzer, was die Verwendung dieser Muster einschränkt.

Diese Dissertation verfolgt im Kern das Ziel der Darstellung von Datenschutz in Geschäftsprozessmodelle zu diversen Zwecken. Design Pattern werden daher im Folgenden nur am Rande betrachtet.

Die beschriebenen Konzepte zur farblichen Markierung, sowie die Nutzung von Icons, bzw. Piktogrammen und Annotationen fließen in diese Arbeit ein und werden weiterentwickelt.

Tabelle 6.1.: Übersicht über verwandte Arbeiten

	Thema	DS in GPM	Farbe	Icons	Ann.	Pattern
[Ag19]	Design Pattern für Betroffenenrechte	x				x
[AMA13]	Security Risk-Aware BPMN	x	x	x	x	
[BCM19; BMS15]	Ontologiebasierter Datenschutztask	x		x		
[Be20]	Bewertung von Videokonferenzdiensten		x			
[BF20]	Design Pattern für Einwilligung	x				x
[Br13]	SecureBPMN			x		
[Ch14; CHR12]	Secure*BPMN	x	x	x	x	
[FI22]	Traffic Light Protocol		x		x	
[GCC17]	Datenschutzrisiken in GPM	x				
[Je09]	Erweiterung des TLP um Icons		x	x		
[KM21]	Darstellung von Risiken in GPM		x	x		
[MSB11]	Sprache für sichere BPMN-Prozesse				x	
[PMB17; Pu19a]	Privacy Enhanced BPMN	x	x		x	
[RFP07]	BPMN-Erweiterung für Sicherheitsanforderungen			x		
[SDG17]	SecureBPMN	x	x	x		
[SZ15]	BPMN-Erweiterung für Sicherheit im Gesundheitswesen			x		
[ZAA20]	BPMN-Erweiterung für Sicherheits-Anforderungsanlyse		x	x		

In Tabelle 6.1 werden alle betrachteten Arbeiten noch einmal aufgeführt und mit den wichtigsten identifizierten Methoden abgeglichen. Außerdem wird angegeben, ob die Arbeit sich direkt mit Datenschutz in Geschäftsprozessmodellen befasst oder nicht.

Es zeigt sich, dass insbesondere für die Verwendung von Farbe und Icons eine Vielzahl von Arbeiten existieren, die auch jeweils relativ ähnlich sind.

Bei den Arbeiten, die in erster Linie auf Icons setzen, fällt auf, dass die fertigen Prozessmodelle schnell recht unübersichtlich werden. Hierbei besteht die Gefahr, den Betrachter auf den ersten Blick eher zu verwirren als ihm zielführend wichtige Informationen zu vermitteln (siehe dazu auch brucker2013a). Außerdem sind die Icons oft wenig intuitiv und gerade bei einer großen Anzahl verschiedener Symbole ist wohl eine recht hohe Einarbeitungszeit notwendig.

Die farblichen Markierungen hingegen dienen auf jeden Fall als „Hingucker“. In mehreren Arbeiten wird mit einer Ampel-Notation gearbeitet, also den Farben Rot, Gelb bzw. Orange und Grün. Diese Notation ist zumindest grundlegend für die meisten Menschen intuitiv verständlich (siehe auch Abschnitt 4.1). Allerdings ist die ganz konkrete Definition der einzelnen Farben in jedem Kontext etwas anders und muss letztlich doch erst mit Hilfe zusätzlicher Quellen erlernt werden.

Als ergänzendes Mittel wird in einigen Arbeiten mit textuellen Annotationen gearbeitet. Diese bieten den großen Vorteil, dass sehr eindeutig und verständlich Informationen dargestellt werden können. Allerdings kann ein Prozessmodell mit vielen solchen Annotationen auch schnell wieder überladen wirken und häufig fallen die Annotationen neben dem übrigen Text im Modell nicht so schnell ins Auge, weshalb der Fokus möglicherweise nicht direkt auf die wichtigsten Aspekte fällt.

7. Beispiele für datenschutzkritische Geschäftsprozesse

Um die folgenden Konzepte besser verdeutlichen zu können, werden an dieser Stelle drei exemplarische Geschäftsprozesse als Anwendungsbeispiele eingeführt. Die ersten beiden Prozesse (siehe Abschnitt 7.1 und Abschnitt 7.2) sind fiktiv und basieren auf Prozessen, die von Carolin Goppelt im Rahmen ihrer Masterarbeit [Go23] modelliert wurden und auch für die Evaluation des Konzepts (siehe Kapitel 10) verwendet wurden. Für diese Arbeit wurden die Prozesse allerdings etwas überarbeitet. Beide Prozesse sind bewusst so gewählt, dass sie einerseits für jedermann verständlich sind und andererseits später möglichst alle Aspekte des in Kapitel 8 beschriebenen Konzepts dargestellt werden können. Wie bereits erwähnt, ist dies ist insbesondere deshalb wichtig, da sie zur Evaluation des Konzeptes verwendet und hierfür verschiedenen Personengruppen präsentiert wurden.

Zusätzlich wird in Abschnitt 7.3 ein realer Geschäftsprozess aus der Verwaltung der Christian-Albrechts-Universität zu Kiel (CAU) eingeführt.

In diesem Kapitel erfolgt zunächst nur eine allgemeine Beschreibung der Prozesse bzw. der entsprechenden Prozessmodelle. Die datenschutzspezifische Betrachtung erfolgt in Verbindung mit der Anwendung des Konzeptes aus Kapitel 8 in Kapitel 9.

7.1. Personalwesen: Einstellung

Der erste Beispielprozess betrifft die Einstellung eines neuen Mitarbeiters inklusive der nötigen Anmeldungen. Das Prozessmodell wird in Abbildung 9.1 dargestellt.

Das Prozessmodell besteht aus vier Beteiligten, jeweils durch einzelne Pools dargestellt. Die Hauptakteure sind der Arbeitgeber und der (zukünftige) Arbeitnehmer. Die beiden Nebenakteure, die Sozialversicherung und das Bundeszentralamt für Steuern, werden nur durch Black-Box-Pools abgebildet.

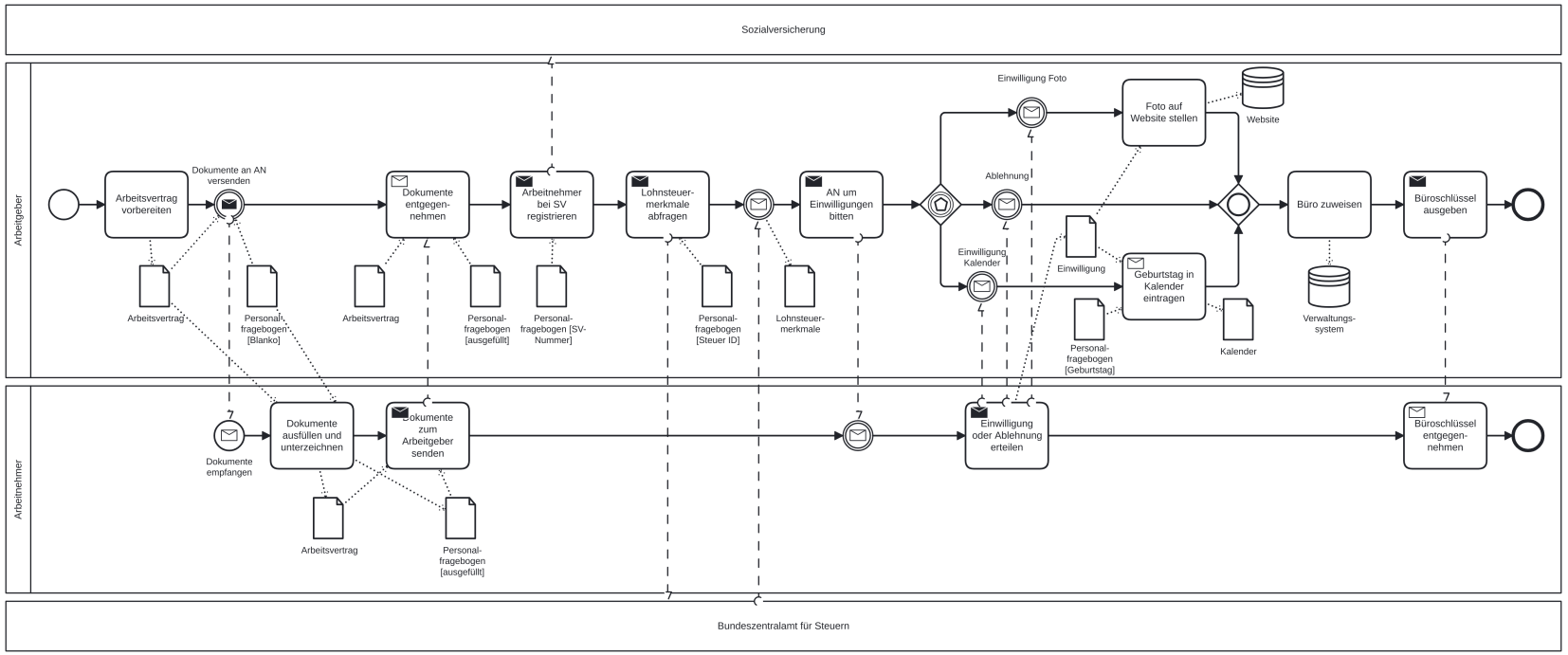


Abbildung 7.1.: Geschäftsprozessmodell zur Einstellung eines Mitarbeiters

Vorbedingung des Prozesses ist ein positiv abgeschlossener Bewerbungsprozess.

Der Prozess beginnt dann mit der Vorbereitung des Arbeitsvertrags, welcher anschließend zusammen mit einem Personalfragebogen zum Arbeitnehmer gesendet wird. Der Arbeitnehmer empfängt die Dokumente, füllt sie aus, unterzeichnet und sendet sie zurück an den Arbeitgeber. Daraufhin meldet der Arbeitgeber den neuen Mitarbeiter mit Hilfe der erhaltenen Sozialversicherungsnummer bei der Sozialversicherung an. Anschließend ruft der Arbeitgeber beim Bundeszentralamt für Steuern mit Hilfe der Steuer-ID des Arbeitnehmers dessen Lohnsteuerabzugsmerkmale ab. Danach bittet der Arbeitgeber den Arbeitnehmer um seine Einwilligungen, um einerseits dessen Foto auf die Unternehmenswebsite zu stellen und andererseits seinen Geburtstag in einen geteilten Kalender einzutragen. Das folgende ereignisbasierte Gateway wartet, bis entweder eine Ablehnung oder die Einwilligung vorliegt. Wenn eine oder beide Einwilligungen vorliegen, werden die jeweiligen Datenverarbeitungen durchgeführt. Anschließend – oder im Falle einer Ablehnung direkt – wird auf Seite des Arbeitgebers dem neuen Mitarbeiter noch ein Büro zugewiesen und der passende Schüssel ausgegeben. Der Prozess endet mit der Entgegennahme des Schlüssels durch den Arbeitnehmer.

Im Prozessmodell wurden bewusst einige Details weggelassen, um nicht zu komplex zu werden. Insbesondere werden keine Ausnahmen behandelt. In der Realität könnte es natürlich beispielsweise sein, dass der Arbeitnehmer den Vertrag nicht unterzeichnet oder es könnten Angaben fehlen, etc. Mindestens für eine sinnvolle (Teil-)-Automatisierung mit Hilfe einer Workflowengine, aber auch für eine wirklich brauchbare Dokumentation im Unternehmen wären entsprechende Ausnahmebehandlungen natürlich sinnvoll.

7.2. Gesundheitswesen: Zahnarztbesuch

Der andere fiktive betrachtete Prozess betrifft den Besuch eines Patienten bei einem Zahnarzt. Das Prozessmodell findet sich in Abbildung 7.2. Auch hier wurden zur Vereinfachung wieder einige Aspekte außer Acht gelassen.

In diesem Prozessmodell sind fünf verschiedene Teilnehmer(-rollen) vertreten. Drei davon sind in einem gemeinsamen Pool zusammengefasst. Dieser bildet die Zahnarztpraxis während des Besuchs ab. Der Pool ist in zwei Swimlanes unterteilt, wobei eine davon wiederum in zwei Lanes unterteilt ist. Eine Lane bildet der Patient. Die andere Lane bildet das eigentliche Praxispersonal ab und besteht aus Lanes für den Arzt und den medizinischen Fachangestellten (MFA). Zusätzlich wird noch eine externe Abrechnungsstelle betrachtet und außerdem die Krankenkasse (KK) des Patienten als Black-Box-Pool.

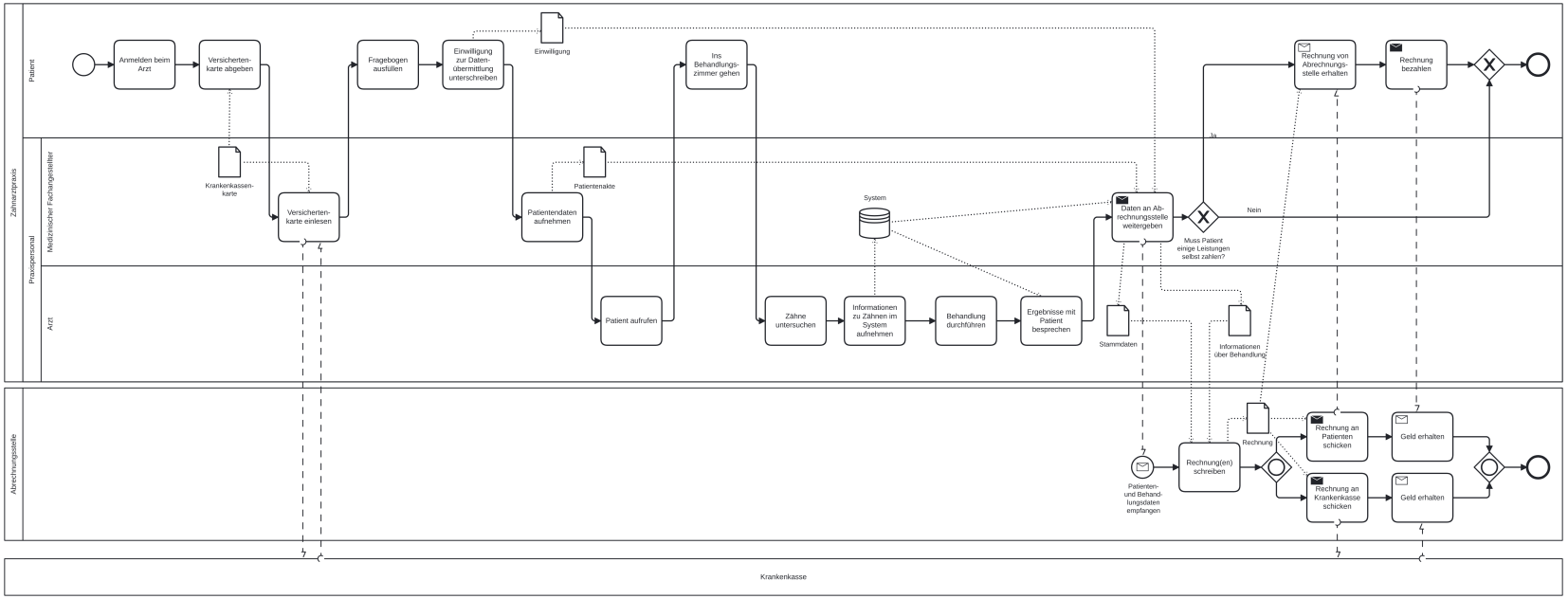


Abbildung 7.2.: Geschäftsprozessmodell eines Zahnarztbesuchs

Bei strenger Beachtung des BPMN-Standards ist die Zusammenführung von Praxispersonal und Patient in einem Pool eigentlich nicht korrekt, da beide eigenständige Entscheidungen treffen können und der Patient nicht wirklich der Arztpraxis untergeordnet ist. Einerseits würde eine Trennung in verschiedene Pools für die Modellierung aber bedeuten, dass keine direkten Kontrollflüsse zwischen beiden Teilnehmern fließen könnten, sondern ständig Nachrichten ausgetauscht werden müssten, was den Prozess unnötig verkompliziert. Andererseits herrscht im vorliegenden Fall schon eine gewisse Weisungsbefugnis des Praxispersonals gegenüber des Patienten, weshalb eine Zusammenfassung evtl. doch als legitim angesehen werden kann. Insgesamt sollte Semantik der Aufteilung so für jeden Betrachter verständlich sein.

Der Prozess beginnt, nachdem der Patient die Praxis betreten hat. Er meldet sich am Empfang an und gibt seine Krankenkassenkarte ab, welche daraufhin vom MFA eingelesen wird. Hierbei erfolgt ein nicht näher definierter, bidirektionaler Datenaustausch mit der KK um den Versichertenstatus zu verifizieren. Anschließend muss der Patient einen Anamnesebogen ausfüllen und eine Einwilligung zur Datenübermittlung unterschreiben. Der Arzt nimmt die Daten auf. Wenn der Patient an der Reihe ist, wird er vom Arzt aufgerufen und geht ins Behandlungszimmer. Dort angekommen untersucht der Arzt die Zähne und dokumentiert alle Befunde in seinem Verwaltungssystem. Anschließend wird die eigentliche Behandlung durchgeführt und mit dem Patienten besprochen. Abschließend gibt der Arzt alle relevanten Daten an die Abrechnungsstelle weiter. Diese schreibt dann eine Rechnung und sendet diese je nach Leistungen an den Patienten und/oder die Krankenkasse. Der Patient bzw. die Krankenkasse erhalten dann entsprechend die Rechnung und zahlen diese. Damit endet der Prozess für alle Beteiligten.

7.3. (Öffentliche) Verwaltung: Erstellung eines Studentenausweises

In diesem Abschnitt wird noch ein reales Geschäftsprozessmodell der CAU vorgestellt. Es handelt sich um einen Prozess zur Erstellung eines Studentenausweises, der sogenannten CAU Card. Der Prozess besteht aus einem übergeordneten Hauptprozess (siehe Abbildung 7.3) und drei Teilprozessen, die in den folgenden Unterabschnitten beschrieben werden. Alle Prozessmodelle wurden ursprünglich durch das Prozessmanagement-Team der CAU in Picture-BPMN modelliert (siehe Abschnitt 3.2) und für diese Arbeit durch die Autorin in den Camunda Modeler (siehe Abschnitt 14.1.2) importiert und dort überarbeitet. Hierbei wurde nur das Layout angepasst und für die späteren Ausführungen in Abschnitt 9.3 Farben ergänzt. Es wurden keine inhaltlichen Änderungen

vorgenommen um das Bild eines realen Prozesses nicht zu verfälschen. Syntaktische und sprachliche Ungenauigkeiten entsprechen somit auch dem Original aus der Praxis[1].

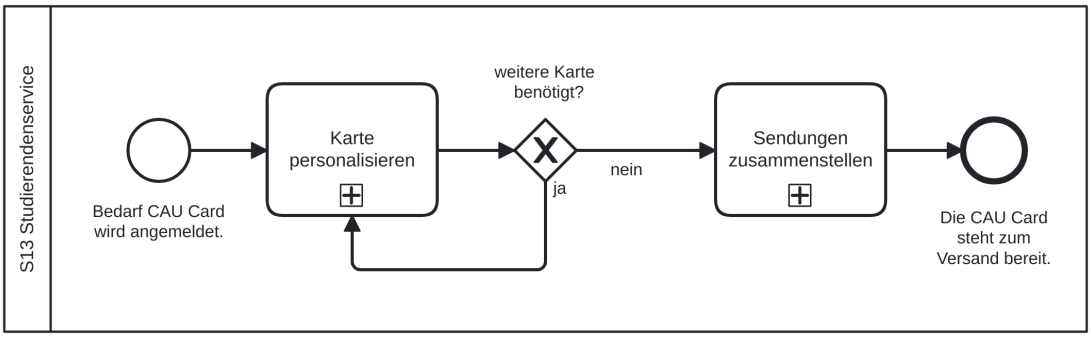


Abbildung 7.3.: Geschäftsprozessmodell zur Erstellung eines Studentenausweises

Der Hauptprozess zur Erstellung der CAU Card ist sehr kurz gehalten. Es existiert nur ein Pool, welcher den Studierendenservice abbildet. Der Prozess wird durch eine Bedarfsmeldung ausgelöst. Daraufhin wird der Teilprozess „Karte personalisieren“ gestartet (siehe Abschnitt 7.3.1). Falls eine weitere Karte benötigt wird, wird dieser Teilprozess solange erneut gestartet, bis keine weitere Karte benötigt wird. Anschließend wird der Teilprozess „Sendungen zusammenstellen“ (siehe Abschnitt 7.3.2) gestartet und der Prozess damit beendet, dass die CAU Card zum Versand bereitsteht.

7.3.1. Karte personalisieren

Der erste aufgerufene Teilprozess, abgebildet in Abbildung 7.4, betrifft die Personalisierung der CAU Card. Zunächst meldet der zuständige Mitarbeiter sich an einem Rechner an und startet die SmartLIFE-Anwendung. Dort werden dann die benötigten Datensätze importiert. Als Schlüssel dienen hierfür die Matrikelnummern der Studenten. Anschließend wird für jeden Studenten an verschiedenen Stellen ein Passfoto gesucht. Wenn kein Foto gefunden werden kann, wird ein neues Foto beim Studenten angefragt und nach einer nicht näher definierten Frist erneut überprüft, ob nun ein Foto vorhanden ist. Wenn ein geeignetes Foto vorhanden ist, wird bei Bedarf der Bildausschnitt angepasst. Anschließend wird die CAU Card codiert und bedruckt. Wenn die Karte in Ordnung ist, endet der Teilprozess, ansonsten wird ein neuer Versuch unternommen.

[1] Hier sind einige interessante (teil-)automatisierte Hilfestellungen für die Korrektur denkbar, die die Anwendung des Färbungskonzepts möglicherweise vereinfachen würden (siehe z. B. [GNV17; Ro19; Sn15]). Dies ist aber nicht Forschungsgegenstand dieser Arbeit und wird daher nicht näher betrachtet.

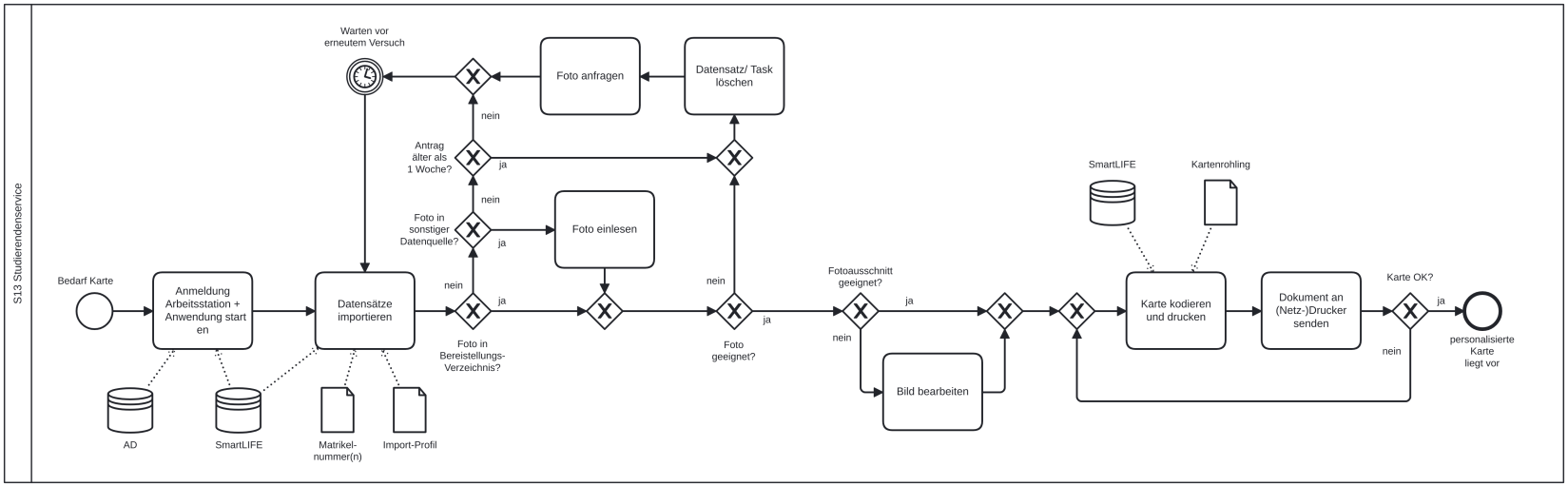


Abbildung 7.4.: Geschäftsprozessmodell „Karte personalisieren“

7.3.2. Sendung zusammenstellen

Nach dem Personalisieren der Karte folgt das Zusammenstellen der Sendung an den Studenten, wie in Abbildung 7.5 dargestellt. Je nach verwendetem Drucker wird zunächst eine Anmeldung bei dessen Betreiber ausgeführt und anschließend der Druck der Anschreiben auf Briefbögen gestartet. Neben dem Anschreiben wird noch ein Beiblatt benötigt, welches bei Bedarf gefalzt wird. Auf das zuvor gedruckte Anschreiben wird im Teilprozess „Karte aufspenden“ (siehe Abschnitt 9.3.3) die Karte aufgebracht, bevor das Anschreiben mit der Karte zusammen mit dem Beiblatt in einem Umschlag verpackt wird. Der Prozess endet mit der versandfertigen Sendung.

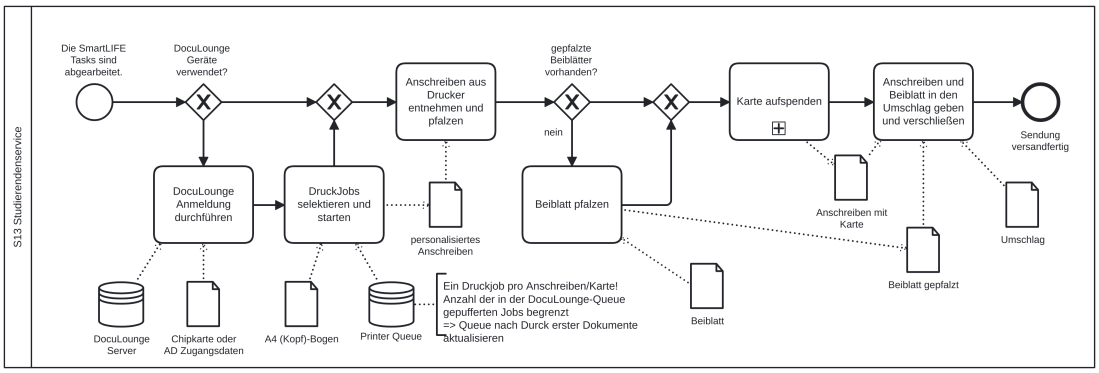


Abbildung 7.5.: Geschäftsprozessmodell „Sendung zusammenstellen“

7.3.3. Karte aufspenden

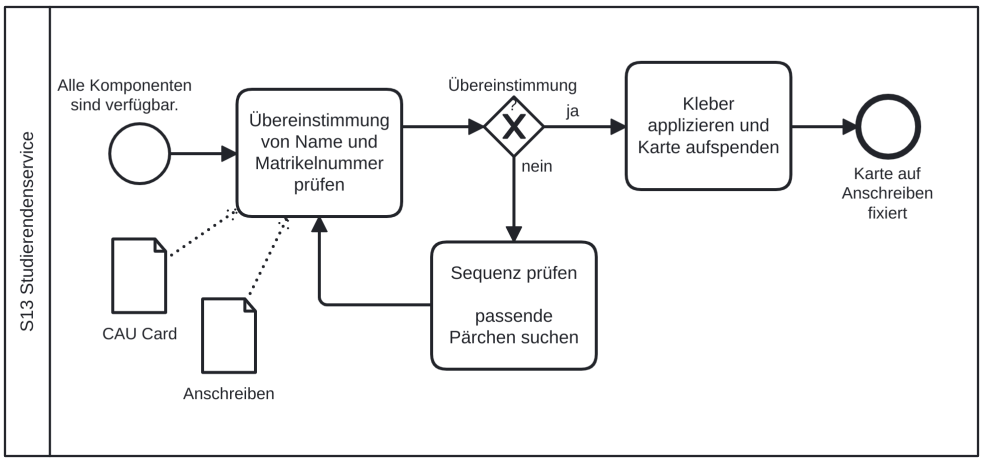


Abbildung 7.6.: Geschäftsprozessmodell „Karte aufspenden“

Als letzter Teilprozess wird noch das „Karte aufspenden“(siehe Abbildung 9.6) betrachtet. Hier werden zunächst die Daten auf Anschreiben und Karte verglichen, wenn diese nicht übereinstimmen, wird nach einem passenden Pärchen gesucht. Wenn eine Übereinstimmung vorliegt wird Kleber appliziert und die Karte auf dem Anschreiben angebracht. Damit endet der Teilprozess.

8. Repräsentation von Datenschutz in Geschäftsprozessmodellen

In der Einleitung dieser Arbeit (siehe Kapitel 1) wird die Notwendigkeit einer Unterstützung für den Umgang mit Datenschutz in Unternehmen und anderen Organisationen gezeigt. Außerdem wird dort dargelegt, dass die Visualisierung von Datenschutzaspekten in Geschäftsprozessmodellen eine Hilfestellung darstellen kann. Kapitel 6 zeigt einige Ansätze, die Problemstellung zu lösen – mit Hilfe von Geschäftsprozessmodellen und auch ohne. Die Ansätze weisen allerdings allesamt einige Schwächen auf. In diesem Kapitel wird ein neuer Ansatz dargestellt, welcher vorrangig Farben für die Visualisierung von Datenschutzaspekten verwendet.

Der Ansatz basiert auf einer Bewertung verschiedener Aspekte des Prozesses einerseits bezüglich der gesetzlichen Bestimmungen, die das Unternehmen einzuhalten hat; und andererseits auf Basis des entstehenden Risikos für Personen, die von den Datenverarbeitungen innerhalb des Prozesses betroffen sind. Der Risikobegriff wird hier aus der Definition des SDM entnommen:

> „Ein Risiko im Sinne der DS-GVO ist das Bestehen der Möglichkeit des Eintritts eines Ereignisses, das einen Schaden für die Rechte und Freiheiten natürlicher Personen (einschließlich ungerechtfertigter Beeinträchtigung der Rechte und Freiheiten) darstellt oder zu einem Schaden für eine oder mehrere natürliche Personen führen kann. Es hat zwei Dimensionen: Erstens die Schwere des Schadens für die Rechte und Freiheiten der betroffenen Personen und zweitens die Wahrscheinlichkeit, dass das Ereignis und der Schaden eintreten.“ [AK22, S. 48]

Das Konzept für die Färbung von BPMN-Prozessen im Zusammenhang mit dem Datenschutz, wie es hier dargestellt wird, wurde erstmals in [WSG21] beschrieben. Im Folgenden werden allerdings noch einige Ergänzungen eingeführt, welche unter Anderem auf den Rückmeldungen zu dem genannten Papier basieren.

Grundlage des Konzepts bietet die Kategorisierung verschiedener Prozesselemente bezüglich deren Kritikalität im Bezug auf den Datenschutz. Entsprechend dieser Kategorisierung werden die Prozesselemente dann auf Basis der Erkenntnisse aus Abschnitt 4.1

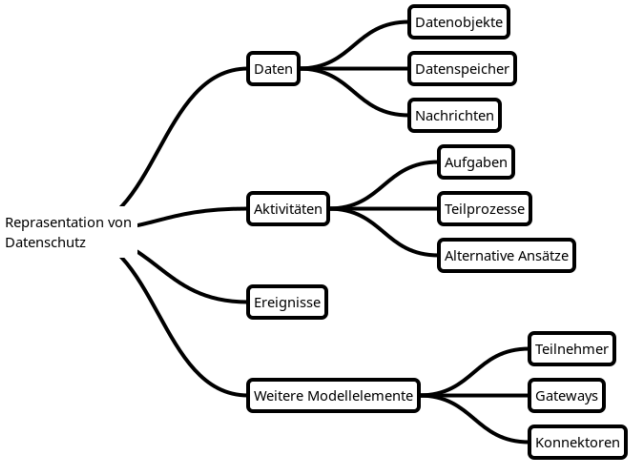


Abbildung 8.1.: Aufbau des Kapitels

in einer Ampelnotation eingefärbt, wie sie etwa auch im Traffic Light Protocol[FI22] verwendet wird.

Als besonders relevante Prozesselemente bezüglich des Datenschutzes wurden Datenobjekte und Speicher, sowie Aktivitäten und Ereignisse identifiziert. Abschnitt 8.1 erläutert daher zunächst die Farbkategorien für Datenobjekte und Datenspeicher. In Abschnitt 8.2 werden dann die Farbkategorien für Aktivitäten eingeführt und in Abschnitt 8.3 diejenigen für Ereignisse. Anschließend geht Abschnitt 8.4 der Frage nach, inwiefern die anderen Modellelemente relevant für die Betrachtung des Datenschutzes sind.

Abbildung 8.1 visualisiert den Aufbau des Kapitels in Form einer Mindmap.

8.1. Daten

In Abschnitt 3.1.2 wurden die verschiedenen BPMN-Elemente beschrieben, die Daten repräsentieren. Im Wesentlichen ist hier zwischen Datenobjekten und Datenspeichern zu unterscheiden. Beides ist offensichtlich hochgradig relevant für die Darstellung von Datenschutzproblematiken in Prozessmodellen, da in beiden Fällen unmittelbar personenbezogene Daten enthalten sein können. Dies wird insbesondere dann problematisch, wenn unzulässige Zugriffe auf die Daten stattfinden. Hier entsteht also ein Risiko für für die betroffene Person. Dieses Risiko ist offensichtlich vom konkreten Inhalt der

Datenobjekte bzw. -speicher abhängig. In Bezug auf die Risikodefinition des SDM wird hier die erste Dimension des Risikos, die Schwere des Schadens, betrachtet.

Dementsprechend sollten sowohl Datenobjekte als auch Datenspeicher kategorisiert und eingefärbt werden. Da Datenspeicher letztlich eine Sammlung von Datenobjekten darstellen, ist die Färbung eng miteinander verbunden. Im Folgenden wird zunächst die Färbung der Datenobjekte und anschließend die der Datenspeicher erläutert.

8.1.1. Datenobjekte

Wie bereits erwähnt, stellen Datenobjekte unmittelbar Daten dar, die in vielen Fällen auch personenbezogen und damit datenschutzrelevant sein können. Die Färbung der Datenobjekte ist auf den ersten Blick recht einfach und wird in Tabelle 8.1 dargestellt.

Tabelle 8.1.: Regeln für die Färbung von Datenobjekten

Farbe	**Inhalt des Datenobjekts**
Grün	Nicht personenbezogene Daten
Gelb	Personenbezogene Daten
Rot	Besondere Kategorien personenbezogener Daten

Datenobjekte, die keinerlei personenbezogene Daten enthalten, werden grün eingefärbt. Beispiele hierfür könnten etwa eine Anforderungsbeschreibung oder ein Handbuch sein. Hier besteht kein direktes Risiko für die Grundrechte natürlicher Personen. Datenobjekte mit personenbezogenen Daten hingegen werden gelb eingefärbt. In diese Kategorie fällt beispielsweise ein Arbeitsvertrag, der Name und Adresse der beschäftigen Person enthält. Unzulässige Zugriffe auf derartige Daten würden durchaus einen Eingriff in das Grundrecht auf informationelle Selbstbestimmung der betroffenen Personen darstellen. Darüber hinaus können Datenobjekte aber auch noch besondere Kategorien personenbezogener Daten enthalten, wie beispielsweise ein Personalfragebogen aus dem etwa die Konfession hervorgeht. Derartige Datenobjekte werden rot eingefärbt, da hier ein noch schwerwiegender Eingriff bestünde.

Abbildung 8.2 zeigt exemplarisch die gefärbten Datenobjekte.

Die Färbung der Datenobjekte bietet den Vorteil, dass auf den ersten Blick klar ist, welche Daten schützenswert sind und wo somit eventuell besondere TOM ergriffen werden müssen (siehe Abschnitt 2.1.8)

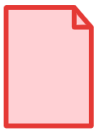
Häufig haben Datenobjekte einen komplexen Aufbau und enthalten Daten aus unterschiedlichen Kategorien. So enthält der Personalfragebogen etwa nicht ausschließlich besondere Kategorien personenbezogener Daten, sondern überwiegend „normale“ personenbezogene Daten, wie etwa die Adresse. In derartigen Fällen wird immer von den

Produktinfos

(a) Kennzeichnung eines grünen Datenobjekts

Lieferschein

(b) Kennzeichnung eines gelben Datenobjekts



Patientenakte

(c) Kennzeichnung eines roten Datenobjekts

Abbildung 8.2.: Gefärbte Datenobjekte

enthaltenen Daten mit dem höchsten Schutzniveau ausgegangen. Selbst ein Datenobjekt, welches fast ausschließlich Daten enthält, welche überhaupt keinen Personenbezug haben, würde also rot eingefärbt werden, wenn auch ein Datum einer besonderen Kategorie personenbezogener Daten enthalten wäre.

Wichtig zu betrachten ist das Thema „Anonymisierung“. Häufig werden personenbezogene Daten aufgenommen und anschließend anonymisiert, um beispielsweise statistische Auswertungen durchzuführen. Diese Daten haben dann – korrekte Anonymisierung vorausgesetzt – keinen Personenbezug mehr und können theoretisch grün eingefärbt werden. Allerdings besteht hier die Frage, wie zuverlässig erkannt werden kann, ob ein Dokument (korrekt) anonymisiert wurde. Grundlegende Überlegungen dazu werden in Abschnitt 12.3 und Abschnitt 12.3.1 erläutert. Die Fragestellung, wie mit – eventuell inkorrekter – Anonymisierung umgegangen werden soll, wurde auch in den durchgeführten Expertenbefragungen angesprochen, deren Ergebnisse in Abschnitt 10.1 zusammengefasst werden.

8.1.2. Datenspeicher

Wie in Abschnitt 3.1.2 beschrieben, existieren neben den Datenobjekten im im BPMN-Standard auch noch Datenspeicher. Die Färbung dieser basiert hier auf der Färbung der Datenobjekte: Grundsätzlich sind Datenspeicher immer so einzufärben, dass sie das strengste Schutzniveau aller beinhalteten Daten abbilden. Wenn also beispielsweise eine Datenbank ausschließlich nicht personenbezogene Daten enthält, oder gar leer ist, dann kann der repräsentierende Datenspeicher grün eingefärbt werden. Wenn allerdings möglicherweise personenbezogene Daten enthalten sind, so muss der Datenspeicher gelb eingefärbt werden. Analog verhält es sich mit Daten aus den besonderen Kategorien personenbezogener Daten und roten Datenspeichern.

Allerdings wird aus einem Prozessmodell nicht immer klar, welche Daten enthalten sind, da gewöhnlich nicht der komplette Lebenszyklus des Datenspeichers abgebildet ist. Selbst wenn in einem Prozessmodell also nur Daten ohne Personenbezug in einem Da-

tenspeicher abgelegt werden, ist es ja durchaus denkbar, dass in einem anderen Prozess auch Daten mit Personenbezug im gleichen Datenspeicher abgelegt werden. Dies lässt sich aber nicht immer nachvollziehen, weshalb die Färbung hier immer nur eine grobe Annäherung an das tatsächliche Schutzniveau sein kann. Eine Ausnahme bilden hier unter Umständen Datenspeicher, deren aktueller Status explizit (in eckigen Klammern, siehe Abschnitt 3.1.2) angegeben ist. Falls hier beispielsweise ein Datenspeicher den Zusatz *[leer]* hat, kann wohl davon ausgegangen werden, dass keine (personenbezogenen) Daten enthalten sind und somit der Datenspeicher grün eingefärbt werden kann.

8.1.3. Nachrichten

Nachrichten stellen im Wesentlichen Daten da, welche zwischen zwei Pools ausgetauscht werden. Sie können daher analog zu den Datenobjekten behandelt werden. Zu beachten ist hier, dass der Austausch der Nachrichten an sich natürlich eine relevante Datenübertragung darstellt, für die auch oft eine Einwilligung nötig ist. Hier ist dann aber die ausführende Aktivität das zu betrachtende Modellelement. Die Bewertung der Aktivitäten wird im folgenden Abschnitt erläutert. Das Thema der Nachrichten-Aktivitäten wird außerdem noch konkreter in Abschnitt 12.3 diskutiert.

8.2. Aktivitäten

Neben den Modellelementen, die Daten repräsentieren, haben Aktivitäten eine hohe Relevanz für den Datenschutz, da diese unter Umständen Verarbeitungstätigkeiten abbilden.

Nur auf Basis der Aktivitäten kann schlecht ein Risiko bewertet werden. Nach Definition des SDM besteht dies aus den Dimensionen der Schwere des Schadens und der Eintrittswahrscheinlichkeit [AK22]. Die Schwere des Schadens wird in erster Linie durch die verarbeiteten Daten bestimmt (siehe Abschnitt 8.1). Die Eintrittswahrscheinlichkeit ist von einer Vielzahl von Faktoren abhängig. Betrachtet man etwa die Speicherung von Daten als Aktivität, so muss hier bewertet werden, in welchem System Daten gespeichert werden, ob diese verschlüsselt abgelegt werden, was der Serverstandort ist, etc. Dies lässt sich aus dem Prozessmodell aber in aller Regel nicht ablesen. Daher kommt eine Bewertung auf Basis des Risikos für den Betroffen nicht in Frage. Stattdessen findet die Bewertung hier auf Basis der einschlägigen Vorschriften statt.

Das Schema zur Färbung der Aktivitäten ist etwas komplexer als jenes im Bereich der Daten. Generell muss hier, wie in Abschnitt 3.1.1 erläutert, zwischen Aufgaben und Teilprozessen unterschieden werden.

8.2.1. Aufgaben

Tabelle 8.2.: Regeln für die Färbung von Aufgaben

Farbe	Art des Aufgaben
Grün	Keine relevante Datenverarbeitung
Gelb	Legitime Datenverarbeitung
Rot	Einwilligung erforderlich

Tabelle 8.2 gibt einen Überblick über mögliche Regeln zur Einfärbung von Aufgaben.

Der einfachste Fall für die Färbung einer Aufgabe ist, dass keine relevante Datenverarbeitung vorliegt. Dies ist einerseits der Fall, wenn schlicht keine personenbezogenen Daten verarbeitet werden. Andererseits ist es aber beispielsweise auch möglich, dass eine betroffene Person selbst ihre eigenen personenbezogenen Daten in einer Aufgabe nutzt. Dies ist datenschutzrechtlich nicht zu betrachten und kann daher genauso behandelt werden, als würden gar keine personenbezogenen Daten verarbeitet werden (siehe hierzu auch Abschnitt 12.3). Alle Aufgaben, in denen also keine relevante Datenverarbeitung nach der obigen Beschreibung vorliegt, werden grün eingefärbt.

Unter allen Aufgaben, in denen hingegen keine relevante Datenverarbeitung stattfindet, wird nach der Rechtsgrundlage der Verarbeitung unterschieden. Wie in Abschnitt 2.1.3 erläutert, gibt es verschiedene Rechtsgrundlagen für die Verarbeitung personenbezogener Daten. Etwas kritisch zu betrachten ist hierbei die Einwilligung (Art. 6 (1) a)), da diese beispielsweise nicht einfach vorausgesetzt und außerdem jederzeit widerrufen werden kann (siehe Abschnitt 2.1.4). Es ist also durchaus ungünstig, an Stellen, die für einen Prozessablauf kritisch sind, sich auf diese Einwilligung zu verlassen. Besser ist es daher, wenn einer der anderen Verarbeitungsgründe aus Art. 6 (1) vorliegen.

Aus diesem Grund werden Aufgaben, die eine Einwilligung erfordern, für die also kein anderer Verarbeitungsgrund vorliegt, als besonders kritisch betrachtet und deshalb rot eingefärbt. Aufgaben hingegen, für die ein anderer Verarbeitungsgrund vorliegt, wie beispielsweise die Vertragserfüllung oder auch eine rechtliche Verpflichtung, werden gelb eingefärbt. Unerheblich bei der Frage, ob eine Aufgabe gelb oder rot einzufärben ist, ist das Vorliegen einer Einwilligung.

Eine beispielhafte Darstellung der gefärbten Aufgaben findet sich in Abbildung 8.3.

8.2.2. Teilprozesse

Relativ einfach verhält es sich, wenn eine Aktivität ein Teilprozess ist. Aufgeklappte Teilprozesse als solches werden zur besseren Übersicht überhaupt nicht eingefärbt.

(a) Kennzeichnung einer grünen Aufgabe

(b) Kennzeichnung einer gelben Aufgabe

(c) Kennzeichnung einer roten Aufgabe

Abbildung 8.3.: Gefärbte Aufgaben

Stattdessen werden alle enthaltenen Aufgaben nach dem oben beschriebenen Schema eingefärbt, sodass auf den ersten Blick zu erkennen ist, ob in dem Teilprozess eine relevante Datenverarbeitung stattfindet.

Bei zugeklappten (aber ausmodellierten) Teilprozessen wird hingegen die Färbung der enthaltenen Aufgabe mit der höchsten Schutzklasse übernommen. Ist also beispielsweise eine rote Aufgabe im Teilprozess enthalten, so wird der zugeklappte Teilprozess ebenfalls rot eingefärbt. Ist keine rote, aber eine gelbe Aufgabe enthalten, so wird der Teilprozess gelb eingefärbt. Grün hingegen ist ein Teilprozess nur dann, wenn lediglich grüne Aufgaben enthalten sind. Problematisch ist hier der Kontext. Bei sehr allgemeinen Teilprozessen, die in vielen verschiedenen Hauptprozessen verwendet werden, sind eventuell mehrere Färbungen denkbar. Hier sollte daher besonders kritisch bewertet werden.

Etwas problematischer verhält es sich mit zugeklappten Teilprozessen, die zum Zeitpunkt des Einfärbens (noch) nicht ausmodelliert wurden, also gewissermaßen eine Black-Box darstellen. Hier besteht nur die Möglichkeit, diese wie einfache Aufgaben zu behandeln.

8.2.3. Alternative Ansätze

Im Gegensatz zur Definition der Farbkategorien für die Datenobjekte ist jene für Aufgaben deutlich weniger intuitiv. Während die Definition *„grün = keine relevante Datenverarbeitung“* noch recht offensichtlich ist, kommen für die Unterscheidung zwischen gelben und roten Aufgaben durchaus mehrere Varianten in Frage.

Zulässige Datenverarbeitungen

Eine Möglichkeit wäre es beispielsweise, zwischen zulässigen und nicht zulässigen Datenverarbeitungen zu unterscheiden. Mit einer Einwilligung ist aber grundsätzlich zunächst jede Datenverarbeitung zulässig. Natürlich könnte im Prozessmodell überprüft werden, ob eine Einwilligung für eine Datenverarbeitung (für die kein sonstiger Verarbeitungsgrund vorliegt) vorhanden ist. Falls nicht, könnte die entsprechende Auf-

gabe dann als unzulässig klassifiziert und somit rot eingefärbt werden. Allerdings ist fraglich, ob die Einwilligung in jedem Fall konkret modelliert wird – auch wenn dies natürlich wünschenswert wäre. Somit könnte es zu vergleichsweise häufigen falschen Rotfärbungen kommen. Für Betrachter, die kein vertieftes Datenschutzwissen haben und nur die Definition der einzelnen Farben zur Hilfe haben, könnte es dann schnell so wirken, als sei eine Datenverarbeitung wirklich in jedem Fall verboten und das könnte unter Umständen komplexe Änderungen der Geschäftsprozesse nach sich ziehen.

Als zusätzlicher Aspekt könnten hier noch die Grundsätze für die Verarbeitung nach Art. 5 (1) betrachtet werden. Beispielsweise gilt es im Sinne der Datenminimierung zu bewerten, ob die verarbeiteten Daten wirklich benötigt werden. Dies – und auch alle anderen Grundsätze – sind allerdings schwierig auf Basis des Prozessmodells zu bewerten. Lediglich die Speicherbegrenzung lässt sich im Prozessmodell wirklich gut abbilden. Allerdings ist hier diskussionswürdig, ob eine Aufgabe beispielsweise rot eingefärbt werden sollte, nur weil im weiteren Prozessverlauf keine (oder eine zu späte) Löschung verzeichnet ist. Das eigentliche Problem liegt dann ja an einer anderen Stelle. Hier bieten sich viel mehr andere Ergänzende Ansätze an.

Interne und externe Datenverarbeitung

Ein weiterer Ansatz wäre die Unterscheidung nach interner Datenverarbeitung und der Weitergabe von Daten an Dritte, da dies für viele Betroffene einen größeren Einschnitt in ihre Rechte darstellt. Allerdings behandelt die DSGVO die Weitergabe von Daten grundlegend genau so, wie jede andere Datenverarbeitung auch. Wenn ein legitimer Verarbeitungsgrund vorliegt, ist die Weitergabe zulässig und auch nicht zwangsweise an strengere Dokumentationspflichten oder Ähnliches geknüpft. Etwas komplizierter ist die Weitergabe personenbezogener Daten an Stellen außerhalb der EU. Dies ist allerdings schon ein relativ spezieller Fall, der in vielen Prozessen gar nicht und wenn doch, dann nur sehr vereinzelt auftreten wird. Daher wäre das Potenzial der dritten Farbe in vielen Prozessen gewissermaßen verschenkt.

Anzahl der Kategorien

Neben den Definitionen der drei verwendeten Farbkategorien ist natürlich auch die Anzahl und Auswahl der Farben diskussionswürdig. Generell liegen einige Arbeiten vor, die für das Ampelschema sprechen (vgl. Abschnitt 4.1 und Abschnitt 6.3). Möglicherweise ist aber die Einführung einer vierten Farbe (z. B. blau oder grau) sinnvoll. Diese könnte etwa für Modellelemente verwendet werden, deren Kategorie unklar ist. Oder die Definitionen der Farben könnten insgesamt angepasst werden. So könnte etwa die zusätzliche Farbe für Aufgaben verwendet werden, die keine Datenverarbeitung beinhalten, also bisher grün kategorisierte. Grün wäre die neue Farbe für bisher gelbe

Aufgaben und gelb für bisher rote Aufgaben. Die frei gewordene Farbe Rot könnte dann z. B. entsprechend eines oben beschriebenen Ansatzes, für komplett unzulässige Datenverarbeitungen (z. B. nach Widerruf einer Einwilligung) verwendet werden.

Fazit

Wegen der genannten Kritikpunkte an den anderen Ansätzen fiel die Entscheidung auf die zu Beginn des Kapitels beschriebenen Definitionen. Um sicherzustellen, dass dieser aber auch tatsächlich gut verständlich und insgesamt sinnvoll ist, wurde in durchgeführten Experteninterviews explizit nachgefragt, ob eine andere Menge oder Definition der Farben sinnvoller wäre.

8.3. Ereignisse

Ereignisse (siehe Abschnitt 3.1.1) müssen etwas differenzierter betrachtet werden. Grundsätzlich haben diese wenig Auswirkungen auf den Datenschutz, da sie in der Regel weder ein Datum als solches noch eine Verarbeitungstätigkeit abbilden.

Eine Ausnahme bilden hier allerdings Nachrichtenereignisse. Diese bilden, wie der Name schon sagt, den Versand bzw. den Empfang von Nachrichten eines Prozessteilnehmers, genauer gesagt eines Pools, an einen anderen ab. Dies kann natürlich durchaus relevant für den Datenschutz sein. Im Wesentlichen können Nachrichtenereignisse im Kontext der Datenschutzklassifizierung aber mit Aufgaben gleichgesetzt werden, da der Nachrichtenversand bzw. -empfang letztlich eine Teilmenge der Aufgaben innerhalb eines Prozesses darstellt (siehe dazu auch Abschnitt 3.1.1). Dementsprechend werden die Regeln aus Abschnitt 8.2 entsprechend übernommen.

8.4. Weitere Modellelemente

Neben den oben genannten Aktivitäten, Ereignissen und Datenobjekten bzw. -speichern, sieht der BPMN-Standard noch einige weitere Modellelemente vor (siehe Abschnitt 3.1), deren Relevanz für den Datenschutz an dieser Stelle kurz diskutiert wird.

8.4.1. Pools/Lanes

In den Grundlagen zum Datenschutzrecht wurde bereits dargelegt, dass die DSGVO zwischen unterschiedlichen Rechtssubjekten unterscheidet (siehe Abschnitt 2.1.5). Und das Subjekt eines Prozessschrittes wird durch den Pool bzw. die Lane abgebildet, in dem

sich eine Aufgabe befindet. Somit liegt es also nahe, dass bei der Betrachtung des Datenschutzes im Geschäftsprozessmodell, die Pools/Lanes die vertretenen Rechtssubjekte darstellen. In der Praxis ist es tatsächlich so, dass in einem Prozessmodell mindestens ein Pool für den *Verantwortlichen* vorhanden ist. Dazu kommt häufig noch ein Pool für den *Betroffenen* (siehe dazu auch die Beispiele in Kapitel 7). In einigen Fällen wird zusätzlich auch noch ein *Auftragsverarbeiter*, häufig als geschlossener Pool, dargestellt. Außerdem können auch *Dritte* und *Empfänger* im Sinne der DSGVO modelliert sein. Empfänger sind hier in der Regel alle Pools, die Nachrichten von einem anderen Pool enthalten, welche personenbezogene Nachrichten enthalten. Und Dritte sind entsprechend alle Empfänger-Pools, die nicht dem Betroffenen, dem Verantwortlichen oder einem Auftragsverarbeiter entsprechen.

Möglicher Ansatz

All diese Teilnehmer eines Prozesses haben offensichtlich eine unterschiedliche Relevanz für den Datenschutz. Rein theoretisch wäre es also beispielsweise denkbar, den Pool des Betroffenen grün einzufärben, da dieser ja gerade die schützenswerte Person ist und daher eigentlich keine Datenschutzverstöße begehen kann. Der Verantwortliche und eventuell vorhandene Auftragsverarbeiter könnten dann gelb markiert werden, da hier sehr wohl relevante Datenverarbeitungen stattfinden und alle Vorgänge genauer betrachtet werden sollten. Für Dritte käme eventuell die Farbe Rot in Frage, da die Weitergabe personenbezogener Daten an Dritte genau geprüft werden muss und oft schwierig zu begründen ist.

Problematik

Diese Einfärbung birgt aber die Gefahr, die eigentlich wichtigeren Aspekte aus Abschnitt 8.1, 8.2 und 8.3 visuell zu überlagern und damit den Fokus negativ zu beeinflussen. Außerdem ist es durchaus denkbar, dass ein Pool (oder eine Lane) im Verlauf des Prozesses nicht durchgehend die gleiche Datenschutzrechtliche Rolle einnimmt. So könnte es etwa sein, dass ein Teilnehmer zeitweise als Auftragsverarbeiter fungiert, zeitweise aber auch als Dritter, weil möglicherweise eine Datenverarbeitung stattfindet, die nicht vom Auftragsverarbeitungsvertrag abgedeckt ist. Besonders auffällig ist dies bei der Rolle des Empfängers. Hier ist es nicht unwahrscheinlich, dass im Verlauf eines Prozesses jeder Teilnehmer diese Rolle temporär einnimmt. Abschnitt 9.3.1 zeigt ein Beispiel, in dem ein Pool, der eigentlich durchgehend den Verantwortlichen abbildet, in einer Aktivität sogar den Betroffenen darstellt.

Umgang mit der Problematik

Somit fällt die Entscheidung, die Rechtssubjekte in einem Prozessmodell nicht farblich zu kennzeichnen. Allerdings sollten diese trotzdem betrachtet werden. Mögliche Ansätze hierfür werden in Abschnitt 12.3 betrachtet.

8.4.2. Gateways

Gateways scheinen zunächst eher keine Implikationen für den Datenschutz zu haben. Allerdings könnten in einem konkreten Modell Sachverhalte so abgebildet werden, dass dem Gateway doch eine gewisse Relevanz zugeschrieben werden kann. Beispielsweise könnte mit einem exklusiven Gateway abgefragt werden, ob eine Einwilligung vorliegt. Oder es könnte beispielsweise mit einem inklusiven Gateway abgefragt werden, ob für eine gewisse Verarbeitungstätigkeit überhaupt irgendein Verarbeitungsgrund vorliegt. In diesen Fällen ist allerdings nicht wirklich das Gateway als solches, sondern es sind eher die ausgehenden Kontrollflüsse interessant. Eine Einfärbung oder sonstige Kenntlichmachung ist daher nicht wirklich hilfreich. Stattdessen bietet es sich aber an, ein solches Vorgehen als eine Art Prozessmuster zu etablieren, ähnlich dem Vorgehen in [Ag19], [BF20], oder auch [Ni22] im Kontext des Versicherungsrechts. Ein entsprechendes Muster könnte dann etwa so aussehen wie in Abbildung 8.4.

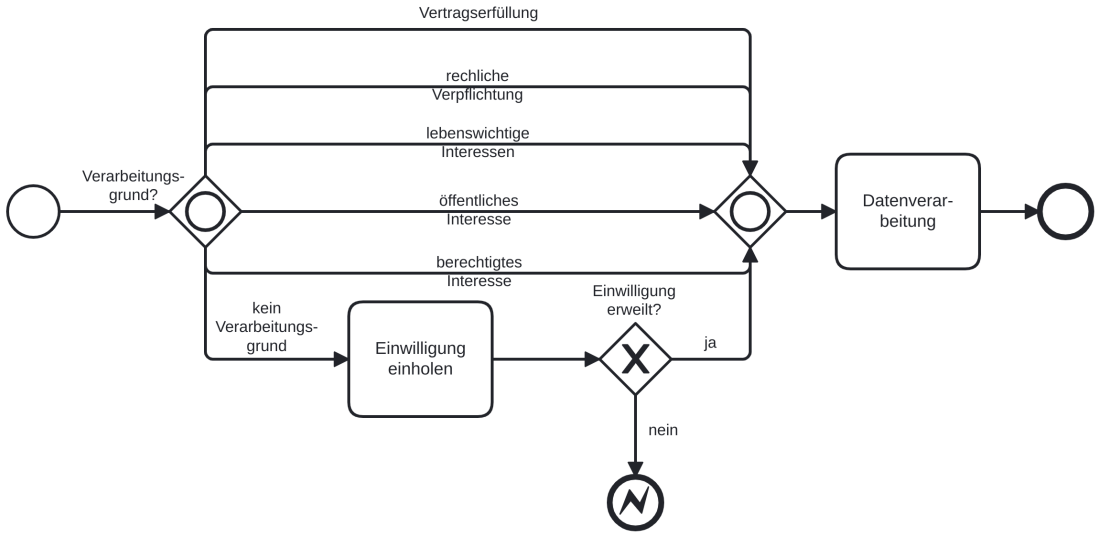


Abbildung 8.4.: Prozessmuster für die Abfrage eines Verarbeitungsgrundes

8.4.3. Konnektoren

Wie oben zum Thema Gateways erwähnt, können Kontrollflüsse durchaus eine gewisse Relevanz für den Datenschutz haben. Im oben erläuterten Beispiel zur Auswahl eines Verarbeitungsgrunds könnte beispielsweise der ausgehende Kontrollfluss für „kein Verarbeitungsgrund“ rot, die Einwilligung als ungünstigste Variante (siehe Abschnitt 2.1.4) gelb und alle anderen Verarbeitungsgründe grün eingefärbt werden.

9. Anwendung des Konzepts auf die Beispiele

In diesem Kapitel soll das in Kapitel 8 eingeführte Konzept anhand einiger Beispiele weiter verdeutlicht werden. Die hierfür verwendeten Prozesse wurden bereits in Kapitel 7 vorgestellt.

9.1. Personalwesen: Einstellung

Abbildung 9.1 zeigt das Geschäftsprozessmodell aus Abbildung 7.1, welches aber entsprechend des vorgestellten Konzeptes eingefärbt wurde.

Die Pools der Sozialversicherung und des Finanzamtes können hier vernachlässigt werden, da sie nur als Black-Box abgebildet sind und somit keine Notationselemente enthalten, welche eingefärbt werden könnten.

Besonders interessant ist der Pool des Arbeitgebers. Dieser ist für alle abgebildeten Datenverarbeitungen der Verantwortliche. Zunächst finden sich in diesem Pool zwei gelb eingefärbte Aufgaben, in welchen also nach der Definition aus Abschnitt 8.2 personenbezogene Daten verarbeitet werden, wofür aber keine Einwilligung benötigt wird, da eine andere Grundlage vorliegt. Bei der Vorbereitung des Arbeitsvertrages wird beispielsweise der Name und das Geburtsdatum des Arbeitnehmers in den Vertrag einfügt. Dies ist offensichtlich eine Verarbeitung personenbezogener Daten. Diese Daten können aber als notwendig „zur Durchführung vorvertraglicher Maßnahmen“ (Art. 6 (1) b) angesehen werden, womit die Verarbeitung ohne Einwilligung rechtmäßig ist. Ähnlich verhält es sich bei der zweiten Aufgabe „Dokumente zum Arbeitnehmer senden“. Auch hier wird beispielsweise der Name, aber auch die Adresse des Arbeitnehmers verarbeitet, die Verarbeitung ist aber natürlich aus dem gleichen Grund legitim. Die folgende Aufgabe, „Dokumente entgegennehmen“, ist grün eingefärbt. Dies ist allerdings durchaus diskussionswürdig. Gemeint ist hier die reine Annahme eines verschlossenen Umschlags. Dabei werden im Grunde keine personenbezogenen Daten verarbeitet. Die Aufgabe könnte aber durchaus auch so verstanden werden, dass hier beispielsweise der Inhalt auch gelesen und vielleicht sogar in ein elektronisches System übertragen wird.

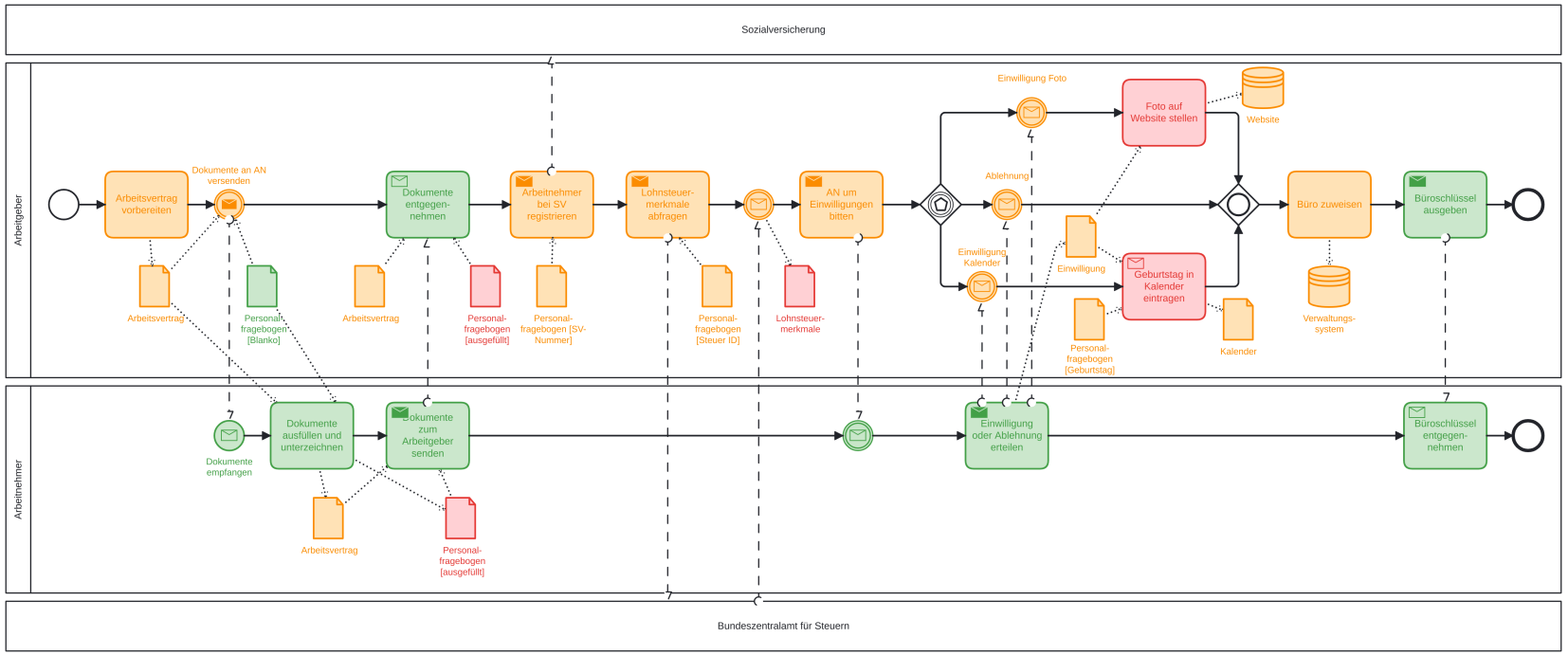


Abbildung 9.1.: Eingefärbtes Geschäftsprozessmodell zur Einstellung eines Mitarbeiters

Dann wäre die Aufgabe selbstverständlich gelb zu färben, da dann eine (notwendige) Datenverarbeitung vorläge. Gelb sind auch die folgenden drei Aufgaben, bei welchen offensichtlich Personenbezogene Daten (insbesondere Sozialversicherungsnummer, Religion, Steuer-ID) verarbeitet werden, bei welchen aber mit der rechtlichen Verpflichtung zur Anmeldung bei Sozialversicherung und Finanzamt auch eindeutig ein Grund für die Verarbeitung vorliegt (Art. 6 (1) c) DSGVO in Verbindung mit §28a (1) 1. SGB IV und §39e (4), Satz 2 EStG). Es folgen zwei rot eingefärbte Aufgaben, die die Eintragung des Geburtstages und das Hochladen eines Fotos betreffen. Hier werden offensichtlich personenbezogene Daten, nämlich Name, Geburtsdatum und Lichtbild, verarbeitet. Allerdings kann hierfür keine Rechtsgrundlage außer einer Einwilligung gefunden werden. Genau dieser Zusammenhang entspricht der Definition einer roten Aufgabe (siehe Abschnitt 8.2). Die Zuweisung eines Büros und Ausgabe eines Schlüssels sind grün, da hier zunächst keine Daten verarbeitet werden. Erst in der letzten Aufgabe des Arbeitgebers, der Eintragung der Bürozuweisung wird wieder ein personenbezogenes Datum (der Name) verarbeitet. Dies kann wohl durchaus als notwendig betrachtet und daher gelb eingefärbt werden.

Im Pool des Arbeitnehmers sind alle Aufgaben grün eingefärbt. Zwar werden hier durchaus personenbezogene Daten verarbeitet, allerdings nur die des Arbeitnehmers selbst. Dies ist datenschutzrechtlich nicht relevant. Darüber hinaus agiert der (zukünftige) Arbeitnehmer während des Prozesses noch als Privatperson und nicht im Auftrag des Unternehmens, weshalb er nicht der DSGVO unterliegt (vgl. Art. 3 (2) Buchst. c)).

Neben den Aufgaben enthält das Prozessmodell auch einige Datenobjekte. Zunächst ist dies der Arbeitsvertrag. Dieser ist gelb eingefärbt, da er, wie weiter oben in diesem Abschnitt erläutert, grundlegende personenbezogene Daten enthält. Interessant ist der Personalfragebogen. Dieser ist zu beginn grün eingefärbt, da er hier – erkennbar an dem Zusatz [blanko] – noch keine Daten enthält. Im Verlauf des Prozesses wird der Personalfragebogen ausgefüllt und wechselt damit die Farbe zu rot, da nun die Religion enthalten ist und diese zu den besonderen Kategorien personenbezogener Daten zählt. Zusätzlich existieren noch Datenobjekte für den (Geburtstags-)Kalender und die Einwilligung des Arbeitnehmers zur Aufnahme des Geburtstags in den Kalender und das Hochladen des Fotos auf die Website. Beide Datenobjekte enthalten Namen und Geburtsdatum und werden daher gelb eingefärbt.

Letztlich finden sich im Prozess auch noch zwei gelb eingefärbte Datenspeicher. Die Website enthält hier offensichtlich Lichtbilder der Mitarbeiter und somit personenbezogene Daten, es sollten aber eigentlich keine Daten der besonderen Kategorien zu finden sein[1]. Die Kategorisierung des Verwaltungssystems ist nicht ganz eindeutig. Grundsätzlich kann hier wohl davon ausgegangen werden, dass personenbezogene Daten,

[1] Lichtbilder zählen nach Erwägungsgrund 51 der DSGVO nicht zu den besonderen Kategorien personenbezogener Daten.

mindestens der Name des Mitarbeiters, verarbeitet werden. Rein theoretisch wäre es auch durchaus denkbar, dass Daten der besonderen Kategorien verarbeitet werden. Ein Beispiel hierfür könnte etwa eine mögliche Gehbehinderung sein, um zu begründen, dass dem Mitarbeiter ein bestimmtes barrierefreies Büro zugeteilt wird. Außerdem ist das Verwaltungssystem nicht näher definiert und es wäre möglich, dass es sich hier um ein generelles Personalverwaltungssystem handelt, in welchem alle Daten der Mitarbeiter und damit beispielsweise Religion, Schwerbehinderungen, Kranheitszeiten, etc. enthalten sind.

9.2. Gesundheitswesen: Zahnarztbesuch

Das Prozessmodell aus Abschnitt 7.2 ist in der eingefärbten Variante in Abbildung 9.2 zu sehen.

In der Lane des Patienten fällt wie im obigen Beispiel (siehe Abschnitt 9.1) beim Arbeitnehmer wieder auf, dass alle Aufgaben grün eingefärbt sind. Der Grund ist auch hier wieder die offensichtlich unkritische Verarbeitung der eigenen Daten. Die Einwilligung ist an dieser Stelle für die Weitergabe der Daten an die externe Verrechnungsstelle notwendig.

Die Aufgaben des medizinischen Fachangestellten sind in zwei Fällen gelb klassifiziert, da hier zwar personenbezogene Daten verarbeitet werden, die Verarbeitungen aber für die Behandlung zwingend erforderlich sind. Einzig die Aufgabe „Daten an Abrechnungsstelle weitergeben“ ist rot klassifiziert.

Hintergrund ist die Beauftragung einer externen Stelle für die Forderungsverwaltung durch die Praxis. Hierfür ist in aller Regel eine Einwilligung einzuholen (siehe [Sc18]).

Beim Arzt hingegen wechseln sich gelb und grün klassifizierte Aufgaben ab. Die beiden grün klassifizierten sind hierbei durchaus diskussionswürdig. Natürlich nimmt der Arzt schon während der Untersuchung ständig Daten über die Zähne seines Patienten auf und verarbeitet diese auch, indem er beispielsweise bewertet inwiefern eine Behandlung notwendig ist. Ähnliches gilt wohl auch während der eigentlichen Behandlung. Im vorliegenden Prozessmodell folgt allerdings jeweils eine weitere Aufgabe auf die beiden grünen Aufgaben, welche sich explizit mit der weiteren Verarbeitung der Daten befassen. Hiermit sollte die Datenverarbeitung für den praktischen Einsatz hinreichend gut repräsentiert sein. Im Übrigen braucht der Arzt zumindest für die Behandlung unter Umständen zwar durchaus auch eine Einwilligung, diese hat allerdings nichts mit der Einwilligung im datenschutzrechtlichen Sinne zu tun und wird daher nicht betrachtet.

Im Pool der Abrechnungsstelle letztlich, fallen fast alle Aufgaben in die gelbe Kategorie, da hier offensichtlich personenbezogene Daten verarbeitet werden.

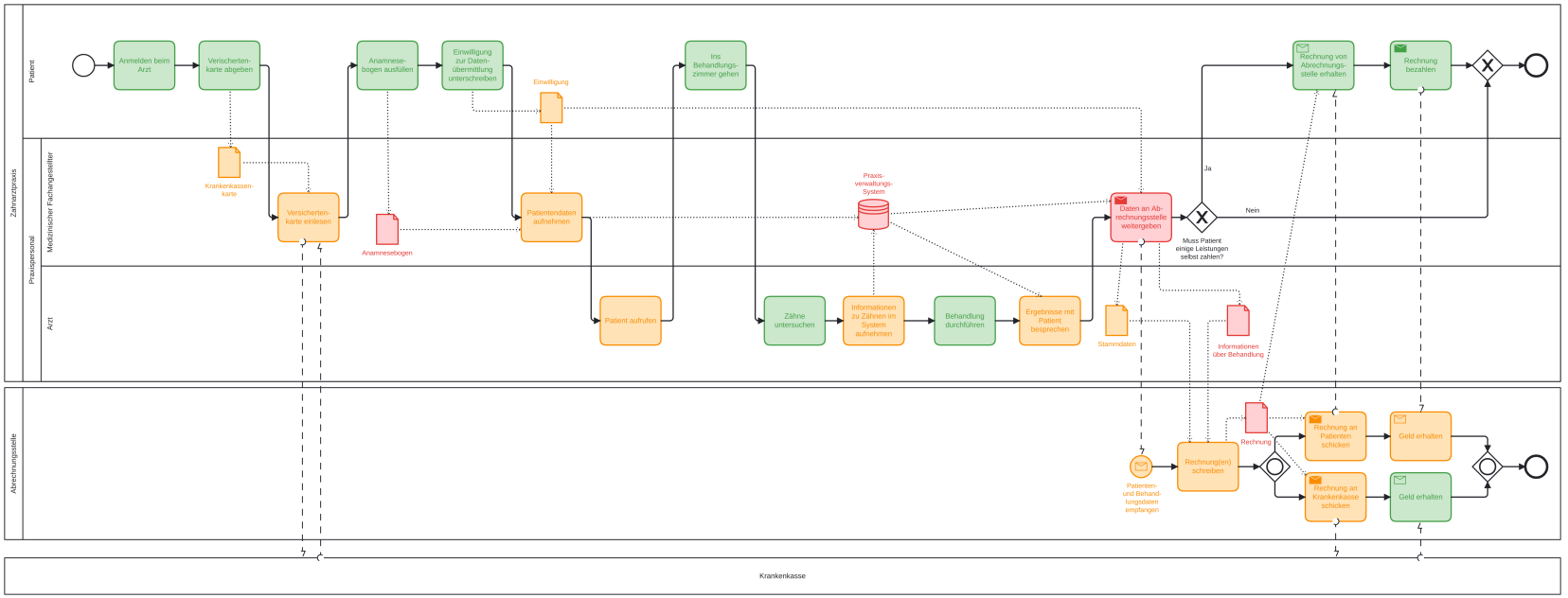


Abbildung 9.2.: Eingefärbtes Geschäftsprozessmodell eines Zahnarztbesuchs

Interessant ist hier, dass eine Aufgabe, welche mit der gleichen Bezeichnung (*Geld erhalten*) zweimal vorkommt, in einem Fall gelb und im anderen Fall grün eingefärbt ist. Dies liegt daran, dass die Abrechnungsstelle im einen Fall das Geld von der Krankenkasse und im anderen Fall direkt vom Patienten erhält (nämlich bei Leistungen, die nicht von der KK bezahlt werden). Bei der Zahlung wird in der Regel einerseits ein Verwendungszweck verarbeitet. Dies ist gewöhnlich eine Rechnungsnummer, die allein betrachtet eigentlich keine Rückschlüsse auf den Patienten zulassen sollte und somit kein personenbezogenes Datum darstellt. Außerdem werden natürlich Kontodaten des Zahlenden verarbeitet. Bei Zahlung des Patienten handelt es sich hier eindeutig um personenbezogene Daten. Bei Zahlung durch die KK wird aber deren Kontoverbindung übertragen, die keinen Personenbezug hat. Somit wird die „gleiche“ Aufgabe nur in Abhängigkeit vom beteiligten Prozessteilnehmer unterschiedlich kategorisiert. In der Realität ist es natürlich durchaus denkbar, dass die Rechnungsnummer Personenbezug hat oder beispielsweise der Name des Patienten mit im Verwendungszweck angegeben wird. Das Beispiel sollte aber den beschriebenen Sachverhalt verdeutlichen und lässt diesen Umstand daher außen vor.

Die im Prozess anfallenden Datenobjekte sind gelb oder rot. Die Krankenkassenkarte und die Einwilligung zur Datenübermittlung an die Abrechnungsstelle und die Stammdaten für selbige sind gelb klassifiziert. Hier handelt es sich im Wesentlichen um eben einige Stammdaten, wie Name, Geburtsdatum und Adresse, sowie die Versichertennummer des Patienten. Im Anamnesebogen, den Behandlungsdaten und der Rechnung sind medizinische Daten enthalten. Bei den ersten beiden Datenobjekten ist dies recht offensichtlich. Bei der Rechnung wird aber in der Regel zumindest eine Kennung für die Behandlung angegeben, welche auch als medizinisches Datum betrachtet werden kann. Das Praxisverwaltungssystem ist auch rot klassifiziert, da natürlich auch hier die sensiblen medizinischen Daten der Patienten enthalten sind.

9.3. (Öffentliche) Verwaltung: Erstellung eines Studentenausweises

Das Hauptprozessmodell zur Erstellung einer CAU-Card besteht nur aus zwei Teilprozessen. Diese werden beide gelb eingefärbt. Dies ergibt sich allerdings nicht direkt, sondern nur aus den Aktivitäten innerhalb der Teilprozesse. Daher müssen zunächst diese betrachtet werden. Es sei vorweggenommen, dass in beiden Teilprozessen nur grün und gelb kategorisierte Aktivitäten zu finden sind. Nach der Definition aus Abschnitt 8.2 werden die Teilprozesse als ganzes damit als gelb kategorisiert, da dies die kritischste aller vorhandener Kategorien ist.

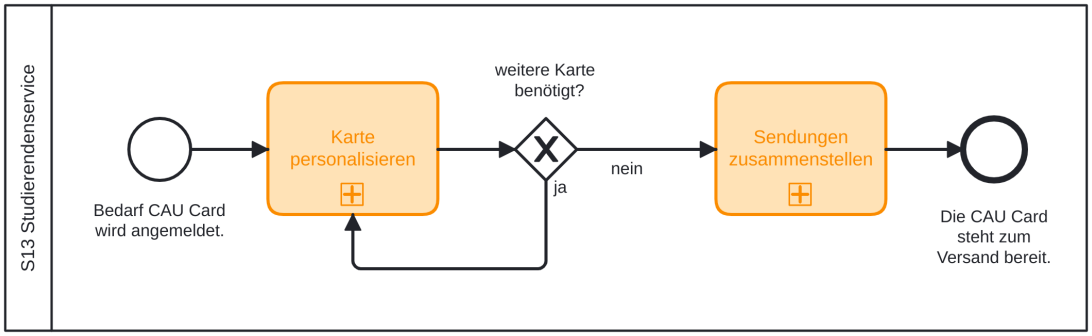


Abbildung 9.3.: Eingefärbtes Geschäftsprozessmodell zur Erstellung eines Studentenausweises

9.3.1. Karte personalisieren

Das Personalisieren der Karte ist zwar recht umfangreich, aus Datenschutzsicht aber relativ übersichtlich. Zwei Aufgaben sind hier grün eingefärbt: Einerseits ist dies das Anmelden an der Arbeitsstation und Starten der Anwendung. Das Starten der Anwendung ist hierbei wohl eindeutig nicht datenschutzrelevant. Beim Anmelden an der Arbeitsstation ist die Bewertung von einigen Faktoren abhängig. Wenn der Benutzername beispielsweise dem Realnamen des Anwenders entspräche, wäre dies ein personenbezogenes Datum. Außerdem könnte es durchaus sein, dass es sich hier um eine Remote-Desktop-Verbindung handelt und möglicherweise auch ein Fernzugriff aus dem Home-Office möglich wäre. Dann käme die übertragene IP-Adresse des Anwenders als personenbezogenes Datum in Frage. Somit käme auch eine gelbe Einfärbung in Frage. Interessanterweise wäre an dieser Stelle dann der ausführende Mitarbeiter auch gleich der Betroffene. Es wird hier nun aber davon ausgegangen, dass es sich um eine lokale Anmeldung handelt und die Benutzerkennung allein keine Rückschlüsse auf den Benutzer zulässt. Die andere grün eingefärbte Aufgabe ist das Senden des Dokuments an den Drucker kurz vor Ende des Prozesses. Auch hier wäre aus ähnlichen Gründen wie oben wieder eine Verarbeitung personenbezogener Daten möglich, wird hier aber nicht angenommen.

Die restlichen Aufgaben des Prozessmodells sind gelb eingefärbt. Im Wesentlichen werden hier das Foto und/oder weitere Daten des Studenten verarbeitet. Interessant ist hierbei die Aufgabe „Datensatz/ Task löschen“. Einerseits ist nicht ganz klar, was genau hier mit „Datensatz/ Task“ gemeint ist, es kann aber wohl davon ausgegangen werden, dass hier personenbezogene Daten enthalten sind. Andererseits wirkt das Löschen intuitiv vielleicht nicht datenschutzrelevant bzw. sogar förderlich. Nichtsdestotrotz ist das Löschen von personenbezogenen Daten laut Art. 4 (2) DSGVO eine Verarbeitung für die alle weiteren Vorschriften genauso gelten, wie für jede andere Verarbeitungstätigkeit.

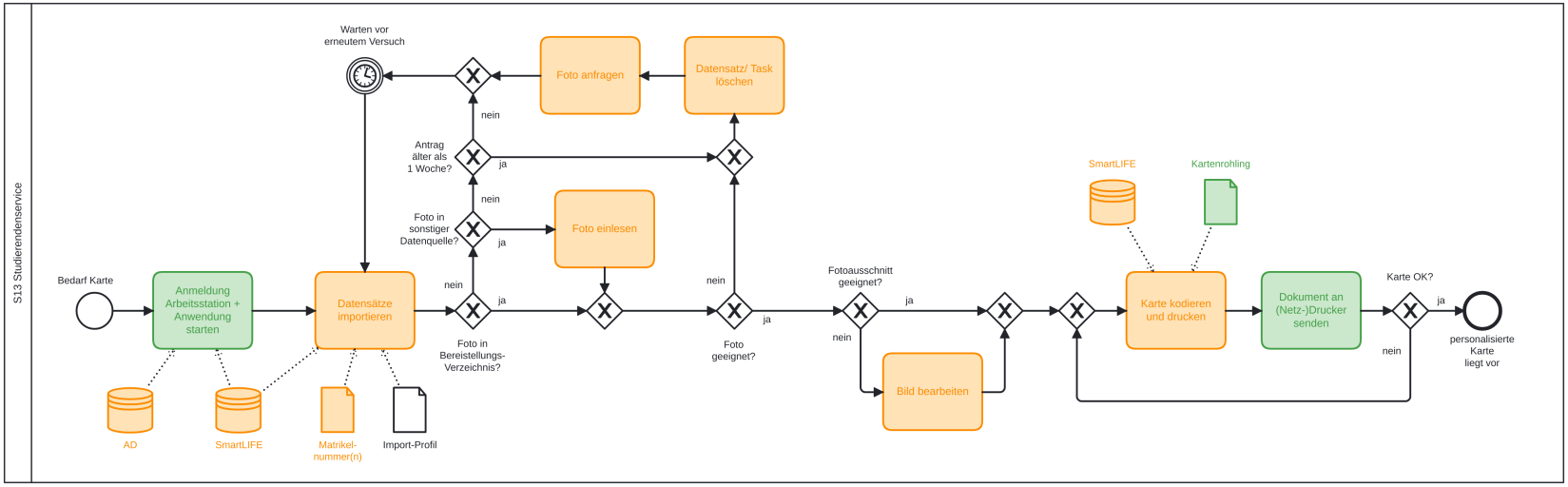


Abbildung 9.4.: Eingefärbtes Geschäftsprozessmodell „Karte personalisieren“

Im Prozessmodell sind zwei verschiedene Datenspeicher zu finden. Mit „AD“ wird hier ein Active Directory bezeichnet, also ein Verzeichnisdienst, der unter anderem für die Benutzerverwaltung verwendet werden kann. Hier werden von den jeweiligen Benutzern auch die Klarnamen gespeichert. Daher ist der Datenspeicher gelb gefärbt. „SmartLIFE“ bezeichnet die Datenbank *smart.LIFE*, welche Kunden der Firma Intercard die Validierung der Chipkarten ermöglicht[In23a]. Hier werden auch personenbezogene Daten, mindestens in Form des Namens, der Studenten gespeichert. Somit resultiert auch hier eine gelbe Färbung.

Neben den Datenspeichern finden sich auch noch drei Datenobjekte. Ein Datenobjekt ist hierbei mit „Matrikelnummern“ bezeichnet. Wenn hier wirklich nur Matrikelnummern ohne weitere Informationen enthalten sind, ist es wohl zumindest diskussionswürdig, ob es sich um personenbezogene Daten handelt. Hier ist aber wohl davon auszugehen, dass auch Namen und evtl. weitere Daten zu den Matrikelnummern enthalten sind, weshalb das Datenobjekt gelb eingefärbt wird. An der gleichen Aufgabe findet sich auch noch ein Datenobjekt „Import-Profil“. Hier ist allerdings völlig unklar, wobei es sich dabei handelt, weshalb keine Färbung vorgenommen wurde. Letztlich existiert am Ende des Prozessmodells noch ein Datenobjekt für den Kartenrohling. Hier handelt es sich offensichtlich nur um die reine Karte ohne aufgespielte Daten. Daher kann dieses Datenobjekt grün gefärbt werden.

9.3.2. Sendung zusammenstellen

Das Zusammenstellen der Sendung ist ebenfalls recht unkritisch in Bezug auf den Datenschutz. Die meisten Aktivitäten können hier grün eingefärbt werden. Bei der „DocuLounge Anmeldung“ besteht natürlich wieder die Frage nach den notwendigen Benutzerdaten. Der Erfahrung nach könnte hier aber sogar ein nicht personenbezogener Account für die ganze Abteilung vorliegen, weshalb nicht von der Verwendung personenbezogener Daten ausgegangen wird. Beim Entnehmen der Anschreiben aus dem Drucker, stehen natürlich durchaus personenbezogene Daten auf dem Anschreiben, welche theoretisch von der ausführenden Person gelesen werden könnten, es ist hier aber wohl trotzdem nicht von einer relevanten Datenverarbeitung auszugehen, da diese eigentlich nicht Gegenstand der Aufgabe ist. Selbiges gilt in der Aufgabe „Anschreiben und Beiblatt in den Umschlag geben und verschließen“. Die einzige gelbe Aktivität im Prozessmodell stellt somit der Teilprozess „Karte aufspenden“ dar, dessen Kategorie sich aus der Bewertung im folgenden Abschnitt 9.3.3 ergibt.

Bei der Betrachtung der Daten ist die Mehrheit der Modellelemente ebenfalls grün eingefärbt. Auch hier stellt sich bei zwei Elementen („DocuLounge-Server“ und „Chipkarte oder AD Zugangsdaten“) wieder die Frage nach eventuell enthaltenen personenbezogenen Daten. Entsprechend der oben genannten Begründung wird dies hier aber auch

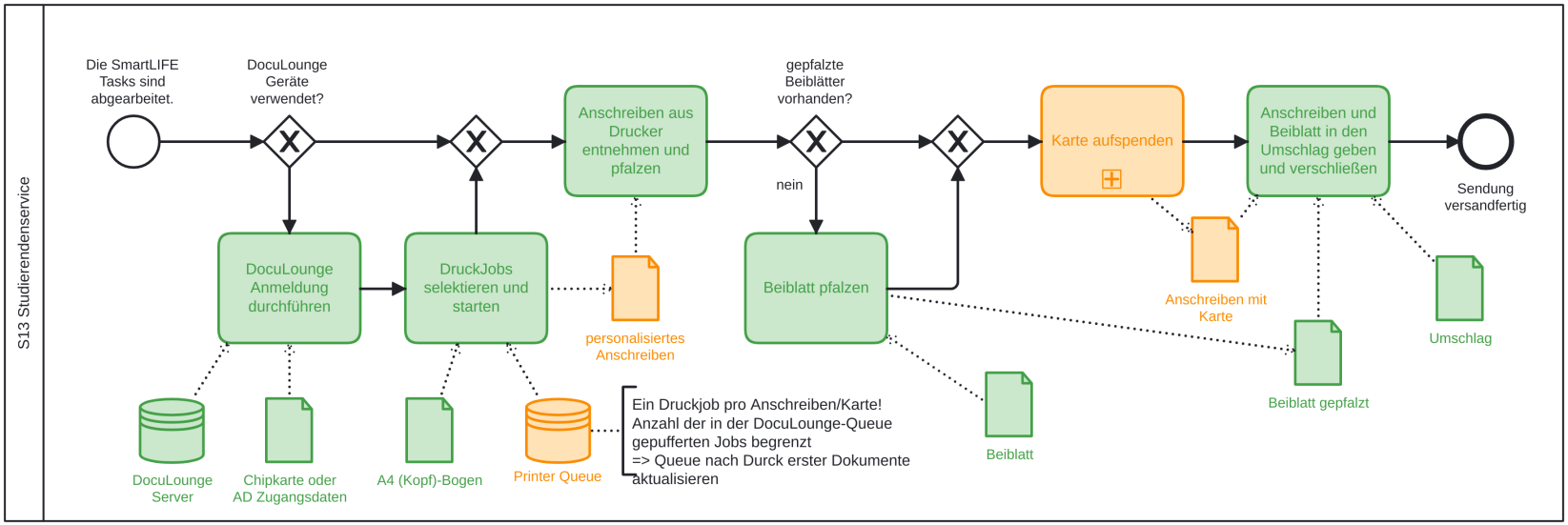


Abbildung 9.5.: Eingefärbtes Geschäftsprozessmodell „Sendung zusammenstellen“

nicht angenommen. Neben den grün eingefärbten Datenobjekten und -speichern existieren auch noch einige Gelbe. Zunächst ist dies die „Printer Queue“. Hier sind wohl die fertigen Anschreiben inklusive. Namen und Adressen der Studenten enthalten. Als Datenobjekte wird zwei mal das Anschreiben verwendet, einmal hiervon in Verbindung mit der Karte. Auch hier finden sich offensichtlich diese Daten.

9.3.3. Karte aufspenden

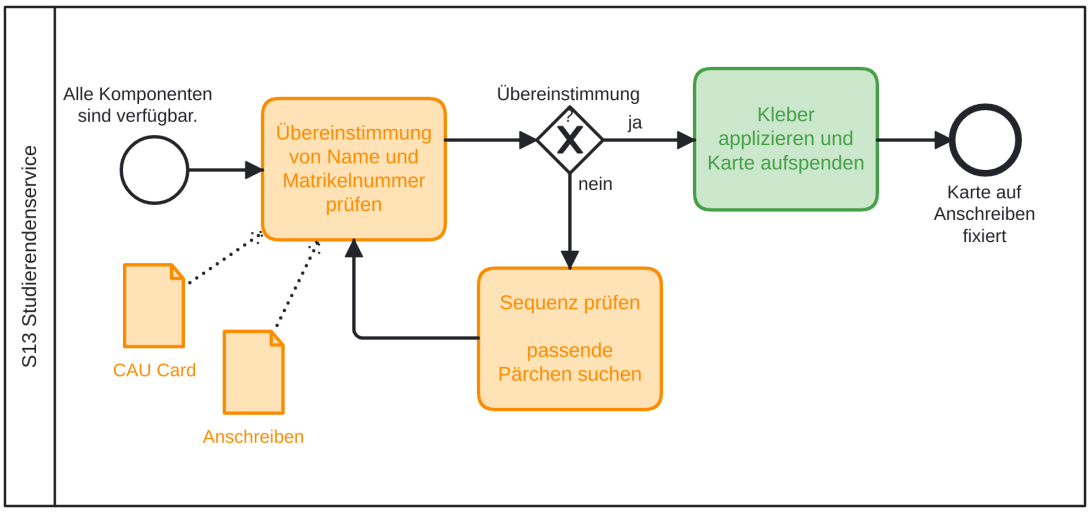


Abbildung 9.6.: Eingefärbtes Geschäftsprozessmodell „Karte aufspenden“

Das (Teil-)Prozessmodell zum Karte aufspenden ist wieder sehr kurz. Zwei der drei Aktivitäten werden gelb kategorisiert, da hier Namen und Matrikelnummern von Studenten verarbeitet werden. Einzig die Aktivität „Kleber applizieren und Karte aufspenden“ ist grün eingefärbt, da hier nicht unmittelbar eine Datenverarbeitung stattfindet.

Beide im Prozess verwendete Datenobjekte (CAU Card und Anschreiben) sind ebenfalls gelb eingefärbt, da hier offensichtlich Daten des Studenten enthalten sind.

9.4. Fazit

Die Anwendung des Konzepts in hat gezeigt, dass es eine große Problematik bei der Bewertung des Datenschutzes auf Basis von Geschäftsprozessmodellen gibt: Die mehrdeutige Modellierung vieler Prozesse. Oft wird für einen externen Betrachter nicht wirklich klar, ob ein Datenobjekt personenbezogene Daten enthält oder ob in einer Aufgabe solche verarbeitet werden. Eine mögliche Lösung des Problems wäre natürlich eine eindeutigere und verständlichere Modellierung. Dies ist aber schwierig umsetzbar,

da sich in Unternehmen häufig ein spezielles Wording entwickelt und dies teilweise sogar im Rahmen einer Corporate Language als Teil der Corporate Identity gewünscht ist[Sc15]. Somit wird es für externe immer schwierig sein, die Bedeutung eines jeden Begriffes und vor allem einer jeden Abkürzung zu verstehen. Allerdings kann der vorgestellte Ansatz hier unterstützen, da hiermit die Kommunikation zwischen Unternehmen und etwa einem externen Datenschutzbeauftragten oder auch einer Aufsichtsbehörde vereinfacht werden kann.

10. Evaluation des Konzepts

Ein zentraler Aspekt der Design Science Mehodik ist die Evaluation des eigenen Vorgehens und der erstellten Artefakte. In diesem Kapitel soll daher das in Kapitel 8 vorgestellte Konzept evaluiert werden.

Bisher gibt es wenig Arbeiten, in denen die Visualisierung von Recht in Prozessmodellen empirisch evaluiert wird [Ni22]. Die generelle Evaluation von konzeptuellen Modellen ist in der Wirtschaftsinformatik aber weit verbreitet [GW04]. Daher richtet sich das Vorgehen an diesen Methoden aus.

Zur Evaluation des beschriebenen Konzepts wurde ein zweistufiges Vorgehen gewählt. Zunächst wurden einige Experten aus dem Datenschutzbereich um ein Feedback und mögliche Verbesserungsvorschläge gebeten (siehe Abschnitt 10.1). Anschließend wurde ein Vergleichsexperiment unter Studenten durchgeführt, bei dem getestet wurde, ob Datenschutzprobleme in Prozessen, die nach dem beschriebenen Vorgehen eingefärbt wurden, einfacher gefunden werden, als in ungefärbten Prozessmodellen (siehe Abschnitt 10.2).

Das komplette Vorgehen wurde auch in der im Rahmen der Dissertation betreuten Masterarbeit von Carolin Goppelt[Go23] beschrieben. Hier finden sich unter anderem nähere Informationen zur Wahl der Methodik.

In Abschnitt 10.3 werden die Ergebnisse zusammengefasst und Schlussfolgerungen für das weitere Vorgehen abgeleitet.

10.1. Expertenbefragung

Für eine erste Einschätzung der Qualität des vorgestellten Konzepts wurden einige Personen befragt, die beruflich auf verschiedene Arten Kontakt mit dem Thema Datenschutz haben. Die Befragungen wurden in Form eines leitfadengesteuerten Interviews durchgeführt. Ziel ist zum einen ein allgemeines Feedback zum Konzept, zum anderen aber auch mögliche Verbesserungsvorschläge.

10.1.1. Vorgehen

Zu den allgemeinen Rahmenbedingungen sei gesagt, dass alle Interviews als Videokonferenz stattfanden. Teilgenommen haben jeweils der Interviewte, Carolin Goppelt

als Befragende und die Autorin dieser Arbeit als Moderatorin, welche aber auch an einzelnen Stellen tiefer gehende Fragen stellte.

Als Gesprächsgrundlage diente in allen Interviews ein Prozessmodell, welches dem Beispiel aus Abschnitt 7.1 bzw. Abschnitt 9.1 ähnelt. Der Prozess wurde jeweils zunächst in der ungefärbten Version erläutert. Anschließend wurde die gefärbte Version gezeigt und anhand dessen der Ansatz erklärt. Daraufhin startete das eigentliche Interview mit der allgemeinen Frage nach der Meinung des Interviewten zu dem vorgestellten Ansatz. Aus den Antworten ergaben sich dann verschiedene Anschlussfragen. Zusätzlich wurden noch einige Themengebiete gezielt angesprochen:

- Auswahl der Farben
- Definition der Farben
- Einwilligung
- Anonymisierung
- Verbesserungs- und Erweiterungsvorschläge
- Automatisierung

Einerseits wurde hier die Auswahl der Farben Grün, Gelb und Rot (und Weiß bzw. ungefärbt) angesprochen und abgefragt, ob eine andere Farbwahl sinnvoller wäre oder ob eine zusätzliche Farbe eingeführt werden sollte. In diesem Zusammenhang wurde auch die konkrete Definition der Farben, wie sie in Kapitel 8 eingeführt wird, diskutiert. Ein Fokus lag hier auf dem Umgang mit Einwilligungen und Anonymisierung.

Andererseits wurden aber auch allgemeine Verbesserungs- und Erweiterungsvorschläge und Ideen für die Automatisierung der Färbung besprochen.

Im Folgenden werden die Interviewten Personen kurz vorgestellt und die Ergebnisse der Interviews erst einzeln dargestellt und anschließend zusammengefasst.

10.1.2. Isabelle Puttrus

Die erste Interviewpartnerin hat Wirtschaftsinformatik studiert und sich unter Anderem in ihrer Masterarbeit intensiv mit dem Thema Datenschutz befasst [Pu19b]. Aktuell arbeitet sie als Dynamics and BI Engineer in einem IT Consulting Unternehmen. Direkt arbeitet sie zur Zeit nicht in den Bereichen Datenschutz und Prozessmodellierung. Allerdings hat sie in beiden Bereichen langjährige Erfahrungen und hat zu beiden Feldern Anknüpfungspunkte und ein solides Grundwissen.

Isabelle Puttrus bezeichnet den Ansatz als sehr gute Idee und stellt die Übersichtlichkeit hervor. Die Farben weisen schnell auf mögliche Problemstellen hin und erleichtern

den Überblick. Auch die Definition der Farben findet sie sinnvoll. Als wichtiger Punkt wird die Abgrenzung zwischen den einzelnen Farbkategorien angesprochen. Zum Thema Einwilligung findet Frau Puttrus es sinnvoll, dass Aktivitäten auch dann rot eingefärbt bleiben, wenn davon ausgegangen wird, dass eine Einwilligung vorliegt, da es jederzeit möglich ist, dass eine Einwilligung widerrufen wird und somit schneller die Teile des Prozesses gefunden werden, an denen eventuell interveniert werden muss. Zur Anonymisierung wird besprochen, dass anonymisierte Datenobjekte nicht datenschutzrelevant sind und daher grün eingefärbt werden können. Allerdings sei es hierbei wichtig, dass entsprechende Kontrollmechanismen existieren, um die korrekte Anonymisierung sicherzustellen.

Als Fazit führt Frau Puttrus auf, dass das Thema Datenschutz bisher kaum in Prozessmodellen betrachtet, sondern eher separat dargestellt wird. Der vorgestellte Ansatz stelle jedoch eine große Hilfe beim Verständnis der Zusammenhänge zwischen beiden Themen dar und könne durchaus auch Unternehmen beim Umgang mit dem Datenschutz helfen. Außerdem sei der Ansatz aus Sicht des Consultings auch eine gute Möglichkeit, das Thema Datenschutz Kunden im Rahmen ihrer Prozesse zu verdeutlichen[GW22b].

10.1.3. Stella Thoben

Stella Thoben, Datenschutzbeauftragte an der CAU ist als einzige Interviewte Juristin. Auf Grund ihrer Ausbildung und ihrer Tätigkeit in dem Bereich hat sie ein großes Wissen im Bereich Datenschutz. Mit Geschäftsprozessmodellierung ist sie nicht direkt betraut. Allerdings werden an der CAU seit längerem typische Geschäftsprozesse modelliert, die auf die Einhaltung der Datenschutzbestimmungen überprüft werden müssen.

Frau Thoben findet die Einfärbung der Geschäftsprozessmodelle sinnvoll. Auch die Anzahl und Auswahl der Farben werden positiv bewertet. Mehr Farben würde sie als unübersichtlich empfinden. Ihrer Einschätzung nach könnte die Einführung des Konzeptes an der Hochschule ihre Arbeit erleichtern. Insbesondere fände der Datenschutz damit wahrscheinlich mehr Beachtung in der Universitätsverwaltung und die Awareness könne erhöht werden. Die rote Färbung von Aktivitäten würde Frau Thoben mit der gleichen Begründung wie Frau Puttrus auch bei vorliegender Einwilligung beibehalten. Zum Thema Anonymisierung führt sie an, dass der Begriff rechtlich deutlich enger gefasst sei als im allgemeinen Verständnis. Dies führe dazu, dass viele Daten, die als anonymisiert angesehen werden, eigentlich nur pseudonymisiert sind. Daher findet Frau Thoben es wichtig, zu prüfen, ob wirklich eine Anonymisierung vorliegt, bevor Daten im Prozess so behandelt werden. Als Erweiterung schlägt sie eine Fußnote oder einen Verweis im Prozess vor, um zu verdeutlichen, ab wann Daten anonymisiert vorliegen[GW22d].

10.1.4. Ricarda Radden

Ricarda Radden ist gelernte Fachinformatikerin und arbeitet zur Zeit als Data Privacy Manager bei einem großen Logistikunternehmen, zuvor auch als externe Datenschutzbeauftragte, sodass sie ein hohes Wissen im Bereich Datenschutz vorweisen kann. Im Bereich der Geschäftsprozessmodellierung sind Grundkenntnisse vorhanden.

Auch sie empfindet das vorgestellte Konzept als sinnvoll und ein Weg der Vereinfachung. Die Auswahl der Farben findet sie wegen der Ampel-Analogie sehr intuitiv. Auch die konkrete Definition der Farbkategorien bezeichnet sie als gut durchdacht und würde keine Änderungen vornehmen. Das Thema Anonymisierung wird in dem Gespräch als problematisch dargestellt, da hier schnell Fehler passieren können, sodass die Anonymisierung immer geprüft werden sollte. Nach erfolgter Überprüfung ist gegen eine grüne Färbung der entsprechenden Datenobjekte aber nichts einzuwenden.

Zusammenfassend stellt Frau Radden fest, dass ihre Arbeit durch den Einsatz des Konzepts durchaus erleichtert werden könnte, weil unmittelbar klar wird, wo in einem Prozess personenbezogene Daten verarbeitet werden. Als groben Zeitansatz schätzt sie 15-30 Minuten Zeitersparnis pro geprüftem Prozess, da die grünen Elemente zumindest für eine Erstprüfung vernachlässigt werden könnten. Voraussetzung hierfür ist natürlich eine korrekte Einfärbung[GW22c].

10.1.5. Georg Rasch

Georg Rasch hat Mathematik studiert, sich während dessen aber auch immer wieder der Informatik gewidmet und auch schon während des Studiums als Datenschutzbeauftragter und IT-Sicherheitsberater gearbeitet. Aktuell ist er Geschäftsführer eines Unternehmens in diesem Bereich, welches unter Anderem externe Datenschutzbeauftragte für Unternehmen stellt. In der Geschäftsprozessmodellierung hat Herr Rasch kaum Erfahrung.

Das Konzept findet auch er sehr gut und übersichtlich. Die Farben werden seiner Aussage nach im Bereich Datenschutz häufig verwendet und sind dementsprechend gut verständlich. Im Gespräch wurde diskutiert, wie mit Modellelementen umgegangen werden sollte, bei denen die korrekte Färbung für denjenigen, der sie übernimmt (oder für ein automatisches System) nicht sicher ist. Herr Rasch vertritt hier die Ansicht, dass derartige Elemente ungefärbt gelassen werden sollten. Alternativ wäre es beispielsweise denkbar, immer die entsprechend kritischere Farbkategorie zu wählen. Dies würde seiner Meinung nach aber den Aufwand erhöhen, da zeitintensiv geprüft werden müsse, welche Einfärbungen korrekt sind. Bei der ungefärbten Variante könnte der Datenschutzexperte in einem ersten Schritt vergleichsweise schnell nur diese Elemente prüfen und selbstständig einfärben. Als großes Problem sieht Herr Rasch den Umstand, dass in vielen

– insbesondere kleineren – Unternehmen gar keine Geschäftsprozessmodelle erstellt werden. Es ist seiner Erfahrung nach in der Regel viel mehr so, dass ein externer Datenschutzbeauftragter in einem Unternehmen die Prozesse beobachtet und sich beschreiben lässt und dabei selbstständig ermittelt, welche Daten an welchen Stellen übermittelt und verarbeitet werden. Die aufgenommenen Prozesse werden dann für gewöhnlich als Fließtext in Kombination mit einigen Diagrammen dokumentiert. Anschließend wird dann geprüft, welche Regeln gelten und was zulässig ist. Schwierig sei hierbei häufig, dass je nach Branche teilweise auch sehr spezielle Gesetze, wie etwa das Luftsicherheitsgesetz oder das Landeskrankenhausgesetz Implikationen für den Datenschutz habe. Hier müsse der Kunde dann entsprechend unterstützen. Eine Prozessmodellierung schon durch den Kunden würde die Arbeit für den Datenschutzbeauftragten deutlich erleichtern. Insbesondere, wenn diese nach dem beschriebenen Konzept eingefärbt wären. Zu bedenken gibt Herr Rasch allerdings noch, dass Prozesse sich unter Umständen. häufig ändern und hier schon für die Anpassung der Prozessmodelle ein enormer Aufwand entsteht. Dementsprechend wäre eine automatisierte Einfärbung wünschenswert, die ohne viel zusätzlichen Aufwand nach jeder Prozessänderung durchgeführt werden könnte.

Abschließend fasst Herr Rasch zusammen, dass aus seiner Sicht alles hilfreich ist, was die Übersicht erleichtert. Hierfür hat er noch einige Ergänzungen zum vorgestellten Ansatz. Einerseits wäre ein automatisches Auslesen von Rechtsgrundlagen interessant. Hierfür könnte an die eingefärbten Prozesselementen die entsprechende Rechtsgrundlage in Form eines Paragraphen bzw. Artikels angeheftet werden. Andererseits wäre die Generierung von Prozessmodellen aus Fließtext sehr hilfreich, damit der vorgestellte Ansatz überhaupt zur Anwendung kommen kann[GW22a].

10.1.6. Zusammenfassung

Die vier Interviews haben insgesamt ein großes Interesse am vorgestellten Konzept gezeigt. Alle fanden die Idee grundsätzlich sehr gut und hatten auch keine grundlegenden Änderungsvorschläge. Insgesamt bestätigt sich die auch in Kapitel 1 dargestellte Motivation der Arbeit. Alle Gesprächspartner sehen aktuell mehr oder weniger große Probleme bei der Betrachtung des Datenschutzes in ihren eigenen Unternehmen bzw. Organisationen oder auch bei Kunden. Mangelnde Übersicht scheint die Arbeit unnötig zu verkomplizieren und zu verlangsamen.

Es wurden einzelne Optimierungs- oder Erweiterungsvorschläge gegeben. Erste Umsetzungsideen dazu werden ausführlich in Kapitel 11 dargestellt.

Die wichtigsten genannten Aspekte zu den einzelnen Fragestellungen werden im Folgenden noch in Stichpunkten zusammengefasst.

Allgemein

- Sinnvoller Ansatz
- Hilft bei erster Einschätzung des Datenschutzes
- Erhöht Awareness
- Problem: Häufig keine Prozessmodelle vorhanden
- Datenflüsse ziehen sich teilweise durch mehrere Prozesse

Farbwahl

- Gut gewählt
- Vermitteln intuitiv das richtige Gefühl
- Mehr Farben eher unübersichtlich
- Im Datenschutzbereich üblich
- Auch in anderen Bereichen verbreitet
- Eventuell zusätzliche Farbe für Unsicherheit

Definitionen

- Aufgaben auch bei vorhandener Einwilligung rot Färben, um Aufmerksamkeit zu erhöhen
- Anonymisierung sollte kritisch betrachtet werden, weil häufig nur pseudonymisiert wird
- Wenn sicher anonymisiert, kann grün gefärbt werden

Erweiterungsvorschläge

- Spezieller Hinweis auf Risiko bei Anonymisierung
- Aufbewahrungsfristen darstellen
- Rechtsgrundlagen abbilden
- Prozessmodell aus Fließtext generieren
- Abgrenzungen der Farben verdeutlichen

Automatisierung

- Besonders Hilfreich und wichtig bei häufigen Änderungen
- Je nach Domäne müssen eventuell spezielle Gesetze beachtet werden
- Unter Umständen müssen andere Prozesse beachtet werden
- Bei Unsicherheit Elemente weiß lassen oder zusätzliche Farbe verwenden

10.2. Vergleichsexperiment

Neben der Expertenbefragung wurde auch noch ein Vergleichsexperiment durchgeführt. Die Grundidee dahinter die war Problemstellung, datenschutzkritische Aspekte in Prozessen zu finden, wobei untersucht wurde, ob dies mit einem eingefärbten Prozessmodell leichter fällt als in einem Standard-Prozessmodell. Der konkrete Versuchsaufbau und die Resultate werden im Folgenden erläutert. Vorweg ist an dieser Stelle anzumerken, dass der Versuchsaufbau sich wegen methodischer Schwächen im Laufe der Untersuchung einige Male verändert hat. Da aber insgesamt auf eine relativ geringe Teilnehmeranzahl zurückgegriffen wurde und auch die Ergebnisse aus den ersten - methodisch ungünstigen - Experimenten eine Aussagekraft haben, werden hier auch die entsprechenden Veränderungen beschrieben.

10.2.1. Versuchsaufbau

Der grundlegende Versuchsaufbau besteht aus Aufgabenblättern mit einer Aufgabenstellung und je zwei verschiedenen Prozessmodellen, auf denen jeweils datenschuztkritische Stellen markiert werden sollten. Hierbei wurde ein Prozessmodell mit dem beschriebenen Konzept eingefärbt, das andere farblos belassen. Die verwendeten Prozesse sind an die Beispiele aus Kapitel 7 bzw. Kapitel 9 angelehnt. Es wurden allerdings einige Änderungen vorgenommen, um einerseits das Verständnis zu erleichtern und andererseits die Prozessmodelle möglichst ähnlich in Umfang und Komplexität zu gestalten.

Es wurden jeweils zwei unterschiedliche Aufgabenvarianten verwendet, bei welchen getauscht wurde, welcher der beiden Prozesse eingefärbt wurde, um sicherzustellen, dass die Probleme nicht ein einem Prozess „leichter“ zu finden seien (siehe Tabelle 10.1). Die Aufgabenstellung an sich und alle weiteren Rahmenbedingungen waren ansonsten identisch.

Die entsprechenden Aufgabenstellungen inklusive kurzem Überblick über die BPMN-Notation finden sich in Anhang A. Die Teilnehmer waren im Wesentlichen Studenten

Tabelle 10.1.: Verwendete Prozesse in den beiden Aufgabenvarianten

Aufgabenvariante	Prozess ungefärbt	Prozess gefärbt
A	Einstellung	Arzt
B	Arzt	Einstellung

verschiedener Studiengänge und teilweise deren Dozenten. Durchgeführt wurde das Experiment jeweils in Lehrveranstaltungen, in Präsenz, online und auch hybrid.

Durchlauf 1

Der erste Durchlauf des Experiments fand an der Universität Oslo in einer Lehrveranstaltung des Moduls *IN5540 - Privacy by Design* statt. Die Teilnehmer der Veranstaltung hatten vor dem Experiment wenig bis keine Vorstellung von Geschäftsprozessmodellierung, insbesondere von BPMN, dafür aber ein fortgeschritteneres Verständnis von Datenschutz und der DSGVO. Die Lehrveranstaltung selbst wurde in Präsenz durchgeführt, die Befragenden waren per Videokonferenz zugeschaltet. Die Aufgabenblätter wurden im Format Din A3 ausgedruckt von einem Helfer ausgeteilt und wieder eingesammelt. Prozesse und Aufgabenstellung wurden für diesen Durchlauf in die englische Sprache übersetzt, welche auch Lehrsprache des Moduls ist. Neben den Studenten nahm auch der Dozent an dem Experiment teil.

Nach einer kurzen Einführung in den Kontext des Experiments sowie in BPMN und Präsentation der Aufgabenstellung, wurden die Aufgabenblätter ausgeteilt. Um einem möglichen Einfluss eines Lerneffekts vorzubeugen, wurde der Hälfte der Teilnehmer die Prozesse in umgekehrter Reihenfolge ausgeteilt - sodass eine Hälfte erst den farblosen und die andere Hälfte erst den gefärbten Prozess bearbeiten sollte. Den Teinehmern wurde zunächst zehn Minuten Bearbeitungszeit gegeben. Da aber scheinbar einige Probleme auftauchten und die Teilnehmer mehr Zeit erbaten, wurden fünf zusätzliche Minuten eingeräumt.

Durchlauf 2

Teilnehmer des zweiten Durchlaufs waren die Teilnehmer einer Vorlesung zum Thema Qualitätsmanagement im Master-Studiengang Wirtschaftsinformatik der CAU. Die Studenten sollten sowohl Grundkenntnisse der Geschäftsprozessmodellierung mit BPMN als auch im Bereich Datenschutz haben. Die Veranstaltung fand als Videokonferenz statt. Bei Betrachtung der Ergebnisse des ersten Durchlaufs waren einige methodische Schwächen aufgefallen, weshalb die Fragestellung für den zweiten Durchlauf etwas angepasst wurde. Außerdem wurde am Ende des Aufgabenblatts eine weitere Seite

mit einer offenen Frage nach eventuellen Problemen bei der Bearbeitung eingefügt. Auch hier gab es wieder die oben beschriebenen zwei verschiedenen Ausführungen des Aufgabenblattes.

Zu Beginn der Veranstaltung wurde nur kurz in den Kontext des Experiments und in das weitere Vorgehen eingeführt. Anschließend wurde die Gruppe zufällig in zwei getrennte Räume aufgeteilt, in denen jeweils eine Variante des Aufgabenblatts zur Verfügung gestellt wurde. Die Studenten wurden gebeten, das Aufgabenblatt herunterzuladen und zu bearbeiten. Anschließend sollte die Lösung wieder hochgeladen werden. Die Teilnehmer hatten zehn Minuten Bearbeitungszeit, wobei etwas zusätzliche Zeit zum Hochladen eingeräumt wurde.

Durchlauf 3

Ein dritter Durchlauf fand ebenfalls an der CAU in einer Vorlesung des Moduls „Grundlagen E-Commerce“ statt. Die Veranstaltung richtet sich an Bachelorstudierende der Wirtschaftsinformatik. Auch diese Veranstaltung fand als Videokonferenz statt.

Im Gegensatz zu Durchlauf 2 wurden nun nicht mehr beide Prozessmodelle (also das gefärbte und das ungefärbte) gleichzeitig ausgegeben. Stattdessen wurde zunächst nur das ungefärbte Modell ausgegeben. Nach einer Bearbeitungszeit von fünf Minuten wurde anschließend das zweite Modell ausgegeben, worauf wieder fünf Minuten Bearbeitungszeit folgten.

Durchlauf 4

Der letzte Durchlauf des Experiments wurde in einer Übung zum Modul „Externes Rechnungswesen“ an der CAU durchgeführt. Zielgruppe des Moduls sind Studenten der Bachelorstudiengänge Wirtschaftschemie, Wirtschaftsinformatik, Wirtschaftsingenieurwesen und Informatik mit Nebenfach BWL. Die planmäßige Semesterlage ist hierbei unterschiedlich. Es können also Studenten mit unterschiedlichen Vorkenntnissen teilnehmen. Generell kann die Erfahrung im Bereich Geschäftsprozessmodellierung und Datenschutz aber eher als niedrig bewertet werden.

Das Experiment wurde in zwei aufeinander folgenden Übungsgruppen mit unterschiedlichen Teilnehmern durchgeführt. Die Veranstaltungen fanden hierbei in Präsenz statt.

Das Verfahren entsprach im Wesentlichen dem Vorgehen in Durchlauf 3, abgesehen davon, dass die Aufgabenblätter in Papierform auf Din A3 ausgedruckt ausgegeben und wieder eingesammelt wurden. Der eingefärbte Prozess wurde jeweils beim Einsammeln des farblosen Prozess ausgegeben. Für jeden Prozess hatten die Teilnehmer fünf Minuten Zeit.

10.2.2. Ergebnisse

Ziel des Experiments war eine quantitative Auswertung, wobei von Anfang an klar war, dass die vergleichsweise geringe Teilnehmerzahl keine wirklich repräsentative Auswertung ermöglichen wird. Zusätzlich sind im Verlauf des Experiments einige Probleme aufgetreten, weshalb die Methodik angepasst wurde. Dies erschwert die Auswertung zusätzlich. Nichtsdestotrotz erfolgt an dieser Stelle eine kurze quantitative Analyse der Ergebnisse. Hierfür werden im Folgenden einige Kennzahlen aufgeführt. Alle absoluten Werte sind auf zwei Nachkommastellen gerundet, die Prozentangaben auf eine Nachkommastelle.

Nach Aufgabenstellung sollten jeweils die „datenschutzrelevanten“ Modellelemente markiert werden. Diese Bezeichnung lässt natürlich einigen Interpretationsspielraum. Mindestens können aber alle rot klassifizierten Modellelemente als datenschutzrelevant bezeichnet werden. Dies sind in beiden ausgeteilten Prozessen jeweils fünf. In einem ersten Schritt wird bestimmt, wie viele der fünf Elemente von den Teilnehmern durchschnittlich markiert wurden.

Tabelle 10.2.: Anzahl markierter rot klassifizierter Modellelemente

	Ungefärbt	**Gefärbt**
Prozess Einstellung	2,72 (54,3%)	3,90 (78,0%)
Prozess Zahnarzt	3,16 (62,9%)	3,96 (79,3%)

Tabelle 10.2 zeigt, wie viele der rot klassifizierten Modellelemente von den Teilnehmern in den verschiedenen Prozessmodellen markiert wurden. Im Einstellungsprozess ist eine deutliche Verbesserung der Bewertung vom ungefärbten zum gefärbten Prozessmodell zu erkennen. Während im ungefärbten Prozess nur etwas über die Hälfte der rot zu klassifizierenden Modellelemente eingefärbt wurden, sind es im gefärbten Prozessmodell mehr als drei Viertel der roten Elemente. Auch im Zahnarzt-Prozess ist eine entsprechende Verbesserung zu erkennen, wobei diese hier mit 62,9% zu 79,3% weniger deutlich ausfällt.

Tabelle 10.3.: Anzahl markierter gelb klassifizierter Modellelemente

	Ungefärbt	**Gefärbt**
Prozess Einstellung	2,66 (26,6%)	3,83 (38,3%)
Prozess Zahnarzt	4,22 (38,4%)	3,16 (28,8%)

In Tabelle 10.3 ist ablesbar, wie viele der gelb kategorisierten Elemente markiert wurden. Insgesamt sind dies im Einstellungsprozess zehn und im Zahnarztprozess elf

Elemente. Interessanterweise wurden hier im Einstellungsprozess in der gefärbten Variante mehr gelb klassifizierte Modellelemente markiert als in der ungefärbten Variante; im Zahnarztprozess verhält es sich aber genau umgekehrt. Hier wurden in der gefärbten Modellvariante weniger Elemente markiert. Insgesamt sind die Werte hier aber auch alle relativ ähnlich (zwischen 26,6% und 38,4%). Zu beachten ist an dieser Stelle auch, dass die Aufgabenstellung sowohl so interpretiert werden kann, als wären die gelb klassifizierten Elemente zu markieren, als auch so, als wären sie nicht zu markieren. Somit sind die Ergebnisse an dieser Stelle eingeschränkt aussagekräftig.

Tabelle 10.4.: Anzahl markierter grün klassifizierter Modellelemente

	Ungefärbt	**Gefärbt**
Prozess Einstellung	2,38 (26,4%)	0,76 (8,4%)
Prozess Zahnarzt	1,11 (13,9%)	0,16 (2%)

Die grünen Modellelemente werden in Tabelle 10.4 betrachtet. Insgesamt sind hier im Einstellungsprozess neun und im Zahnarztprozess acht Elemente vorhanden, die so kategorisiert werden. Die Anzahl der markierten Elemente ist in allen vier Modellvarianten recht gering, weißt aber doch eine relativ hohe Spannweite auf (2% bis 26,4%). Bei beiden Prozessen wurden in den gefärbten Modellen weniger grün kategorisierte Elemente markiert, als in den ungefärbten Modellen.

Interpretation Das Vergleichsexperiment legt nahe, dass die Einfärbung des Prozessmodells die gewünschte Wirkung zeigt und den Betrachter auf datenschutzrelevante Aspekte eines Prozessmodells hinweist. Grundsätzlich scheint die Semantik der Farben verständlich zu sein und die Interpretation des Prozessmodells in Bezug auf den Datenschutz zu erleichtern.

10.2.3. Limitationen

Das beschriebene Experiment hat nur eine begrenzte Aussagekraft. Ein Grund hierfür ist die vergleichsweise geringe Teilnehmerzahl. Außerdem bilden die Teilnehmer eine relativ homogene Gruppe. Fast alle Teilnehmer sind Studenten. Auch die Studienfächer sind relativ ähnlich.

Außerdem wurde zwischen den einzelnen Experiment-Durchläufen die Methodik mehrfach verändert, weshalb die Ergebnisse untereinander nicht wirklich vergleichbar sind. Der Grund für die veränderte Methodik waren scheinbare Probleme beim Verständnis der Aufgabenstellung. Dies zeigte sich in den ersten Durchläufen durch unerwartete

Abgaben. Ziel war eigentlich eine Markierung der datenschutzkritischen Modellelemente, beispielsweise durch Einkreisen. Stattdessen haben aber einige Teilnehmer z. B. versucht das Färbekonzept selbst auf das ungefärbte Prozessmodell anzuwenden. Dies ist zwar durchaus auch interessant, entspricht aber nicht der Fragestellung und lässt sich daher auch schlecht gemeinsam mit den anderen Ergebnissen statistisch auswerten. Andere Teilnehmer haben nur Aktivitäten und keine Datenobjekte betrachtet. Auch dies verfälscht die Aussagekraft.

10.3. Schlussfolgerungen

Auch wenn die bisherigen Untersuchungen nicht als repräsentativ betrachtet werden können, geben sie doch einen guten ersten Eindruck vom Nutzen des Konzepts.

Im qualitativen Teil der Untersuchung haben alle Interviewten bestätigt, dass der präsentierte Ansatz durchaus hilfreich sein könnte. Dies spiegelt auch die quantitative Analyse der durchgeführten Experimente wider. Hier konnte gezeigt werden, dass durch die Einfärbung eine bessere Erkennung von datenschutzkritischen Bereichen eines Prozesses und auch von entsprechend unkritischen Bereichen erreicht werden kann. Die Kategorisierung und Farbwahl scheinen wie erwartet recht intuitiv zu sein.

Allerdings haben die Untersuchungen auch einige negative Aspekte aufgezeigt. Das größte Problem ist, dass in der Praxis – zumindest bei kleinen und mittelständischen Unternehmen – wenig Geschäftsprozesse überhaupt modelliert werden. Das bedeutet, dass das Konzept nur in wenigen Fällen überhaupt zur Anwendung kommen kann. Allerdings stellt sich die Frage, ob bei einem größeren erkannten Nutzen eventuell auch mehr Unternehmen ein entsprechendes Geschäftsprozessmanagement einführen. Ein anderer Aspekt sind mögliche Verständnisprobleme. Während ein grundsätzliches Verständnis durch die intuitiv wahrgenommenen Farben zwar vorliegt, ist die Abgrenzung der einzelnen Farben im Detail nicht offensichtlich. Dies kann – insbesondere bei neuen Anwendern – schnell zu Fehlern führen. Diese könnten unter Umständen fatale Folgen haben, wenn die Färbung durch eine Person falsch umgesetzt wird und eine andere Person basierend darauf Entscheidungen trifft.

10.3.1. Optimierungen/Erweiterungen

Die Experteninterviews haben einige Hinweise auf mögliche Optimierungen und Erweiterungen des Konzepts ergeben.

Besonders interessant ist die Idee, **Rechtsgrundlagen** abzubilden. Hier könnte für Aufgaben beispielsweise die Verarbeitungsgrundlage angegeben werden. Außerdem

könnte bei Datenobjekten angegeben werden, um welche Art von (personenbezogenen) Daten es sich handelt. Also beispielsweise Stammdaten, medizinische Daten usw.

Als kritisch wurde von allen Befragten das Thema **Anonymisierung** angesehen, da es hier häufig zu Fehlern kommt. Wenn die Anonymisierung allerdings wirklich korrekt abgelaufen ist, wird es als durchaus sinnvoll betrachtet, die entsprechenden Datenobjekte grün einzufärben, da dann kein Personenbezug mehr hergestellt werden kann. Da dies aber nur an Hand des Prozessmodells nicht wirklich sichergestellt werden kann, bietet sich ein Hinweis an derartigen Modellelementen an, dass die Anonymisierung unbedingt geprüft werden muss.

Eine weitere Idee ist die Abbildung von gesetzlichen **Aufbewahrungsfristen** im Prozessmodell, etwa aus dem Steuerrecht. Diese haben durchaus einen Zusammenhang zu den Kriterien des vorgestellten Konzepts, da sie an einigen Stellen Verarbeitungsgründe darstellen können.

Neben diesen Ergänzungen wurde noch angemerkt, dass die konkreten **Definitionen der Farben** nicht ganz intuitiv sind und daher die Abgrenzung zwischen den verschiedenen Kategorien möglichst direkt bei der Betrachtung des Prozessmodells ersichtlich sein sollten.

Die bisher genannten Ideen werden in Kapitel 11 näher betrachtet.

Ein Kernproblem, das sich in den Gesprächen gezeigt hat ist, dass häufig gar keine Prozessmodelle vorliegen, die entsprechend kategorisiert und eingefärbt werden könnten. Daher wurde eine automatische **Generierung von Prozessmodellen** aus textuellen Prozessbeschreibungen vorgeschlagen. Dieses Fragestellung betrifft allerdings nur am Rande das Thema dieser Arbeit und es bestehen auch bereits einige Ansätze (siehe z. B. [Ep15; FMP11; GKC07; No20]). Daher wird das Thema nicht weiter verfolgt.

10.3.2. Prototyp/Automatisierung

Die Befragungen haben ergeben, dass neben den oben beschriebenen Erweiterungen des Konzepts, unbedingt ein erster Prototyp entwickelt werden sollte, der alle Funktionalitäten umsetzt. Teil III beschreibt ausführlich die Entwicklung eines derartigen Prototyps. Ein wichtiger Aspekt ist hierbei die automatische Einfärbung der Prozessmodelle. In den Interviews wünschen sich alle Beteiligten eine Arbeitserleichterung. Mit der manuellen Einfärbung entsteht aber zumindest initial ein höherer Aufwand. Selbst unter der Annahme, dass die späteren Effizienzsteigerungen diesen Aufwand ausgleichen, kann er zunächst trotzdem abschreckend wirken. Darüber hinaus fällt je nach Organisationsstruktur dieser zusätzliche Aufwand eventuell bei einer anderen Person an, als bei derjenigen, die im Anschluss die Vorteile nutzen kann. Dies würde recht wahrscheinlich zur Ablehnung des Konzepts auf dieser Seite führen. Daher

soll zumindest eine teilweise Automatisierung ermöglicht werden. Der Prototyp kann anschließend für weitere Evaluationen des Konzepts genutzt werden.

10.3.3. Weitere Erhebungen

Es können noch weitere Erhebungen durchgeführt werden, um das Konzept weiter zu bestätigen und zu optimieren. Hierfür bietet es sich an, den in Abschnitt 10.3.2 angesprochen Prototyp zu verwenden. Auch für die zukünftige Evaluation bietet es sich an, verschiedene (qualitative und quantitative) Verfahren anzuwenden.

Einerseits können weitere Gespräche geführt werden, da hierbei am besten erkannt werden kann, wie gut die Gesprächspartner die Idee wirklich verstanden haben. Außerdem entwickeln sich im direkten Gespräch weitere Ideen für mögliche Optimierungen. Eine Option wäre es, den Gesprächspartnern vorab den Prototyp zur Verfügung zu stellen[1] und sie zu bitten, diesen zu testen und – wenn möglich – in ihren Arbeitsalltag zu integrieren. Nach einer vorgegebenen Zeitspanne könnte dann etwa besprochen werden, inwiefern das System den Benutzer unterstützen konnte, ob Probleme aufgetreten sind, welche Optimierungen wünschenswert wären, etc. Bei der Auswahl der Gesprächspartner sollte dann darauf geachtet werden, dass zumindest einige davon in ihrem Berufsalltag mit Geschäftsprozessmodellen arbeiten.

Andererseits ergibt es aber durchaus auch Sinn, weitere quantitative Erhebungen durchzuführen. Auch hier könnte der Prototyp eingesetzt werden. Eine interessante Fragestellung wäre, ob mit Hilfe des Konzepts bzw. des Prototyps eine bestimmte Arbeit schneller umgesetzt werden kann. Getestet werden könnte beispielsweise die Erstellung einer Datenschutzfolgeabschätzung oder eines Verzeichnis von Verarbeitungstätigkeiten. Die Probanden sollten hier bestenfalls Personen sein, welche mit derartigen Aufgabestellungen auch in ihrem Berufsalltag konfrontiert sind, um ein möglichst aussagekräftiges Ergebnis zu erzielen. Außerdem sollte beachtet werden, ob die Teilnehmer auch vorher schon mit Geschäftsprozessmodellen gearbeitet haben, um eine Verzerrung der Ergebnisse durch diesen Faktor zu vermeiden.

[1] Hier sollte mindestens eine Dokumentation mitgeliefert werden. Eventuell ist auch eine praktische Einweisung sinnvoll.

11. Erweiterung der Visualisierung

Das in Kapitel 8 beschriebene Färbungskonzept bringt zwar viele Vorteile wie die große Übersichtlichkeit mit sich, hat aber durchaus auch einige Erweiterungsmöglichkeiten (siehe Evaluation in Kapitel 10). So ist beispielsweise die Aussagekraft erweiterbar. Die Färbung trifft eine Aussage darüber, ob etwas datenschutzrelevant ist, zeigt aber keine Gründe dafür auf. Gerade für die weiteren Arbeitsschritte eines Datenschutzbeauftragten, der beispielsweise ein Verzeichnis von Verabeitungstätigkeiten oder eine Datenschutzfolgeabschätzung erstellen möchte, sind diese Informationen aber sehr wichtig. Hier stellt sich nun die Frage, wie das vorgestellte Konzept erweitert werden kann, um mehr Informationen zu vermitteln, welche das sein könnten und wie diese visualisiert werden könnten.

Die Inhalte in diesem Kapitel sind teilweise im Rahmen der Betreuung der Masterarbeit von Paul Hilge [Hi23] entstanden. Hier finden sich weitere Details zu einigen Überlegungen.

11.1. Zusatzinformationen

Erste Hinweise darauf, welche Informationen zusätzlich zum bereits beschriebenen Konzept interessant wären, können der Zusammenfassung der Experteninterviews in Abschnitt 10.1 entnommen werden. Sinnvolle Zusatzinformationen sind demnach und einigen weiteren Überlegungen zufolge folgende Aspekte:

Verarbeitungsgrund: Gelb und rot eingefärbte Aufgaben haben gemeinsam, dass sie sich mit der Verarbeitung personenbezogener Daten befassen. Die Farbe gibt einen gewissen Aufschluss über die Verarbeitungsgrundlage, da rot immer *Einwilligung* bedeutet. Mit einer zusätzlichen Visualisierung kann bei den bei gelb eingefärbten Aufgaben die Färbung begründet und auch weiter differenziert werden.

Art der Datenverarbeitung: Unter Umständen ist auch die Art der Datenverarbeitung interessant für die spätere Betrachtung.

Art dargestellter Daten: Unter Umständen kann es sinnvoll sein, zu hinterlegen, um welche Art (personenbezogener) Daten es sich bei einem Datenobjekt bzw.

-speicher handelt. Die Färbung gibt lediglich Aufschluss darüber, ob personenbezogene Daten oder besondere Kategorien personenbezogener Daten enthalten sind. Unter Umständen ist es aber relevant zu wissen, ob es sich hier beispielsweise um medizinische Daten oder Daten, die die Religion betreffen, handelt. Dies ist insbesondere dann wichtig, wenn aus der Bezeichnung eines Objekts nicht direkt klar wird, um welche Art von Daten es sich konkret handelt.

Dargestelltes Rechtssubjekt: Für die weitere Betrachtung und auch für die automatische Kategorisierung (siehe Abschnitt 12.3) macht es Sinn festzuhalten, welche Art von Rechtssubjekt ein Prozessteilnehmer darstellt. So könnte beispielsweise der Betroffene oder auch der für die Datenverarbeitung Verantwortliche kenntlich gemacht werden. Interessant ist dies insbesondere, wenn aus dem Prozess nicht unbedingt hervorgeht, welche Rollte ein Teilnehmer einnimmt. Im Beispiel des Zahnarzt-Prozesses aus Abschnitt 7.2 und Abschnitt 9.2, ist beispielsweise nicht eindeutig, welche Rolle die Abrechnungsstelle einnimmt. Hier wäre einerseits denkbar, dass diese als Auftragsverarbeiter im Sinne von Art. 4 Nr. 8, 28 DSGVO agieren. Andererseits könnte die Abrechnungsstelle aber im Rahmen einer gemeinsamen Datenverarbeitung im Sinne von Art. 26 DSGVO auch als Verantwortlicher in Betracht kommen [Mö20].

Außerdem haben die Befragungen noch weitere Aspekte ergeben, die an dieser Stelle nur am Rande betrachtet werden sollen:

- Spezieller Hinweis auf Risiko bei Anonymisierung
- Aufbewahrungsfristen
- Legende mit Abgrenzungen der Farben

11.2. Visualisierungsarten

Grundsätzlich muss bei der Visualisierung immer zwischen grafischer und textueller Darstellung unterschieden werden. Wie in Kapitel 4 erörtert, bietet die grafische Darstellung häufig den Vorteil eines schnelleren und intuitiven Verständnisses, aber auch den Nachteil einer begrenzten Aussagekraft. Sinnvoll ist daher oft die Kombination beider Möglichkeiten, wie es letztlich auch in der Prozessmodellierung umgesetzt wird.

11.2.1. Grafische Elemente

Da das in diesem Kapitel beschriebene Konzept hauptsächlich auf Farben basiert, bietet es sich an, für zusätzliche Informationen eine andere Darstellungsart zu wählen. Ein zur

grafischen Darstellung häufig verwendeter Ansatz ist die Verwendung von Piktogrammen. Zur Darstellung der Art einer Datenverarbeitung verwendet beispielsweise die an BPMN angelehnte „PICTURE-Methode“[BAF07; BAR09; Be07a; Be07b] verschiedene Aufgabentypen, an denen jeweils durch ein Piktogramm an der oberen linken Ecke dargestellt wird, um welche Art von Tätigkeit es sich handelt (siehe auch Abschnitt 3.2).

11.2.2. Textuelle Darstellung

Im Bereich der textuellen Darstellung bietet es sich an, die vom BPMN-Standard vorgegebene Möglichkeit der Annotation zu nutzen. In [BF20] wird diese etwa genutzt, um an Datenflüssen zu hinterlegen, welche Daten verwendet werden und was der Verwendungszweck ist. Hierfür wird ein konkretes Benennungsschema eingeführt: [Zweck,{Attributnamen}]. Ein Beispiel hierfür findet sich in Abbildung 11.1.

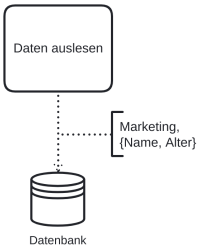


Abbildung 11.1.: Verwendung einer Annotation für die Darstellung von verwendeten Daten und Verwendungszweck (in Anlehnung an [BF20])

11.2.3. Interaktion

Die Visualisierung im Rahmen einer interaktiven Software, bietet die Möglichkeit einer dynamischen Änderung in Abhängigkeit von Benutzereingaben. Das hat den Vorteil, dass auf der gleichen Fläche insgesamt mehr Informationen untergebracht werden können, ohne die Übersichtlichkeit zu verringern, da einige Informationen nur temporär eingeblendet werden.

Möglichkeiten zur Interaktion sind etwa:

- der Klick direkt auf ein schon bestehendes Visualierungselement,
- Mouseover über ein Visualierungselement, oder
- Klick auf eine separate Schaltfläche oder ein Menüelement.

Allerdings bietet sich bei allen Varianten unter Umständen ein Problem mit der Benutzerfreundlichkeit, da der Nutzer hier aktiv eingreifen muss. Hier stellt sich erneut

die Frage nach einer intuitiven Umsetzung. Alternativ oder zusätzlich können Hinweise auf Interaktionsmöglichkeiten erstellt werden, die dem Nutzer mindestens bei Erstbenutzung angezeigt werden.

11.3. Umsetzungsideen

Dieser Abschnitt bietet einen Überblick über erste Ideen zur Visualisierung der Aspekte aus Abschnitt 11.1. Ein Teil der hier erwähnten Ideen ist im Rahmen einer Masterarbeit [Hi23] entstanden, welche im Kontext dieser Dissertation betreut wurde.

11.3.1. Verarbeitungsgrund

Neben der Art der Datenverarbeitung ist auch der Grund hierfür interessant, da dieser über die Rechtmäßigkeit mit entscheidet. Die Einfärbung einer Aufgabe gibt hier einen ersten Aufschluss, da zumindest bei roten Aufgaben klar ist, dass der einzige vorliegende Verarbeitungsgrund die Einwilligung ist. Bei gelben Aufgaben kommen allerdings mehrere Verarbeitungsgründe in Frage, die aus dem Kontext teilweise nicht unbedingt klar werden. Hier bietet sich eine zusätzliche Darstellungsform an. Abbildung 11.2 zeigt erste Entwürfe für eine derartige Darstellung.

(a) Kennzeichnung einer gelben Aufgabe (b) Kennzeichnung einer roten Aufgabe

Abbildung 11.2.: Kennzeichnungen für Aufgaben

Die Aufgaben werden hier mit zusätzlichen Symbolen versehen, die auch in der jeweiligen Farbe der Aufgabe gehalten sind. Die Formen orientieren sich an Verkehrszeichen: Gelbe Aufgaben erhalten ein Quadrat, welches so ausgerichtet ist, dass eine Spitze nach oben zeigt. Dies erinnert an das Verkehrszeichen, welches eine Vorfahrtsstraße ankündigt. Rote Aufgaben hingegen werden mit einem Achteck versehen, welches an ein Stoppschild erinnert. Beides passt auch zur Semantik der Farbkategorisierung: Bei gelben Aufgaben liegt ein guter Grund für die Verarbeitung vor. Es herrscht gewissermaßen „freie Bahn“ für die Verarbeitung. Trotzdem müssen natürlich die gesetzlichen

Gegebenheiten beachtet werden. Dies trifft aber auch auf die Vorfahrtsstraße zu. Bei roten Aufgaben hingegen ist die Einwilligung der einzige Verarbeitungsgrund. Dies ist aus verschiedenen Gründen problematisch. Hier ist besondere Vorsicht geboten. Vor der Ausführung gründlich überprüft werden, ob wirklich eine Einwilligung vorliegt. Für rote Aufgaben bietet sich alternativ ein Kreis an. Dieser kann in der Verkehrszeichen-Analogie mit einem „Einfahrt verboten“-Schild assoziiert werden.

Beide zusätzlichen Symbole erhalten außerdem einen Text: Hier wird die Stelle der DSGVO angegeben, welche den jeweiligen Verarbeitungsgrund darstellt. Bei roten Aufgaben ist dies immer Art. 6 (1) a. Bei gelben Aufgaben kommen verschiedene Stellen in Frage. Für ein besseres Verständnis wäre hier eigentlich ein Stichwort besser, als der Verweis auf die DSGVO. Dies benötigt allerdings deutlich mehr Platz und ist daher schwierig in den Formen unterzubringen. Eine weitere Alternative wäre ein Piktogramm, welches den Verarbeitungsgrund darstellt. Hier ist es aber wieder schwierig, eindeutige Symbole zu finden.

Um das Verständnis zu erleichtern, können interaktive Ansätze genutzt werden. Die folgenden beiden Abbildungen zeigen zwei mögliche Ansätze.

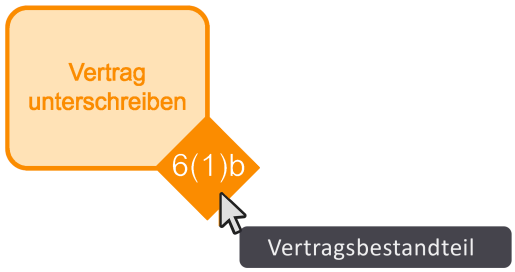


Abbildung 11.3.: Erster Entwurf: Mouseover an einer Aktivität

In Abbildung 11.3 wird eine Mouseover-Interaktion dargestellt. Wenn der Nutzer mit dem Cursor auf das Quadrat zeigt, wird eine Tooltipp dargestellt, welcher die Verarbeitungsgrundlage zeigt.

Einen umfangreicheren Ansatz zeigt Abbildung 11.4. Hier öffnet sich nach einem Klick auf das Quadrat eine Infobox. In dieser wird einerseits in einem Dropdown-Menü ein Stichwort zur Verarbeitungsgrundlage angegeben, wobei selbige hiermit auch verändert werden kann. Außerdem wird der entsprechende Gesetzestext angezeigt. Ein Klick auf den Button *„Okay“* schließt die Infobox und speichert eventuelle Änderungen im Dropdown-Menü.

Beide Ansätze können natürlich auch kombiniert werden.

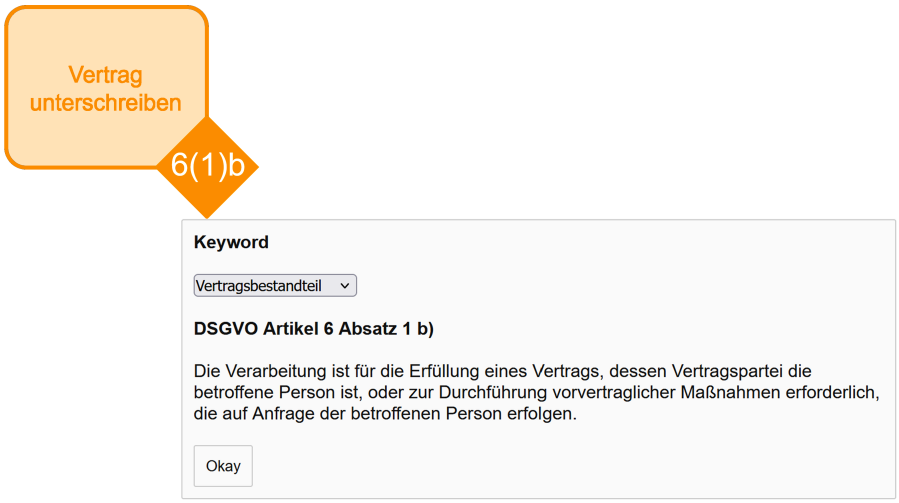


Abbildung 11.4.: Zweiter Entwurf: Zusatzfenster an einer Aktivität

11.3.2. Art der Datenverarbeitung

Die Art der Datenverarbeitung geht in aller Regel aus dem Inhalt der Aufgabe hervor. Daher muss diese nicht direkt dargestellt werden. Für Fälle, in denen die Art der Datenverarbeitung unklar ist, bietet sich aber eine Mouseover-Interaktion an, vergleichbar mit der Darstellung aus Abbildung 11.3. Je nach Güte der automatischen Klassifizierung, wäre aber auch ein Dropdown-Menü, wie es in Abbildung 11.4 zur Anwendung kommt, eine sinnvolle Darstellungsart.

Inhaltlich bieten sich hier die Begrifflichkeiten für Verarbeitungstätigkeiten aus Art. 4 DSGVO an. Diese wurden bereits in Abschnitt 2.1.3 eingeführt und werden in Abschnitt 12.2.3 erneut aufgenommen und in einer Taxonomie klassifiziert.

Eine grafische Darstellung gestaltet sich an dieser Stelle schwierig, da für die meisten Arten der Datenverarbeitung keine eindeutig verständlichen Symbole existieren. Hierfür liegen die Kategorien semantisch zu dicht bei einander. Möglich wären hier Symbole für die Oberkategorien, aber auch hierbei muss die Verständlichkeit in Frage gestellt werden.

11.3.3. Art der Daten

Bei der Art der verwendeten Daten sind zwei Optionen denkbar. Einerseits kann argumentiert werden, dass diese Information nicht unbedingt auf den ersten Blick erkenntlich sein muss. Hier bietet sich dann eine reine Interaktionsdarstellung an. Andererseits ist an den Datenobjekten (und Datenspeichern) bisher noch keine Zusatzinformationen

vorhanden. Es besteht also auch die Möglichkeit einer dauerhaften Darstellung, ohne das Modell zu überfrachten.

Für eine einheitliche Darstellung bietet es sich an, diese an die Visualisierung des Verarbeitungsgrunds anzulehnen und die zusätzlichen Symbole an der unteren rechten Ecke des Objekts zu verwenden. An dieser Stelle besteht wieder die Frage nach der grafischen oder textuellen Form. Für eine textuelle Form können die Begrifflichkeiten sich ebenfalls an Art. 4 und außerdem Art. 9 (1) DSGVO orientieren. Auch hierfür findet sich eine Taxonomie in Abschnitt 12.2.1. Die kompletten Begriffe lassen sich allerdings schlecht in den Symbolen unterbringen. Abkürzungen sind nicht offensichtlich und daher ebenfalls ungünstig. Es bietet sich daher eine grafische Darstellung an. Für die häufigsten Kategorien lassen sich hier recht einfach gut verständliche Piktogramme finden. Eine erste Anregung liefert [Bi23]. Um alle Kategorien abbilden zu können, müssen aber noch weitere Studien durchgeführt werden.

(a) Kennzeichnung eines gelben Datenobjekts (b) Kennzeichnung eines roten Datenobjekts

Abbildung 11.5.: Kennzeichnungen für Datenobjekte

Abbildung 11.5 zeigt erste Entwürfe für die zusätzlichen Markierungen an Datenobjekten inklusive einer Mouseover-Interaktion. Hierbei fällt allerdings auf, dass auf diese Art jeweils nur eine Datenart angegeben werden kann. In vielen Datenobjekten sind allerdings verschiedene Kategorien von Daten enthalten. Hier ist eine Umsetzung denkbar, in welcher ein spezielles Piktogramm für „mehrere Kategorien“ eingeführt wird und bei der Interaktion dann alle entsprechenden Kategorien angezeigt werden. Beim roten Datenobjekt in Abbildung 11.5b wird ein Kreis an Stelle eines Achtecks verwendet. Diese Option wird bereits in Abschnitt 11.3.1 thematisiert.

Die Darstellung lässt sich auf Datenspeicher übertragen.

11.3.4. Rechtssubjekt

Ein weiterer Aspekt, der visualisiert werden sollte, sind die beteiligten Rechtssubjekte. Für die Rechtssubjekte bietet sich eine dauerhafte Darstellung an, da in den Pools bzw. Lanes hinreichend viel Platz besteht, um eine entsprechende Darstellung einzufügen, ohne dabei die gesamte Darstellung zu überfrachten oder den Blick zu sehr auf diesen Aspekt zu lenken.

Auch hier stellt sich aber die Frage, ob eine grafische oder eine textuelle Darstellung gewählt werden sollte. Gut verständliche grafische Symbole sind wieder recht schwer zu entwickeln. Allerdings ist die Anzahl der abzubildenden Kategorien hier vergleichsweise klein (Betroffene, Verantwortliche, Auftragsverarbeiter, Dritte). Das erleichtert die grafische Darstellung, da auch bei nicht komplett selbsterklärenden Symbolen der Lernaufwand für den Nutzer entsprechend geringer ist, als bei deutlich höherer Anzahl von verschiedener Symbole. Für einen ersten Entwurf fällt die Wahl wegen des verfügbaren Platzes auf einen kombinierten Ansatz. In der Titel-Spalte des Pools wird jeweils ein Personen-Symbol und der Anfangsbuchstabe des jeweiligen Rechtssubjekts angezeigt. Hier wären alternativ auch komplett verschiedene Symbole denkbar.

Da die automatische Bestimmung des Rechtssubjekt recht komplex ist (insbesondere die Unterscheidung zwischen Auftragsverarbeiter und Dritten), wird außerdem ein Dropdown-Menü in den Pool eingefügt, in welchem die komplette Bezeichnung des Rechtssubjekts angegeben wird.

Abbildung 11.6.: Darstellung des Rechtssubjekts im Pool am Beispiel „Betroffener“

In Abbildung 11.6 wird die Idee zur Darstellung des Rechtssubjekts im Pool visualisiert. Das Vorgehen kann analog auch auf Lanes angewendet werden.

11.3.5. Weitere Aspekte

Als weitere Aspekte wurden genannt:

1. Spezieller Hinweis auf Risiko bei Anonymisierung
2. Aufbewahrungsfristen
3. Legende mit Abgrenzungen der Farben

Hier bieten sich folgende Ansätze an, die an dieser Stelle aber nicht weiter diskutiert werden:

1. An Aufgaben, die Anonymisierungsvorgänge darstellen und an anonymisierten Datenobjekten könnte ein Ausrufezeichen angebracht werden. Bei einem Mouseover könnte ein entsprechender Hinweis eingeblendet werden, dass die Anonymisierung kontrolliert werden muss.

2. Für Aufbewahrungsfristen könnte an den entsprechenden Datenobjekten eine Angabe in Jahren oder Monaten, z.B. in die obere linke Ecke, eingefügt werden. Auch hier wäre eine zusätzliche Erklärung bei Mouseover denkbar. Diese könnte sogar die Gesetzesgrundlage anzeigen.

3. Eine Legende könnte im Modellierungs- bzw. Betrachtungswerkzeug an einer Ecke angebracht werden. Es bietet sich hier eine Funktion zum Ein- und Ausblenden an, damit die Legende zwar bei Bedarf jederzeit einsehbar ist, für erfahrende Nutzer aber nicht unnötig den verfügbaren Platz einschränkt.

12. Kategorisierung von Prozesselementen

Die Kategorisierung der Modellelemente nach den in Kapitel 8 beschriebenen Maßgaben kann auf zwei verschiedene Arten geschehen.

Einerseits kann diese selbstverständlich manuell geschehen, also durch einen entsprechend versierten Menschen. Dies kann entweder diejenige Person sein, die einen Prozess modelliert oder aber beispielsweise auch der Datenschutzbeauftragte.

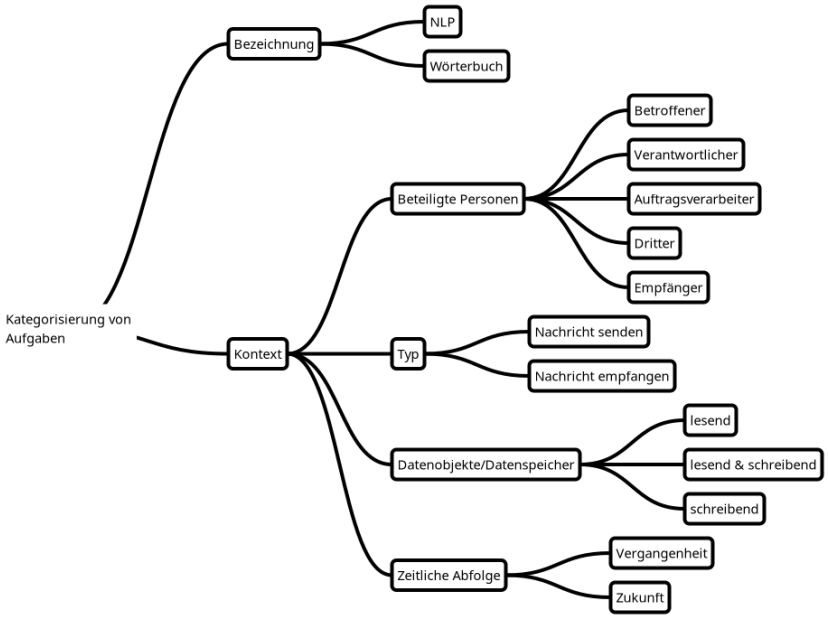


Abbildung 12.1.: Überblick über die wichtigsten Aspekte für die Datenschutzkategorisierung von Aufgaben

Andererseits bedeutet dies aber natürlich zusätzliche Arbeit für die entsprechenden Personen. Daher ist eine automatisierte Lösung wünschenswert, bei der für die Nutzer ein möglichst geringer zusätzlicher Aufwand entsteht (siehe dazu auch die Ergebnisse der Evaluation des Konzepts in Abschnitt 10.1). An dieser Stelle werden drei verschiedene Ansätze hierfür erläutert. In Teil III wird dann darauf aufbauend die Entwicklung eines Prototyps beschrieben, der zumindest eine Unterstützung bei der Kategorisierung

darstellt und diskutiert, welche Probleme dabei auftreten und inwiefern diese behoben werden können.

In Abbildung 12.1 werden die wichtigsten Aspekte dargestellt, die bei der Datenschutzkategorisierung von BPMN-Elementen betrachtet werden sollten. Die Abbildung spiegelt im Wesentlichen auch die Gliederung dieses Kapitels wider. Grundsätzlich kann zwischen der Bezeichnung des jeweiligen Elements (siehe Abschnitt 12.1) und kontextabhängigen Kriterien (siehe Abschnitt 12.3) unterschieden werden.

12.1. Kategorisierung über die Bezeichnung

Das offensichtlichste Kriterium zur automatisierten Kategorisierung ist die Bezeichnung des entsprechenden Modellelements. Auch bei der manuellen Kategorisierung ist diese der erste Anhaltspunkt.

Besonders gut möglich ist das bei den Datenobjekten. Eine Patientenakte wird – völlig Kontextunabhängig – in den allermeisten Fällen besondere Kategorien personenbezogener Daten enthalten. Eine Produktbeschreibung hingegen wird in aller Regel ohne personenbezogene Daten auskommen. Natürlich ist dies nicht immer so einfach. Ob ein „Vertrag“ oder eine „Akte“ Daten der besonderen Kategorien enthält, wird teilweise erst aus dem Kontext klar. Mit hoher Wahrscheinlichkeit wird beides aber zumindest personenbezogene Daten enthalten. Daher ist die Kategorisierung über die Bezeichnung hier durchaus ein betrachtenswertes Vorgehen.

Für Datenspeicher gilt grundsätzlich das Gleiche. Allerdings muss hier jeweils der aktuelle Zustand des Datenspeichers betrachtet werden (siehe Abschnitt 3.1.2). Hilfreich ist es hier, wenn der Zustand direkt angegeben ist. Hängt an einem Datenspeicher beispielsweise der Zusatz „[leer]“, enthält dieser mit ziemlicher Sicherheit keine (personenbezogenen) Daten und kann damit grün kategorisiert werden. Allerdings kann natürlich auch nicht aus jeder Statusangabe ein Rückschluss auf den konkreten Inhalt gezogen werden. So könnte der Status etwa auch „[gesichert]“ lauten um auszudrücken, dass ein digitaler Speicher z. B. passwortgeschützt oder ein Aktenschrank abgeschlossen ist. Hier ist es nun natürlich durchaus denkbar, dass der Grund für die Sicherung besonders schützenswerte Daten sind. Es kann sich aber auch schlicht um Geschäftsgeheimnisse handeln, welche unter Umständen überhaupt keine personenbezogene Daten enthalten.

Die Kategorisierung von Aufgaben nur auf Basis der Bezeichnung ist deutlich schwieriger. Die Beispiele in Kapitel 7 zeigen, dass gegebenenfalls zwei Aufgaben mit exakt der gleichen Bezeichnung unterschiedlich kategorisiert werden müssen. Offensichtlich ist die Kategorisierung hier also vom Kontext abhängig. Näheres zur kontextabhängigen Kategorisierung findet sich in Abschnitt 12.3. Nichtsdestotrotz ist die Analyse der

Bezeichnung einer Aufgabe ein erster Hinweis, da hierüber teilweise eine Datenverarbeitung ausgeschlossen werden kann.

Sprachlich besteht der Hauptunterschied von Aufgaben gegenüber Datenobjekten darin, dass diese in aller Regel mit mehreren Worten bezeichnet werden. Meist findet sich hier mindestens ein Substantiv und ein Verb im Infinitiv (z. B. “Formular ausfüllen“)[GL13]. Das Substantiv stellt hier gewöhnlich ein Akkusativobjekt und das Verb ein Prädikat dar. Die Bezeichnungen können aber durchaus auch noch komplexer werden, wenn beispielsweise noch weitere Objekte hinzukommen (siehe hierfür die Beispiele in Kapitel 7). Nun stellt sich die Frage, ob es für die Datenschutzbewertung ausreichend ist, einen Teil der Bezeichnung zu betrachten, oder ob immer alle Begriffe der Bezeichnung betrachtet werden müssen. Und falls alle Begriffe betrachtet werden sollten, stellt sich darüber hinaus die Frage, ob dies sinnvollerweise getrennt von einander geschehen sollte, oder ob sie in Kombination betrachtet werden müssen.

Da Aufgaben Verarbeitungstätigkeiten abbilden können, bietet es sich an, zunächst nur das Verb zu betrachten, da dieses der Verarbeitungstätigkeit entspricht. Hieraus müsste sich ableiten lassen, ob die Tätigkeit einer Verarbeitung entspricht. Es zeigt sich allerdings, dass durchaus das Objekt eine entscheidende Rolle spielt. Betrachtet man etwa die Aufgabe „Bewerbungsunterlagen vernichten“, handelt es sich offensichtlich um eine Verarbeitungstätigkeit, da das „Vernichten“ in Art. 4 Nr. 2 DSGVO explizit als solche genannt wird. Ein Prozess kann aber auch die Aufgabe „Produkt vernichten“ enthalten. In diesem Fall wird es sich höchstwahrscheinlich nicht um eine Verarbeitungstätigkeit handeln, da das Produkt keine personenbezogenen Daten enthält.

Der nächste Ansatz ist die Überlegung, ob statt des Verbs das Akkusativobjekt als Kriterium ausreicht. Denn nur wenn das Akkusativobjekt personenbezogene Daten abbildet, kann überhaupt eine Verarbeitungstätigkeit vorliegen. Teilweise stimmt das auch. Denn zumindest immer dann, wenn das Akkusativobjekt definitiv keine personenbezogene Daten enthält, kann die Aufgabe in die Kategorie „grün“ einsortiert werden. Wenn das Objekt aber personenbezogene Daten abbildet, muss noch unterschieden werden, ob die Aufgabe als gelb oder rot kategorisiert wird. Und hierfür ist wiederum das Verb entscheidend. Als übertriebenes Beispiel macht es etwa einen enormen Unterschied, die Bewerbungsunterlagen vernichtet oder veröffentlicht werden. Ersteres ist natürlich gelb zu kategorisieren, da hier sogar eine gesetzliche Verpflichtung besteht, letzteres hingegen bräuchte definitiv eine Einwilligung und wäre somit rot zu kategorisieren. Abgesehen davon ist aus dem Akkusativobjekt aber auch nicht immer herauszulesen, ob hier personenbezogene Daten enthalten sein können oder nicht. Als Beispiel diene hier etwa ein „Formular“, welches unausgefüllt keine personenbezogenen Daten enthält, ausgefüllt möglicherweise aber schon.

Letztlich muss noch beachtet werden, dass sich nicht alle Modellierer an die Bezeichnungskonvention halten. Teilweise werden etwa mehrere Aktivitäten in einer Aufgabe

verbunden, was dann separat betrachtet werden muss. Auch grammatikalisch werden in einigen Fällen Änderungen vorgenommen, sodass die Bezeichnung beispielsweise mit dem Verb beginnt.

Insgesamt müssen also definitiv alle Begriffe der Bezeichnung betrachtet werden, wobei einzelne Teile unter Umständen schon zur Kategorisierung ausreichen können.

12.1.1. Wörterbuch

Die technisch wohl einfachste Idee zur Klassifizierung über die Bezeichnung ist die Verwendung einer Art Wörterbuch. Der Begriff Wörterbuch soll an dieser Stelle eine Sammlung von Zeichenketten und der zugehörigen Kategorie bezeichnen.

Hier könnte für mögliche Bezeichnungen jeweils die entsprechende (Farb-) Kategorie gespeichert werden. Ein Klassifizierungsalgorithmus müsste die Kategorie dann gewissermaßen nur „nachschlagen“. Allerdings besteht hier die Problematik der Initialisierung dieses Wörterbuchs. Hierfür ist ein vergleichsweise hoher manueller Aufwand erforderlich.

Tabelle 12.1.: Wörterbuch für Datenobjekte

Begriff	**Kategorie**
Arbeitsvertrag	gelb
Personalfragebogen [blanko]	grün
Personalfragebogen [ausgefüllt]	rot
Personalfragebogen [SV-Nummer]	gelb
Personalfragebogen [SteuerID]	gelb
Lohnsteuermerkmale	rot
Einwilligung	gelb
Personalfragebogen [Geburtstag]	gelb
Kalender	gelb

Tabelle 12.1 stellt ein beispielhaftes Wörterbuch für Datenobjekte auf Basis des Einstellungsprozesses aus Abschnitt 9.1 dar. Für die praktische Umsetzung muss dies selbstverständlich deutlich umfangreicher sein und Bezeichnungen aus verschiedenen Prozessen enthalten.

Ein möglicher Ansatz zur Realisierung wird in Abschnitt 14.3.1 beschrieben.

12.1.2. Natural Language Processing (NLP)

Ein technisch deutlich komplexerer, aber für den Anwender weniger Aufwand verursachender Weg ist NLP (siehe Abschnitt 5.2). Hierbei wird mittels Machine Learning ein Modell entwickelt, welches die Kategorisierung übernimmt.

Grundlage für die Erstellung des Machine Learning Modells bildet eine Menge von Klassen, welche als Ausgabewerte fungieren sollen. Im vorliegenden Fall gibt es hierfür nun zwei Möglichkeiten:

Einerseits können einfach die drei definierten Farbklassen zur Anwendung kommen, sodass den jedem Prozesselement bzw. dessen Bezeichnung direkt eine Farbe zugewiesen wird. Ähnlich wird im Grunde in Abschnitt 14.3.1 verfahren. Allerdings ist ein zentraler Punkt beim NLP die (semantische) Ähnlichkeit von Begriffen. Diese ist bei den Farbkategorien aber nicht unbedingt gegeben. Beispielsweise müssten die beiden Datenobjekte „Patientenakte“ und „Lohnabrechnung“ beide rot eingefärbt werden, da sie besondere Kategorien personenbezogener Daten enthalten[1]. Die semantische Nähe zwischen den beiden Begriffen ist zumindest für den Menschen auf den ersten Blick aber eher nicht gegeben.

Im Hinblick auf die semantische Nähe bietet sich daher eine deutlich feingranularere Untergliederung der Klassen an. Diese kann in Form einer Taxonomie aufgebaut werden, welche unterschiedliche Klassen von Daten und Verarbeitungstätigkeiten strukturiert. Eine solche Taxonomie kann auch für den in Abschnitt 12.1.1 beschrieben Wörterbuch-Ansatz genutzt werden. Die erstellte Taxonomie wird im Folgenden erläutert.

12.2. Taxonomie

Die Taxonomie lässt sich grundsätzlich in drei Teile separieren, welche einerseits entsprechend der einzufärbenden Objekte Daten sowie Aktivitäten abbilden. Andererseits bietet es sich aber auch an, die Teilnehmer eines Prozesses abzubilden, in BPMN also Pools bzw. Swimlanes, um den Prozess möglichst umfassend beschreiben zu können. Grundsätzlich basieren alle drei Teile direkt auf der DSGVO.

12.2.1. Daten

Der Teil der Taxonomie, der die Daten betrifft, leitet sich unmittelbar aus der DSGVO ab. Genauer sind dies die Artikel 4:

> Im Sinne dieser Verordnung bezeichnet der Ausdruck [...] „personenbezogene Daten“ alle Informationen, die sich auf eine identifizierte oder

[1] In der Lohnabrechnung ist dies beispielsweise häufig die Religionszugehörigkeit, die für die Abführung der Kirchensteuer benötigt wird.

> identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann; [...]

und 9 (1):

> Die Verarbeitung personenbezogener Daten, aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person ist untersagt.

Die Taxonomie der Daten ist in Abbildung 12.2a dargestellt. Die Klassen dieser Taxonomie können nun unmittelbar auf die drei Farbkategorien abgebildet werden. Das Ergebnis ist in Abbildung 12.2b dargestellt.

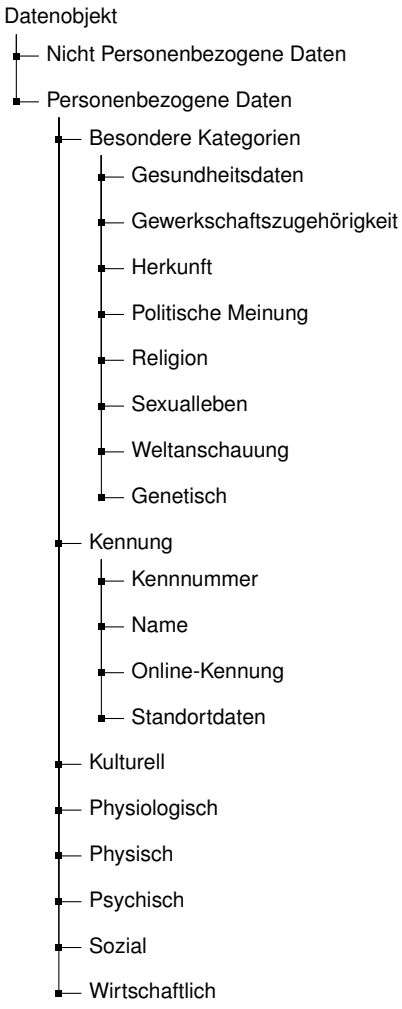
Zu beachten ist, dass *Besondere Kategorien* eine Kindklasse von *Personenbezogene Daten* ist. Es gilt also, dass alle Datenobjekte, die in Kindklassen von *Besondere Kategorien* (oder in die Klasse selbst) einsortiert werden, rot zu färben sind und alle Datenobjekte, die in eine der übrigen Kindklassen von *Personenbezogene Daten* einsortiert werden, gelb zu färben sind.

12.2.2. Teilnehmer

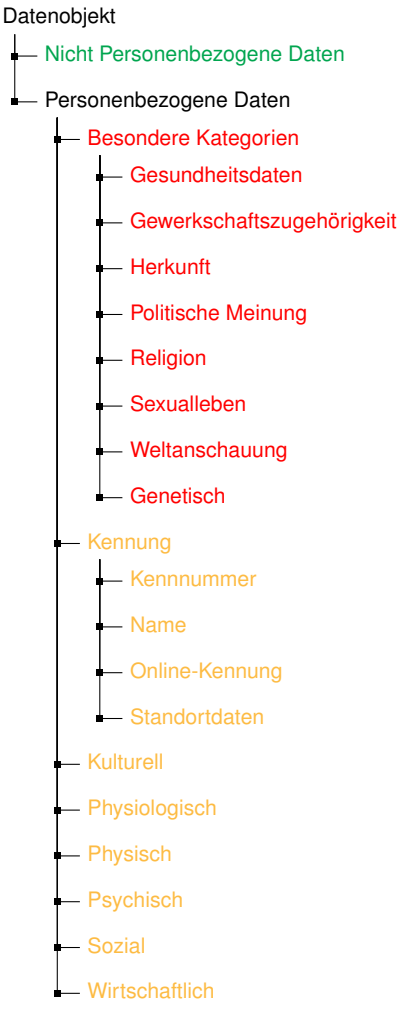
Auch die Taxonomie der Teilnehmer lässt sich direkt aus Art. 4 der DSGVO ableiten (siehe Abschnitt 2.1.5). Sie wird in Abbildung 12.3 dargestellt.

12.2.3. Aktivitäten

Bezüglich der Aktivitäten lassen die einzelnen Klassen der Taxonomie sich ebenfalls im Wesentlichen der Auflistung von Verarbeitungstätigkeiten in Artikel 4 DSGVO entnehmen, wie sie auch in Abschnitt 2.1.3 beschrieben werden. Die Gruppierung basiert im Wesentlichen auf [AK22]. Eine direkte Abbildung auf die Farbkategorien ist hier allerdings nicht möglich. Lediglich die Klasse *Nicht datenschutzrelevant* könnte



(a) Taxonomie für Daten



(b) Farbkategorien in der Datentaxonomie

Abbildung 12.2.: Taxonomie für Daten

offensichtlich grün markiert werden. Bei allen anderen Klassen ist die Färbung aber kontextabhängig.

Es ergibt sich die in Abbildung 12.4 abgebildete Taxonomie.

12.3. Kontextabhängige Kategorisierung

Der oben vorgestellte Ansatz richtet sich nur nach den Bezeichnungen der jeweiligen Elemente. Das ist allerdings nicht immer optimal, da zwei Modellelemente mit der gleichen Bezeichnung durchaus unterschiedliche Bedeutungen bzw. Datenschutzrelevanz

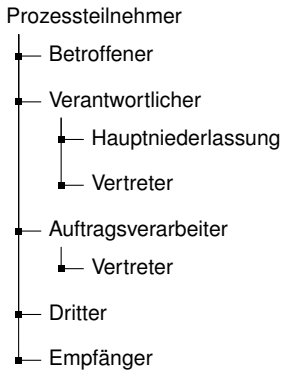


Abbildung 12.3.: Taxonomie für Prozessteilnehmer

haben können. Ein Beispiel hierfür findet sich in Abschnitt 9.2. Offensichtlich spielt also zumindest bei den Aufgaben der Kontext eine deutliche Rolle.

Dieser Kontext lässt sich in fünf Teilaspekte unterteilen, die im Folgenden näher betrachtet werden:

1. Die ausführende Person bzw. Organisationseinheit
2. Der Typ einer Aufgabe
3. Verknüpfte Datenobjekte
4. Verknüpfte Datenspeicher
5. Zeitliche Abfolge

Zu beachten ist hier, dass all diese Kriterien für sich allein betrachtet jeweils eher nicht zur Kategorisierung verwendet werden. Es viel mehr so, dass die Betrachtung zweier Kriterien häufig ein völlig unterschiedliches Ergebnis erzielt. Bei einigen Kriterien (insbesondere bei den verknüpften Datenobjekten und Datenspeichern) kann es darüber hinaus vorkommen, dass mehrere Möglichkeiten für eine Aufgabe in Frage kommen. Die einzelnen Kriterien sprechen also immer nur mit einer bestimmten Wahrscheinlichkeit für eine bestimmte Klasse und müssen in Kombination und zusätzlich mit der Bezeichnung betrachtet werden.

Beteiligte Personen

Neben der Art der Verarbeitung ist auch relevant, wer diese ausführt und welche weiteren Prozessteilnehmer an einer Aufgabe beteiligt sind. Unterschieden wird hier zwischen den fünf Rollen, die in Abschnitt 2.1.5 erläutert werden.

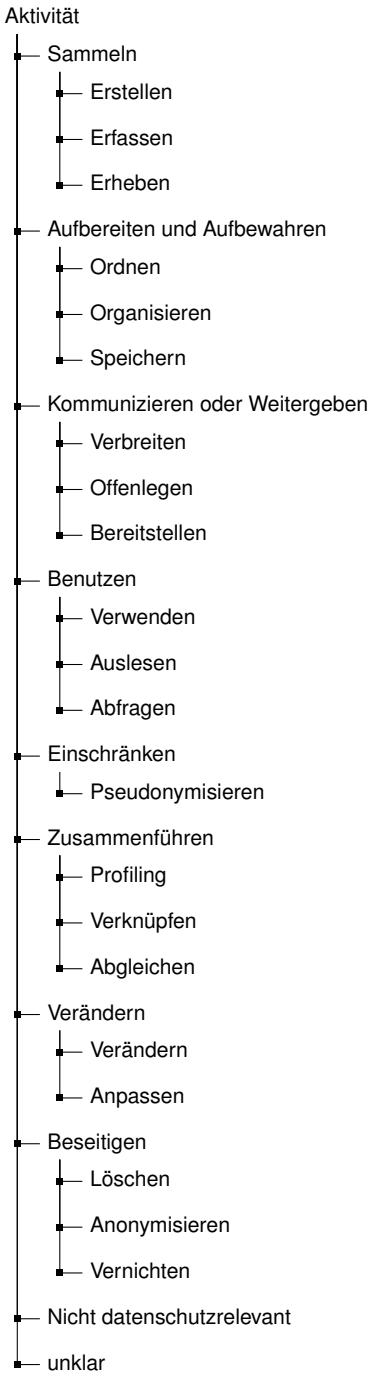


Abbildung 12.4.: Taxonomie für Aktivitäten

Betroffene

Verarbeitungstätigkeiten, die die betroffene Person selbst ausführt, sind datenschutzrechtlich im Wesentlichen irrelevant, da es sich hier um natürliche Personen handelt, für welche die entsprechenden Tätigkeiten in der Regel ausschließlich dem persönlichen Bereich zugeordnet werden können. Nach Art. 2 (2) c) DSGVO sind diese nicht Gegenstand der Verordnung. Daher können grundsätzlich alle Aktivitäten, die die betroffene Person ausführt, mit der Farbe grün markiert werden. Zu beachten ist hier allerdings, dass in einem Prozess unter Umständen mehrere betroffene Personen existieren können. Dann ist nicht mehr offensichtlich, ob in einer Aktivität, in der personenbezogene Daten verarbeitet werden, die ausführende Person auch die betroffene Person ist. Das ist allerdings eher selten der Fall, da die meisten Prozessmodelle eine Interaktion zwischen einem Unternehmen (oder einer sonstigen juristischen Person) und einer einzelnen natürlichen Person (z. B. Kunde, Mitarbeiter, Bewerber, Patient) abbilden. Auf die Daten der juristischen Person findet die DSGVO gemäß Erwägungsgrund 14 keine Anwendung.

Verantwortlicher

Der Verantwortliche kann grundsätzlich alle Arten von Aufgaben ausführen und wird in der Regel wohl im Fokus der Modellierung stehen und somit auch der Teilnehmer mit den meisten Aufgaben sein.

Auftragsverarbeiter

Auftragsverarbeiter sind daher interessant, da keine gesonderte Rechtsgrundlage für die Datenverarbeitung durch diesen notwendig ist[EMS19]. Demnach kann in einem Auftragsverarbeiter-Pool nie eine rote Aufgabe liegen, da ohne notwendige Rechtsgrundlage natürlich auch nie eine Einwilligung benötigt wird.

Dritter

Dritte sind all diejenigen, die nicht Betroffene, Verantwortliche oder Auftragsverarbeiter sind. Dieser Umstand ist insbesondere bei der Weitergabe personenbezogener Daten und damit bei gleichzeitiger Rolle als *Empfänger* relevant.

Empfänger

Schließlich spielt auch noch die Person, welche eine Aktivität ausführt, eine entscheidende Rolle. Besonders relevant wird diese, wenn nur die eigenen Daten verarbeitet werden, die verarbeitende Person also der betroffenen Person entspricht. Denn dann ist die Datenverarbeitung in jedem Fall irrelevant.

Außerdem ist bei allen Aktivitäten aus den Klassen der Kategorie *Kommunizieren oder Weitergeben* der jeweilige Empfänger relevant. Insbesondere ist hier zu beachten, ob die Weitergabe der Daten intern geschieht, also innerhalb des gleichen Unternehmens, oder aber eine externe Stelle der Empfänger der Daten ist, da nur für die Weitergabe an Dritte eine Rechtsgrundlage vorliegen muss. Zu erkennen ist dies häufig daran, dass für jedes Unternehmen ein Pool modelliert wird. Es kann aber auch sein, dass beispielsweise für verschiedene Abteilungen verschiedene Pools genutzt werden. Hier muss dann auf andere Weise entschieden werden, ob es sich um einen internen oder externen Empfänger handelt. Eine Möglichkeit hierfür wird in Kapitel 11 dargestellt.

Typ der Aufgabe

Aufgaben können neben ihrer Bezeichnung auch durch einen vergebenen Typ beschrieben werden (siehe Abschnitt 3.1.1). Hieraus lassen sich Rückschlüsse darauf ziehen, in welche Kategorie die Aufgabe einzuordnen ist. So ist beispielsweise eine *Sende-Aufgabe* häufig in die Kategorie *Kommunizieren oder Weitergeben* einzuordnen, während ein eine *Empfangs-Aufgabe* eher in die Kategorie *Erheben* einzuordnen ist.

Voraussetzung hierfür ist offensichtlich die Verwendung dieser konkreten Typen von Aufgaben. In der Praxis wird allerdings in vielen Fällen überwiegend oder sogar ausschließlich mit abstrakten Aufgaben gearbeitet. Dann kann aber trotzdem oft aus dem Kontext erkannt werden, dass es sich um eine sendende (oder empfangende) Nachricht handelt, da hier Nachrichtenflüsse zwischen zwei Pools ausgetauscht werden.

Verknüpfte Datenobjekte/-speicher

Eine weitere Rolle spielen die angehängten Datenobjekte und Datenspeicher. Hier ist zunächst zu betrachten, ob diese von einer Aufgabe gelesen oder erstellt bzw. bearbeitet werden. Erkannt werden kann dies an der Richtung des Datenflusses: Verläuft dieser von dem Datenobjekt bzw. -speicher zu der Aktivität, ist von einem lesenden Zugriff auszugehen, bei der entgegengesetzten Richtung von einem schreibenden. Darüber hinaus kann es auch vorkommen, dass zwischen einer Aufgabe und einem Datenobjekt bzw. -speicher sowohl ein lesender als auch ein schreibender Zugriff besteht. Dies sollte in der Regel bedeuten, dass die Daten verändert werden. Letztlich ist auch noch denkbar, dass mehrere Datenobjekte und/oder -speicher mit einer Aufgabe verknüpft sind. Wenn hier entweder ausschließlich lesende oder ausschließlich schreibende Zugriffe stattfinden, können diese analog zu dem entsprechenden Fall mit nur einem Datenelement bewertet werden. Wenn allerdings auf (mindestens) ein Element lesender Zugriff und auf (mindestens) ein weiteres schreibender Zugriff besteht, bedeutet das in aller Regel, dass Daten aus einem Datenobjekt bzw. -speicher extrahiert und in ein anderes Datenobjekt bzw. einen anderen Datenspeicher übertragen werden. Das kann beispielsweise bedeuten,

dass Daten veröffentlicht werden, aber auch eine reine Umorganisation/-strukturierung ist denkbar.

Tabelle 12.2.: Mögliche Klassen einer Aufgabe bei angehängten Datenobjekten

Klasse	**L**	**S**	**Bidirektional**	**L & S**
Erstellen		x		
Erfassen	x	x		x
Erheben	x			x
Ordnen			x	x
Organisieren			x	x
Speichern		x		
Verbreiten	x			x
Offenlegen	x			x
Bereitstellen				x
Verwenden	x			
Auslesen	x			
Abfragen	x			x
Pseudonymisieren			x	
Profiling	x			
Verknüpfen	x		x	
Abgleichen	x			
Verändern			x	
Anpassen			x	
Löschen	x		x	
Anonymisieren			x	
Vernichten	x		x	

L = lesender Zugriff;
S = schreibender Zugriff;
B = bidirektionaler Zugriff auf ein Datenobjekt;
L&S = lesender und schreibender Zugriff auf verschiedene Datenobjekte

Bei der Klasse *Erstellen* ist zu beachten, dass hier insbesondere dann eine hohe Wahrscheinlichkeit für diese Art der Verarbeitungstätigkeit vorliegt, wenn das Datenobjekt bzw. der Datenspeicher (oder ein anderes entsprechendes Element mit der gleichen Bezeichnung) im Prozess bisher noch nicht auftaucht. Bei den Klassen *Löschen* und *Vernichten* gilt entsprechend eine höhere Wahrscheinlichkeit für deren Vorliegen, wenn das Element im weiteren Verlauf des Prozesses nicht mehr auftaucht.

Außerdem kann hier auch noch der Status eines Datenobjekts genutzt werden. Wenn bei einer Aufgabe beispielsweise ein Datenobjekt mit dem Status [leer], [blanko], [unausgefüllt] etc. verwendet wird und dabei ein Datenobjekt mit der gleichen Be-

zeichnung aber einem anderen Status erzeugt wird, liegt es nahe, dass es sich etwa um die Verarbeitungsklassen *Erfassen* oder *Erheben* handelt. Im Gegenzug kann es sich bei entgegengesetzten Datenflüssen um die Klassen *Löschen* oder *Vernichten* handeln. Wenn der Status eines ausgehenden Datenobjekts beispielsweise [anonymisiert] oder [pseudonymisiert] lautet, handelt es sich bei erstellenden Aufgabe um die Klassen *Anonymisieren*, respektive *Pseudonymisieren*.

Zeitliche Abfolge

Neben den bereits erwähnten Aspekten kann auch noch der zeitliche Ablauf innerhalb des Prozessmodells eine Hilfestellung bei der automatisierten Bewertung des Datenschutzes darstellen.

Beispielsweise werden häufig Begrifflichkeiten im Prozess mehrfach verwendet und eventuelle Abkürzungen an anderen Stellen ausgeschrieben. Daher macht es bei Unklarheiten Sinn, andere Aktivitäten zu betrachten.

Recht offensichtlich verhält es sich, wenn in einer Aufgabe eine Einwilligung eingeholt wird. Dann liegt es nahe, dass in einer der folgenden Aufgaben diese auch benötigt wird und diese Aufgabe somit rot zu kategorisieren ist. In vielen Fällen wird es sich hier um die folgende Aufgabe handeln.

Teilweise können auch aus zukünftigen Abläufen Rückschlüsse auf die Verarbeitungsgrundlage und somit auch die Kategorie einer Aufgabe gezogen werden. Als Beispiel kann hier ein Prozess dienen, in welchem in einer Aufgabe etwa die Postadresse eines Kunden abgefragt wird. Nun ist zunächst unklar, ob hierfür eine entsprechende Verarbeitungsgrundlage vorliegt. Wenn aber in einer späteren Aufgabe eine Ware an den Kunden versendet wird, dann wird die Adresse zur Vertragserfüllung benötigt und infolgedessen kann die Aufgabe, in der die Adresse abgefragt wird, gelb kategorisiert werden.

12.3.1. Kriterien für Datenobjekte

Bei Datenobjekten ist die Klassifizierung grundsätzlich einfacher. In vielen Fällen reicht hier die reine Betrachtung der Bezeichnung aus. Ein weiterer wichtiger Aspekt ist allerdings der Status eines Datenobjekts. So macht es bei einem Formular für die Datenschutzbetrachtung etwa häufig einen Unterschied, ob dieses leer oder ausgefüllt ist. Beispiele hierfür finden sich auch in Kapitel 9. Der Status sollte also grundsätzlich mit in die Bewertung einfließen.

Allerdings wird dieses Mittel häufig gar nicht benutzt, also einfach generell kein Status vergeben. Hier können dann aber eventuell entsprechende Rückschlüsse aus den erstellenden oder verwendenden Aufgaben gezogen werden. Wenn eine Aufgabe beispielsweise der Klasse *Anonymisieren* zugeordnet wird, ist es ziemlich naheliegend,

dass einem ausgehenden Datenobjekt anschließend der Status [anonymisiert] und somit die Farbe Grün zugewiesen werden kann. Darüber hinaus kann dann aber auch davon ausgegangen werden, dass in einem eingehenden Datenobjekt auf jeden Fall personenbezogene Daten enthalten sind, sonst ergäbe das Anonymisieren keinen Sinn. Beachtet werden muss hier immer, ob die Anonymisierung auch korrekt ausgeführt wurde (siehe hierzu auch Abschnitt 10.1).

In Abschnitt 12.3 wurden bereits die Implikationen vom Status eines Datenobjekts auf die verbundenen Aufgaben beschreiben. Die dort beschriebenen Fälle gelten hier analog. Zu beachten ist allerdings, dass keine Zirkelschlüsse auftreten. Es sollte also klar definiert werden, welches Kriterium angewendet wird. Das lässt sich aber recht einfach erreichen, wenn zu Beginn einmalig überprüft wird, ob ein Status vorhanden ist. Wenn dies nicht der Fall ist, kann der Status als Kriterium für die Klassifizierung der Aufgabe herangezogen werden, wenn dem nicht so ist, kann die Aufgabe als Kriterium für den Status herangezogen werden.

Teil III.

Prototyp

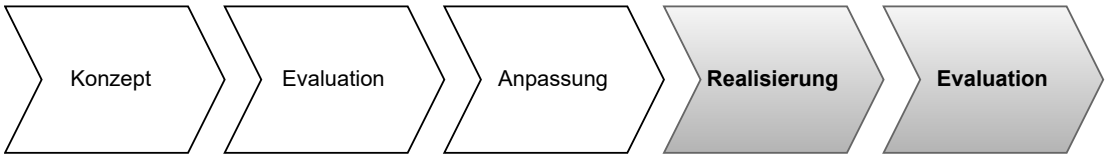


Abbildung 12.5.: Im Teil „Realisierung“ betrachtete Schritte des Vorgehens der Arbeit

Dieser Teil beschreibt die Realisierung des Konzepts aus Teil II. Dies betrifft die letzten beiden Schritte des in Abschnitt 1.2 erläuterten Vorgehens: Die eigentliche Realisierung und die anschließende Evaluation des erstellen Prototyps.

Hierfür werden in Kapitel 13 zunächst einige funktionale und nichtfunktionale Anforderungen an den Prototyp definiert.

Anschließend werden in Kapitel 14 die Grundzüge der eigentlichen Implementierung beschrieben. Der Fokus liegt hier auf dem verwendeten System, der Erweiterung des BPMN-Standards und der automatischen Kategorisierung der Modellelemente.

Abschließend wird der Prototyp in Kapitel 15 mit Bezug auf die zuvor definierten Anforderungen evaluiert.

13. Anforderungen

Die Grundlage für jeden Softwareentwicklungsprozess ist eine detaillierte Anforderungsanalyse[So16]. Dieses Kapitel beschreibt die Anforderungen an ein System, welches das in Kapitel 8 eingeführte Konzept nutzbar macht, aus der Sicht eines potentiellen Kunden bzw. Anwenders. Die inhaltliche Grundlage hierfür bilden in erster Linie viele Gespräche mit Personen aus Forschung und Anwendung, wie beispielsweise den Gesprächspartnern, die in Abschnitt 10.1 vorgestellt wurden. Die Anforderungsanalyse bezieht sich nur auf das Kernkonzept der Einfärbung. Die in Kapitel 11 beschriebenen Erweiterungen werden nicht näher betrachtet, da hier für eine sinnvolle Analyse der Anforderungen zunächst weitere Gespräche geführt werden müssen.

Formal orientiert sich das Vorgehen und die Gliederung des Kapitels an den Vorschlägen von Andrea Herrmann in [He22], die sich bezüglich der Gliederung wiederum auf die Erläuterungen des entsprechenden IREB[1]-Standards in [PR15] bezieht. Auf einige vorgeschlagene Abschnitte, wie beispielsweise einen eigenen Anhang oder einen Index wird allerdings auf Grund des gegebenen Kontextes dieser Arbeit verzichtet. Das Vorgehen für die Anforderungsanalyse orientiert sich außerdem an den Ausführungen zum IREB-Standard in [Gl20].

13.1. Einleitung

An dieser Stelle werden zunächst die allgemeinen Ziele des Projekts und der geplante Umfang des zu entwickelnden Systems erläutert.

13.1.1. Projektziele und -zweck

Wie in Kapitel 1 erläutert wird, besteht eine große Herausforderung von Unternehmen und anderen Organisationen darin, Datenschutzprobleme in den eigenen Abläufen zu erkennen und anschließend intern zu kommunizieren, damit diese behoben werden können. Teil II dieser Arbeit stellt ein Konzept zur Visualisierung von Datenschutzaspekten in Geschäftsprozessmodellen vor, welches die beteiligten Personen im Unternehmen bei der oben beschriebenen Problematik unterstützen könnte.

[1] International Requirements Engineering Board (`https://www.ireb.org`)

Generell ist es als sinnvoll zu betrachten, das beschriebene Konzept in Form einer Software umzusetzen, die insbesondere eine automatische Einfärbung von Geschäftsprozessmodellen ermöglicht. Das haben verschiedene Gespräche ergeben (siehe Abschnitt 10.1).

Eine fertige Software kann aber nicht das Ziel dieser Arbeit sein. Viel mehr soll hier zur besseren Demonstration und weiteren Evaluation des Konzepts ein erster Prototyp entwickelt werden. Dieser kann – bei entsprechender Evaluation – in der Zukunft auch als eine Art erstes Inkrement eines agilen Softwareentwicklungsprozesses mit dem Ziel eines fertigen Produkts betrachtet werden.

13.1.2. Systemumfang

Der zu entwickelnde Prototyp soll es ermöglichen, Geschäftsprozesse zu in BPMN zu modellieren und sowohl manuell als auch automatisiert einzufärben. Einzufärbende Modellelemente sollen Aktivitäten, Datenobjekte und Datenspeicher sein. Für die manuelle Einfärbung müssen mindestens die Farben Rot, Gelb und Grün auswählbar sein. Außerdem sollen Modellelemente auch wieder in ihren farblosen Ursprungszustand zurückversetzt werden können. Die automatisierte Einfärbung soll sich auf die Datenschutzrelevanz der entsprechenden Modellelemente beziehen und den Kriterien aus Kapitel 8 folgen.

13.2. Übersicht

Dieser Abschnitt erläutert die zunächst die Rahmenbedingungen, die bei der Entwicklung betrachtet werden müssen. Hierfür werden Anforderungen an die Systemarchitektur, den Systemkontext und die zu erwartenden Nutzergruppen dargestellt. Anschließend werden die Kernfunktionalitäten aufgeführt, die das System haben soll.

13.2.1. Systemarchitektur

Die Systemarchitektur kann flexibel gewählt werden. Wichtig ist die Möglichkeit einer plattformübergreifenden Nutzung. So sollten mindestens die Betriebssysteme Windows, macOS und Linux, sowohl in den jeweils aktuellsten[2] als auch in den aktuell verbreitetsten Versionen unterstützt werden.

[2] im Januar 2023

13.2.2. Systemkontext, Randbedingungen, Annahmen

Da das Ziel der Entwicklung nur ein erster Prototyp als Proof of Concept ist, werden explizit keine Anforderungen bezüglich Stabilität, Verfügbarkeit usw. gestellt.

13.2.3. Nutzer und Zielgruppen

Das System hat vier wesentliche Ziel- bzw. Anwendergruppen. Im Folgenden findet sich für jede dieser Gruppen je eine kurze Beschreibung ihrer generellen Aufgaben, der Ziele, die die Einführung des Systems mit sich bringen soll, sowie ihres relevanten Vorwissens. Diese Anforderungsanalyse soll nicht die Perspektive eines konkreten Unternehmens darstellen, sondern die einer möglichst großen Menge potenzieller Anwender abdecken. Daher kann an dieser Stelle aber keine Auflistung erstellt werden, die sowohl vollständig ist als auch auf jeden einzelnen potentiellen Nutzer genau in dieser Form zutrifft. Insbesondere bei großen Unternehmen sind einige beschriebene Rollen sicherlich weiter unterteilt, während in kleineren Organisationen möglicherweise eine Person mehrere Rollen einnimmt.

1. Prozessmanager

Der Prozessmanager ist insbesondere für die Modellierung der Prozesse und deren Optimierung verantwortlich.

Ziele

- Einfache Modellierung
- Verbesserte Kommunikation mit anderen Beteiligten
- Deutlicher Hinweis auf Datenschutzprobleme

Wissen/Erfahrung/Fähigkeiten

Tabelle 13.1.: Kenntnisse des Prozessmanagers

Bereich	**Kenntnisse**
Prozessmodellierung/-management	umfangreich
Aufgaben Fachabteilungen	gering
Datenschutz	gering

2. Datenschutzbeauftragter

Der Datenschutzbeauftragte berät die Unternehmensleitung und die Beschäftigten bezüglich des Datenschutzrechts und überwacht dessen Einhaltung. Außerdem sensibilisiert er die Beschäftigten und führt Schulungen zum Thema Datenschutz durch (Art. 39 DSGVO).

Ziele

- Einfaches und schnelles Verständnis der Prozesse und Datenflüsse im Unternehmen
- Verbesserte Kommunikation mit allen Beteiligten
- Lösung von Datenschutzproblemen
- Verbesserte und vereinfachte Dokumentation

Wissen/Erfahrung/Fähigkeiten

Tabelle 13.2.: Kenntnisse des Datenschutzbeauftragten

Bereich	**Kenntnisse**
Prozessmodellierung/-management	gering
Aufgaben Fachabteilungen	gering
Datenschutz	umfangreich

3. Unternehmensleitung

Die Unternehmensleitung ist einerseits für die grundsätzliche Ausgestaltung der Prozesse im Unternehmen verantwortlich. Andererseits ist sie aber auch für die Einhaltung der Datenschutzregularien verantwortlich. Unter Umständen kann die Unternehmensleitung (z.B. in Form des Geschäftsführers) neben dem Unternehmen an sich sogar als „Verantwortlicher“ im Sinne von Art. 4 Ziffer 7 DSGVO gelten und könnte damit für Datenschutzverstöße haftbar gemacht werden[Di22; OL21].

Ziele

- Überblick über Prozesse im Unternehmen
- Grober Überblick über Datenschutzprobleme

- Vereinfachte Kommunikation mit anderen Beteiligten
- Rechtliche Absicherung durch Dokumentation

Wissen/Erfahrung/Fähigkeiten

Tabelle 13.3.: Kenntnisse der Unternehmensführung

Bereich	Kenntnisse
Prozessmodellierung/-management	mittel
Aufgaben Fachabteilungen	mittel
Datenschutz	mittel

4. Prozessbeteiligte

Als Prozessbeteiligte werden an dieser Stelle alle Mitarbeiter eines Unternehmens bezeichnet, die an den Prozessen im Unternehmen mitwirken bzw. diese ausführen. Diese werden hier zusammengefasst. Es muss aber natürlich unterschieden werden zwischen den Beteiligten der einzelnen Prozesse. Insbesondere kann von einem Beteiligten eines Prozesses A nicht unbedingt auch (Detail-)Kenntnis eines Prozesses B vorausgesetzt werden.

Ziele

- Übersichtliche Handlungsanweisung
- Einfache Informationen, wie Datenschutzbestimmungen eingehalten werden können

Wissen/Erfahrung/Fähigkeiten

Tabelle 13.4.: Kenntnisse der Prozessbeteiligten

Bereich	Kenntnisse
Prozessmodellierung/-management	gering
Aufgaben Fachabteilungen	hoch (für den jeweiligen Bereich)
Datenschutz	gering

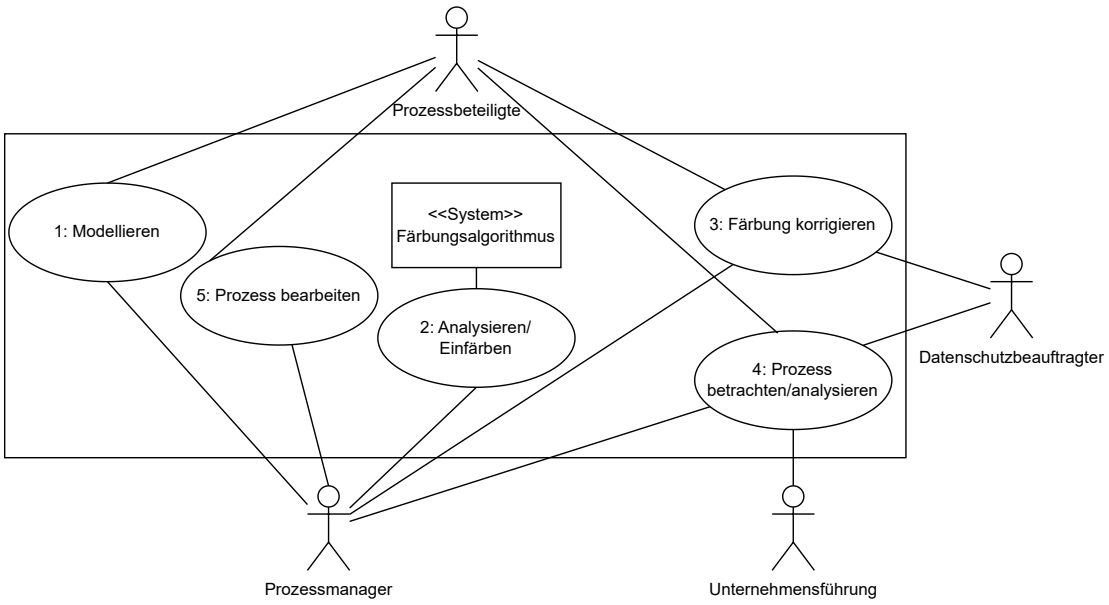


Abbildung 13.1.: Use Case Diagramm des zu entwickelnden Prototyps

13.2.4. System-Funktionalität

Die wichtigsten Anwendungsfälle des Prototyps werden als Use Case Diagramm basierend auf der Notation der Unified Modeling Language (UML) in Abbildung 13.1 dargestellt. Im Folgenden werden diese kurz aus Sicht der beteiligten Nutzer erläutert. Die Beschreibung der Anwendungsfälle in der Ich-Form orientiert sich in ihrer Form an den Beispielen aus [He22]. Eine ausführlichere Darstellung folgt in Abschnitt 13.3.2.

Use Case 1: Modellieren

Als **Prozessmanager** möchte ich die Prozesse des Unternehmens anschaulich in BPMN modellieren können.

Als **Prozessbeteiligter** möchte ich die Prozessmodellierung durch mein Domänenwissen unterstützen, um meine Prozesse möglichst korrekt dokumentiert zu bekommen.

Use Case 2: Analysieren/Einfärben

Als **Prozessmanager** möchte ich datenschutzkritische Bereiche in meinen modellierten Prozessen automatisiert identifizieren und kennzeichnen.

Use Case 3: Färbung korrigieren

Als **Datenschutzbeauftragter** möchte ich die Korrektheit der Einfärbung der Prozesse sicherstellen.

Als **Prozessmanager** möchte ich mit Hilfe des Datenschutzbeauftragten die Färbung der Prozesse anpassen, falls Fehler vorliegen.

Use Case 4: Prozess betrachten/analysieren

Als **Prozessmanager** möchte ich alle Prozesse im Unternehmen analysieren und bei Bedarf optimieren. Hierbei können verschiedenste Kriterien zugrunde liegen.

Als **Datenschutzbeauftragter** möchte ich die Prozesse im Unternehmen bezüglich des Datenschutzes optimieren. Außerdem möchte ich Prozessmodelle nutzten, um kritische Aspekte anderen Stellen zu erläutern. Des Weiteren möchte ich die Geschäftsprozessmodelle nutzen, um meinen Dokumentationspflichten nachzukommen.

Als **Unternehmensleitung** möchte ich einen Überblick über alle Prozesse im Unternehmen erlangen. An Hand der Prozessmodelle möchte ich mögliche Optimierungen mit dem Prozessmanager besprechen. Außerdem möchte ich auf den ersten Blick sehen, wie es um den Datenschutz im Unternehmen bestellt ist.

Use Case 5: Prozess bearbeiten

Als **Prozessmanager** möchte ich die modellierten Prozesse verändern können, um sie zu optimieren oder Fehler zu korrigieren.

Als **Prozessbeteiligter** möchte ich die Prozessbearbeitung durch mein Domänenwissen unterstützen, um meine Prozesse möglichst korrekt dokumentiert zu bekommen.

Mit Blick auf das Prozessmanagement existieren noch weitere Anwendungsfälle, wie etwa die Optimierung des Prozesses. Hieraus ergeben sich aber keine zusätzlichen Anforderungen an das System, weshalb hierauf nicht näher eingegangen wird.

13.3. Funktionale Anforderungen

Dieser Abschnitt beschreibt die funktionalen Anforderungen an das zu entwickelnde System aus verschiedenen Perspektiven. Nach einer kurzen Beschreibung der Strukturperspektive wird hier umfassend die Funktionsperspektive erläutert.

13.3.1. Strukturperspektive (fachliches Datenmodell)

Um die oben erläuterten Anwendungsfälle abzubilden, ist kein wirkliches Datenmodell als Grundlage einer Datenbank für die persistente Speicherung notwendig. Die persistente Speicherung der Prozessmodelle erfolgt in Form einer .bpmn-Datei, welche einem bestimmten XML-Schema folgt (siehe Abschnitt 3.1.5). Dieses muss an einigen Stellen um ein zusätzliches Attribut erweitert werden, welches die Farbe der Modellelemente beinhaltet. Darüber hinaus muss ein Weg gefunden werden, die Daten zu speichern, welche für die automatisierte Klassifizierung der Modellelemente notwendig sind. Das Datenmodell hängt hier aber von dem Verfahren zur Klassifizierung ab und kann daher an dieser Stelle noch nicht definiert werden.

13.3.2. Funktionsperspektive

Dieser Abschnitt beschreibt die Funktionsperspektive, also die konkreten Funktionalitäten der einzelnen Anwendungsfälle. Für jeden Anwendungsfall werden die wichtigsten Eckdaten in tabellarischer Form und der Ablauf in Diagrammform angegeben. Abweichend vom üblichen Vorgehen werden hierfür keine UML Aktivitätsdiagramme, sondern BPMN Prozessmodelle verwendet, da diese Notation ein Kernthema der Arbeit darstellt.

Use Case 1: Modellieren

Tabelle 13.5.: Use Case 1

Beschreibung	Modellieren
Priorität	mittel
Hauptakteur	Prozessmanager
Weitere Akteure	Prozessbeteiligte
Vorbedingungen	Auftrag zur Modellierung des Prozesses ist im Prozessmanagement eingegangen; Prozessmanagement hat Beschreibung von Prozessbeteiligten angefordert.
Auslösendes Ereignis	Prozessbeschreibung geht bei Prozessmanagement ein.
Ergebnis	Prozessmodell ist fertig, korrekt modelliert und persistent gespeichert.
Nachbedingung	Anforderer (z.B. Unternehmensleitung) bekommt Information über Fertigstellung.

Tabelle 13.5 gibt einen Überblick über die wichtigsten Eckpunkte des Anwendungsfalls *Modellieren*. Der Anwendungsfall ist relativ unkritisch für das Gesamtsystem, weil generell einige entsprechende Anwendungen auf dem Markt verfügbar sind (siehe auch Abschnitt 14.1.1) und BPMN ein gut standardisiertes Austauschformat bietet. Somit könnten Prozesse auch mit einem anderen Werkzeug modelliert und in das hier zu entwickelnde System importiert werden. Allerdings ist eine umfassende Lösung für alle Arbeitsschritte in der Regel komfortabler für den Anwender und sollte auch einen geringeren Administrationsaufwand darstellen. Daher wird dem Anwendungsfall die Priorität *mittel* zugeordnet.

Der grobe Ablauf des Anwendungsfalls ist in Abbildung 13.2 abgebildet. Auf einzelne offensichtliche Prozessschritte, wie etwa das Speichern des Prozesses, wird aus Gründen der Übersichtlichkeit verzichtet.

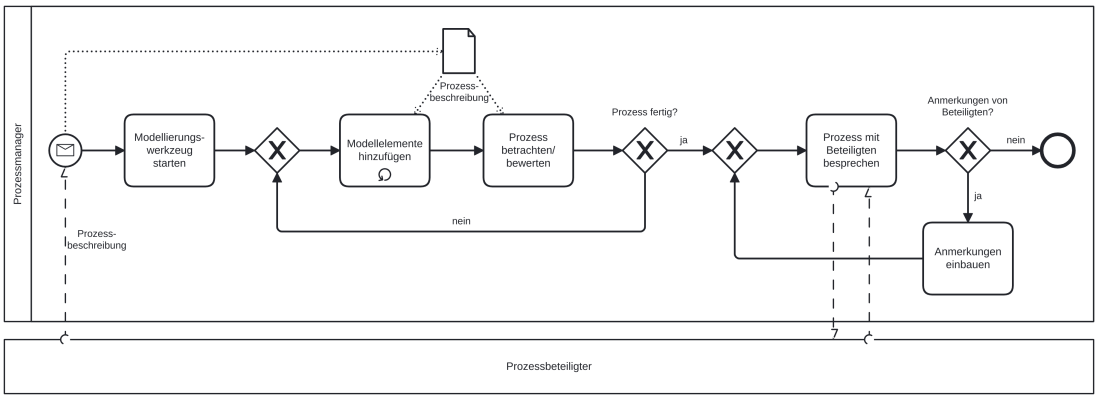


Abbildung 13.2.: BPMN Diagramm zu Use Case 1: Modellieren

Der Prozess beginnt mit einer Nachricht der Prozessbeteiligten, die eine Prozessbeschreibung enthält. Zuvor muss natürlich ein entsprechender Auftrag an die Prozessbeteiligten erfolgt sein, diese Beschreibung zu erstellen. Dies ist aber nicht Bestandteil des Anwendungsfalls.

Nach Eingang der Nachricht startet der Prozessmanager das Modellierungswerkzeug und beginnt mit der Modellierung. Hierzu werden nach und nach immer mehr Prozesselemente zum Prozess hinzugefügt. Zwischendurch betrachtet der Prozessmanager den Prozess und überprüft, ob noch Elemente fehlen oder der Prozess - seiner Meinung nach - bereits vollständig ist.

Wenn der initiale Modellierungsvorgang abgeschlossen ist, bespricht der Prozessmanager sein Ergebnis mit den Prozessbeteiligten. Werden hierbei Probleme gefunden, wird der Prozess entsprechend korrigiert und anschließend erneut den Beteiligten vorgelegt.

Sobald keine Probleme mehr gefunden werden, endet der Modellierungsprozess.

Use Case 2: Analysieren/Einfärben

Der zweite Anwendungsfall *Analysieren/Einfärben* bildet die eigentliche Kernfunktionalität des Systems ab.

Wie in Tabelle 13.6 zu sehen, erhält er daher die Priorität *hoch*. Der einzige beteiligte Akteur ist wie schon im vorhergehenden Anwendungsfall der *Prozessmanager*. Der Anwendungsfall kann auf dessen Initiative ausgelöst werden, sobald ein Prozessmodell im System vorhanden ist. Hierbei ist es unerheblich, ob dieses im System modelliert oder von einer externen Quelle importiert worden ist. Ziel des Anwendungsfalls ist ein eingefärbtes Prozessmodell, welches im Folgenden unbedingt noch kontrolliert werden muss.

Tabelle 13.6.: Use Case 2

Beschreibung	Einfärben
Priorität	hoch
Hauptakteur	Prozessmanager
Weitere Akteure	–
Vorbedingungen	Prozessmodell in BPMN ist im System vorhanden und aufgerufen.
Auslösendes Ereignis	Entscheidung des Prozessmanagers
Ergebnis	Prozessmodell ist eingefärbt.
Nachbedingung	Färbung wird kontrolliert.

In Abbildung 13.3 ist der Ablauf des Anwendungsfalls abgebildet. Der Prozessmanager beginnt den Prozess mit dem Klick auf eine Schaltfläche mit der Beschriftung *Prozess einfärben*. Daraufhin wird die XML-Repräsentation des aktuell geladenen BPMN-Modells an die entsprechende Systemkomponente gesendet. Hier werden nun die relevanten Bestandteile des Modells herausgefiltert. Dies sind im Wesentlichen die Repräsentationen der Datenobjekte (und Datenspeicher) und der Aktivitäten. Viele weitere enthaltene Informationen, wie z.B. Gateways oder Kontrollflüsse sind an dieser Stelle nicht relevant. Die gefilterten Elemente werden dann jeweils in eine Kategorie eingeteilt, die einer Farbe entspricht. Die einzelnen Elemente werden dann jeweils um die entsprechende Kategorie-Information erweitert. Anschließend werden alle Informationen wieder zu einer kompletten .bpmn-Datei zusammengefasst und letztlich wieder an das Frontend des Systems gesendet, wo das eingefärbte Modell angezeigt werden kann.

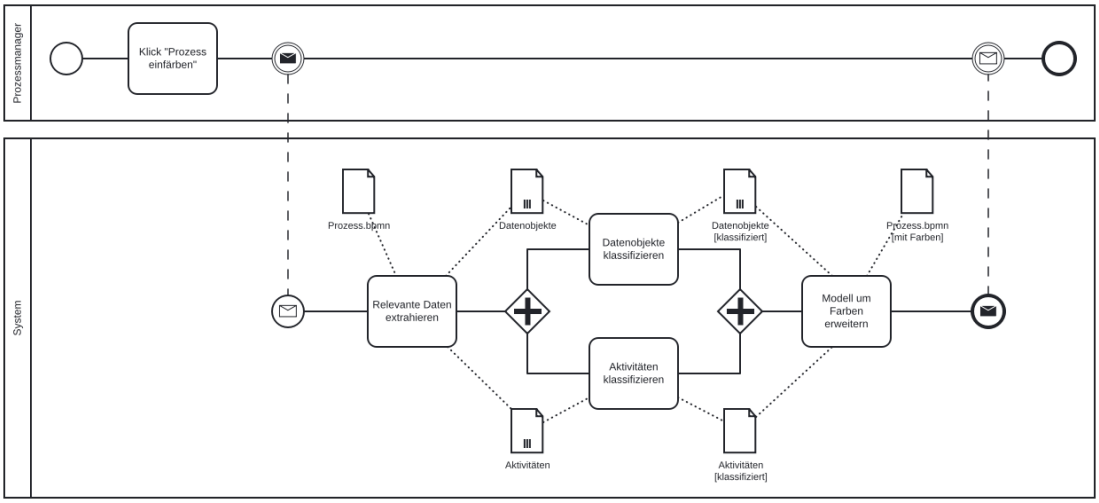


Abbildung 13.3.: BPMN Diagramm zu Use Case 2: Einfärben

Neben einer automatisierten Färbung muss auch noch eine manuelle Färbung möglich sein, für Fälle, die beispielsweise uneindeutig sind und nur von Menschen entschieden werden können. Dieses entspricht dann dem folgenden Use Case 3.

Use Case 3: Färbung korrigieren

Tabelle 13.7.: Use Case 3

Beschreibung	Färbung korrigieren
Priorität	hoch
Hauptakteur	Prozessmanager
Weitere Akteure	Datenschutzbeauftragter
Vorbedingungen	Eingefärbtes Prozessmodell in BPMN ist vorhanden.
Auslösendes Ereignis	Abgeschlossene Färbung.
Ergebnis	Prozessmodell ist korrekt eingefärbt.
Nachbedingung	–

Der dritte Anwendungsfall *Färbung korrigieren* ergänzt Anwendungsfall 2 um die zwingend notwendige Funktionalität, die automatisierte Färbung manuell zu korrigieren, da von der automatischen Färbung keine absolute Korrektheit erwartet werden kann. Auch die komplett manuelle Färbung wird hiermit abgedeckt, da dies letztlich auch nur eine Korrektur von der Nichtfärbung in weiß zu der jeweils korrekten Farbe darstellt.

In Tabelle 13.7 sind die wichtigsten Grundlagen des Anwendungsfalls aufgeführt. Wie oben bereits erwähnt, ist der Anwendungsfall extrem wichtig und erhält deshalb eine hohe Priorität. Auslösender Akteur ist auch hier wieder der Prozessmanager. Allerdings erhält dieser Unterstützung des Datenschutzbeauftragten (siehe Abschnitt 13.2.4). Voraussetzung für die Korrektur ist offensichtlich die bereits erfolgte Einfärbung eines Prozessmodells.

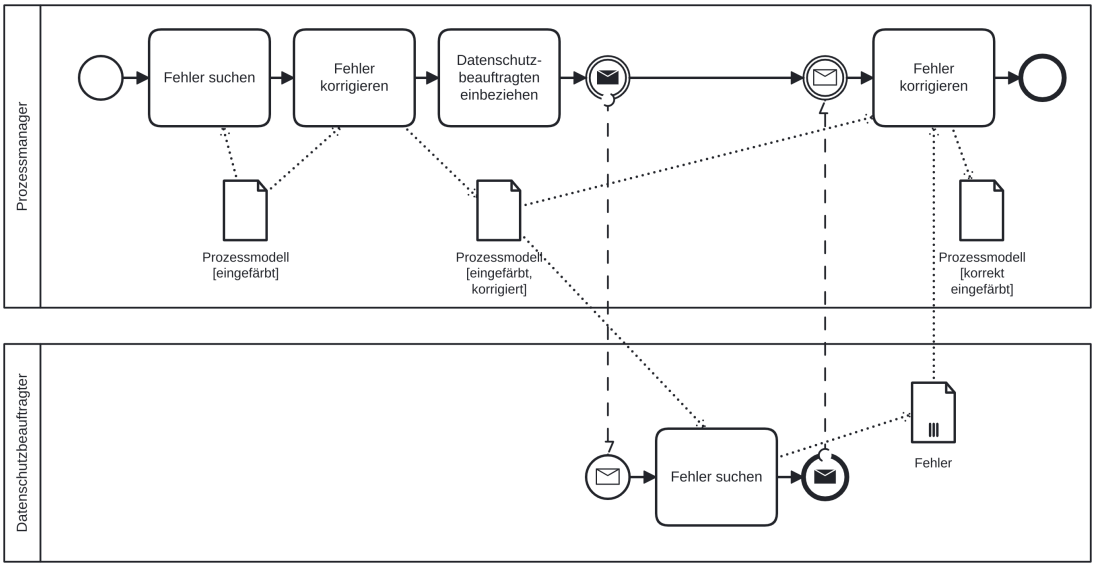


Abbildung 13.4.: BPMN Diagramm zu Use Case 3: Färbung korrigieren

Abbildung 13.4 zeigt den Ablauf des Anwendungsfalls. Zunächst sucht hier der Prozessmanager nach offensichtlichen Fehlern im eingefärbten Prozessmodell und korrigiert diese anschließend. Daraufhin leitet er das Prozessmodell an den Datenschutzbeauftragten weiter. Dieser überprüft das Modell ebenfalls mit seiner Datenschutzexpertise. Sollten Fehler gefunden werden, so werden diese vom Datenschutzbeauftragten notiert und an den Prozessmanager weitergegeben, der die Fehler dann wiederum korrigiert. Generell ist es auch immer möglich, die Farbe eines Objekts komplett zu entfernen, es also in den Ursprungszustand zu versetzen. Das kann z.B. bei Unklarheiten, die weiterer Diskussion bedürfen, sinnvoll sein.

Use Case 4: Prozess betrachten/manuell analysieren

Nach der Korrektur der Färbung sollte der Prozess von allen Beteiligten analysiert werden. Mit Hilfe der Färbung können Prozessteile mit einer hohen Datenschutzrelevanz schnell identifiziert werden. Anschließend kann diskutiert werden, ob an einzelnen Stellen datenschutzfreundliche Optimierungen denkbar sind oder sogar ein Prozess komplett neu gestaltet werden sollte.

Tabelle 13.8.: Use Case 4

Beschreibung	Prozess betrachten/analysieren
Priorität	hoch
Hauptakteur	Prozessmanager
Vorbedingungen	Korrekt eingefärbtes Prozessmodell in BPMN ist vorhanden.
Weitere Akteure	Datenschutzbeauftragter, Unternehmensleitung, Prozessbeteiligte
Auslösendes Ereignis	Färbung wurde besprochen und bei Bedarf korrigiert.
Ergebnis	Datenschutzprobleme wurden identifiziert. Mögliche Optimierungen wurden definiert.
Nachbedingung	Gefundene Optimierungen werden umgesetzt.

Für diesen Anwendungsfall wird kein Prozessmodell angegeben, da der Ablauf hier einerseits sehr variabel sein kann und andererseits aus diesem Anwendungsfall auch eher keine zusätzlichen Anforderungen an die Software resultieren. Softwareseitig muss hier nur eine Möglichkeit bestehen, dass verschiedene Personengruppen den Prozess betrachten können, was aber beispielsweise auch schon aus Use Case 3 hervorgeht. Die manuelle Analyse kann auch noch durch eine Erweiterung der automatischen Analyse ergänzt werden, die Use Case 2 zugrunde liegt. Es könnten etwa Optimierungsmöglichkeiten vorgeschlagen werden. Das Thema wird in Kapitel 17 aufgegriffen.

Use Case 5: Prozess bearbeiten

Tabelle 13.9.: Use Case 5

Beschreibung	Prozess bearbeiten
Priorität	mittel
Hauptakteur	Prozessmanager
Weitere Akteure	Prozessbeteiligte
Vorbedingungen	BPMN-Prozessmodell liegt vor.
Auslösendes Ereignis	Notwendigkeit zur Überarbeitung erkannt.
Ergebnis	Prozessmodell ist überarbeitet und persistent gespeichert.
Nachbedingung	Beteiligte bekommen Information über Fertigstellung.

Tabelle 13.9 zeit die wichtigsten Eckpunkte des Anwendungsfalls *Prozess bearbeiten*.

Im Wesentlichen entspricht dieser Anwendungsfall dem Use Case 1 „Modellieren“. Hier besteht nur die Voraussetzung, dass schon ein Prozessmodell vorliegt. Ansonsten gelten die Ausführungen aus Abschnitt 13.3.2 entsprechend.

13.4. Nichtfunktionale Anforderungen

Neben den funktionalen Anforderungen, sollten an eine Software auch nichtfunktionale Anforderungen definiert werden, welche die erwartete Qualität des Systems beschreiben. Einen Standard hierfür bietet die ISO/IEC 25010:2011 Norm. Die folgenden Ausführungen basieren auf den Ausführungen zur Anwendung dieses Standards in [EG18] und [Tr22]. Die nichtfunktionalen Anforderungen werden an dieser Stelle nicht formal mit Hilfe einer Satzschablone angegeben, da unter Anderem noch gar nicht zu allen Bereichen klare Anforderungen formuliert werden können. Außerdem muss an einigen Stellen zwischen dem fertigen Produkt und dem zunächst zu entwickelnden ersten Prototyp unterschieden werden.

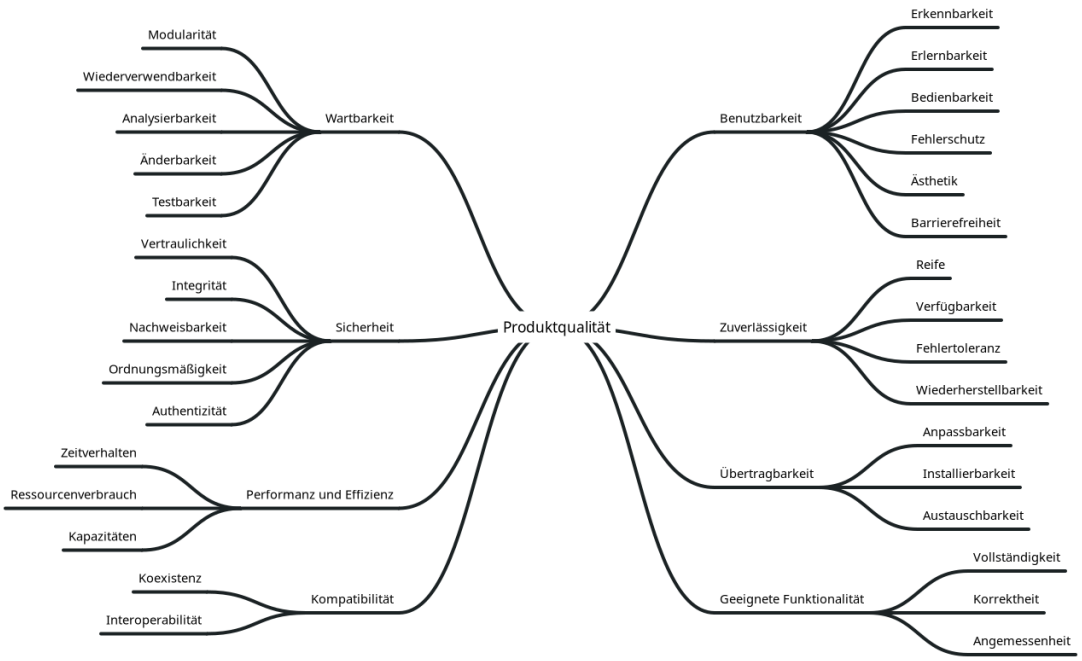


Abbildung 13.5.: Qualitätsmerkmale nach ISO/IEC 25010:2011 (in Anlehnung an [Tr22, S. 106])

Abbildung 13.5 stellt die in der Norm definierten Qualitätsmerkmale dar, auf die im Folgenden näher eingegangen wird.

13.4.1. Wartbarkeit

Die Wartbarkeit eines Systems ist in erster Linie für den Entwickler relevant, da dieser hiervon unmittelbare Kostenvorteile z.B. bei der Weiterverwendung zu erwarten hat. Sie betrifft aber natürlich auch den Kunden, da dieser bei der Behebung eventuell vorhandener Fehler und auch bei gewünschten Weiterentwicklungen tendenziell schneller ein Ergebnis geliefert bekommt.

Modularität beschreibt die Aufteilung eines Systems auf mehrere einzelne Komponenten gemäß des Prinzips *Seperation of Concerns*. Im vorliegenden Fall ist definitiv eine Trennung der reinen Modellierung und der automatischen Färbung sinnvoll. Einerseits kann die Modellierung auch ohne die Färbung sinnvoll verwendet werden. Andererseits steht die Konzeption der Färbung noch relativ am Anfang, sodass Änderungen wahrscheinlich sind. Hier ist ein Austausch einer einzelnen abgetrennten Komponente in der Regel einfacher, als wenn an verschiedenen Stellen im Gesamtsystem Änderungen vorgenommen werden müssen.

Wiederverwendbarkeit betrifft die Weiterverwendung einzelner Komponenten in zukünftigen Systemen. Das kann gerade im Prototyping viel Arbeit ersparen, wenn ein gewisser Teil eines Systems nicht den Anforderungen entspricht, einzelne Komponenten aber trotzdem noch genutzt werden können.

Analysierbarkeit resultiert aus einer guten Softwarearchitektur und Dokumentation. Auch das ist im Prototyping sehr wünschenswert, da beispielsweise höchstwahrscheinlich Probleme auftreten, deren Ursache und Behebungsmöglichkeiten analysiert werden müssen.

Änderbarkeit spielt offensichtlich gerade im Prototyping eine enorm wichtige Rolle, da mit Änderungen in jedem Fall zu rechnen ist. Aber auch bei einem fertigen Produkt können Änderungen immer wieder wünschenswert oder gar notwendig sein. Eine gute Modularität und Analysierbarkeit helfen hier.

Testbarkeit setzt insbesondere eine gute Testinfrastruktur voraus, die einfache Tests nach Änderungen am System ermöglichen. Auch das ist im Prototyping definitiv sinnvoll.

13.4.2. Sicherheit

Vertraulichkeit sichert vor unberechtigtem Zugriff auf Daten. Hier muss unterschieden werden zwischen personenbezogenen Daten und Firmendaten bzw. Geschäftsgeheimnissen. Die Verarbeitung Personenbezogener Daten mit der Anwendung

ist generell unwahrscheinlich. Ein unberechtigter Zugriff wäre hier daher maximal denkbar, wenn durch eine Sicherheitslücke in der Anwendung Zugriff auf externe Daten des Unternehmens ermöglicht werden würde. Das ist natürlich zu verhindern. Ein Zugriff auf Geschäftsgeheimnisse ist wahrscheinlicher. Dieser könnte z.B. möglich sein, wenn die modellierten Prozesse an den Entwickler bzw. Betreiber der Software weitergeleitet werden. Dies könnte beispielsweise durchaus sinnvoll sein, wenn die Färbung auf externen Servern berechnet werden sollte, wofür eine Übermittlung des Prozesses nötig wäre. Daher ist entweder eine komplett lokale Lösung, oder sind entsprechende Maßnahmen wie etwa Verschlüsselung anzustreben.

Integrität stellt den Schutz der Daten vor Veränderungen dar. Hierfür kann z.B. mit einem Rechtemanagement gearbeitet werden. Die Integrität hat eine eher niedrige Relevanz für das System, da unzulässige Änderungen recht schnell auffallen und eher geringe Auswirkungen auf das Unternehmen haben.

Nachweisbarkeit wird auch als Nichtabstreitbarkeit bezeichnet und bedeutet, dass überprüft werden kann, ob ein Ereignis wie z.B. eine Dateiänderung stattgefunden hat. Hierfür sind insbesondere Protokolldateien sinnvoll. Die Relevanz im betrachteten Fall ist aber als eher gering anzusehen, da kaum Fälle existieren, in welchen eine derartige Prüfung wirklich wichtig wäre.

Ordnungsmäßigkeit bezieht sich auf die Möglichkeit, eine Änderung auf eine bestimmte Person zurückzuführen. Auch hierfür ist ein Rechtemanagement in Verbindung mit Protokolldateien sinnvoll. Die Relevanz ist aber auch hier als eher gering anzusehen.

Authentizität bezeichnet die Sicherstellung der Identität eines Nutzers oder eines verwendeten Systems. Eine Rechteverwaltung legt auch hier den Grundstein, wobei die Relevanz des Merkmals wieder als eher gering angesehen werden kann.

13.4.3. Performanz und Effizienz

Zeitverhalten spielt eine wichtige Rolle für die sinnvolle Benutzbarkeit eines Systems. Hier müssen unterschiedliche Funktionen betrachtet werden: Die Prozessmodellierung und -modifikation sollte in „Echtzeit“, also ohne erkennbare Verzögerung möglich sein, wie es in anderen Modellierungswerkzeugen üblich ist. Auch das Speichern und Laden eines durchschnittlichen Prozesses sollte maximal im niedrigen einstelligen Sekundenbereich liegen, da diese Operationen sehr häufig durchgeführt werden und lange Wartezeiten hier zu deutlichen Einschränkungen führen. Etwas anders ist die automatische Färbung zu betrachten. Diese soll

nur vergleichsweise selten ausgeführt werden und darf daher zeitintensiver sein. Generell ist eine kurze Berechnungszeit zwar auch hier erstrebenswert, hat aber niedrigere Priorität. Wichtig wäre aber bei einer längeren Ladezeit eine entsprechende Information des Nutzers, z.B. durch Anzeige der erwarteten Rechenzeit oder einen Fortschrittsbalken.

Ressourcenverbrauch bezeichnet den Bedarf an Rechenleistung und Speicherplatz. Hier ist zwar immer ein möglichst geringer Wert wünschenswert, hat aber in diesem Fall auch nicht die oberste Priorität. Die Anwendung sollte aber auf einem durchschnittlichen Bürorechner problemlos ausführbar sein.

Kapazitäten bezieht sich auf die maximal möglichen Nutzer, erwartbaren nebenläufigen Prozessen usw. Die Anforderungen sind hier relativ gering, da auch in einem großen Unternehmen nur eine recht geringe Anzahl an Mitarbeitern einen Zugriff auf das System braucht. Trotzdem muss aber zumindest deren parallele Nutzung getestet und problemlos machbar sein. Bei einer Lösung mit einem zentralen Server für mehrere Unternehmen muss mit entsprechend höheren Nutzerzahlen gerechnet werden und die Lösung sollte daher skalierbar sein.

13.4.4. Kompatibilität

Koexistenz betrifft die parallele Ausführung des Produkts und anderer Software auf dem gleichen Rechner. Hier sollten möglichst keine Abhängigkeiten bestehen. Relevant ist der Punkt insbesondere bei bereits vorhandenen Workflowmanagementsystemen. Hier darf die Nutzung des zu entwickelnden Systems die bisherigen Abläufe nicht (negativ) beeinflussen.

Interoperabilität bezieht sich auf die notwendige Zusammenarbeit mit anderen Systemen. Neben den Standardanforderungen, wie beispielsweise Betriebssystemkompatibilität, hängen die meisten Anforderungen hier von der konkreten Umsetzung ab. So sollte bei einer browserbasierten Lösung z.B. selbstverständlich die Kompatibilität mit den verbreitetsten Browsern gegeben sein. Schnittstellen zu anderer Software ist im Wesentlichen nicht notwendig. Unter Umständen kann aber eine Schnittstelle zu dem oben schon erwähnten Workflowmanagementsystem sinnvoll sein. Vorteilhaft wäre es hier, die geforderten Funktionen in dieses System zu integrieren. Als Mindestanforderung sollte ein gemeinsames Datenaustauschformat gegeben sein. Hier bietet sich das verbreitete BPMN-XML-Format an.

13.4.5. Benutzbarkeit

Erkennbarkeit hängt eng mit dem Wiedererkennungswert einzelner Softwarekomponenten zusammen. Im vorliegenden Fall ist wohl insbesondere die Modellierung zu betrachten. Einerseits sollte die Modellierungsnotation klar dem Standard entsprechen und andererseits sollte sich der Aufbau der Benutzeroberfläche an der vergleichbarer Werkzeuge orientieren.

Erlernbarkeit ist ein wichtiger Faktor für eine möglichst geringe Einarbeitungszeit. Für die Modellierung sollte die Erlernbarkeit durch die oben beschriebene Erkennbarkeit für entsprechend erfahrende Benutzer ohnehin gut sein. Für Anfänger sollte eine gute Dokumentation vorliegen, die aber nicht im Umfang des ersten Prototyps vorhanden sein muss. Die automatische Färbung sollte recht einfach und intuitiv sein, da hier nur ein Klick auf eine Schaltfläche nötig sein soll. Etwas schwieriger verhält es sich mit dem generellen Konzept der Einfärbung. Zwar ist hier grundsätzlich auch eine hohe Erkennbarkeit gegeben (siehe Abschnitt 4.1), die den Einarbeitungsaufwand verringert, allerdings ist zumindest für die manuelle Korrektur der Färbung doch ein sehr klares Verständnis der Semantik der einzelnen Farbkategorien notwendig. Diese ist insbesondere bei eher geringem Datenschutzwissen nicht ganz offensichtlich. Daher sollte die Erlernbarkeit hier unterstützt werden. Dies könnte durch eine gute Dokumentation, aber auch durch Tooltips oder ähnliches erreicht werden. Für den Prototyp hat aber auch das nicht die oberste Priorität.

Bedienbarkeit kann unter Anderem durch „ein konsistentes und durchgängig einheitliches Visual- und Interaktionsdesign“ [Tr22, S. 109] verbessert werden. Das sollte frühzeitig in der Entwicklung betrachtet werden. Außerdem sollte ein einfacher und direkter Zugriff auf die wichtigsten Funktionen gegeben sein.

Fehlerschutz hilft dem Benutzer, möglichst wenig Fehler in seine Arbeitsergebnisse einzubauen. Für die Modellierungskomponente sollte daher beispielsweise eine Syntaxüberprüfung vorhanden sein. Bei der Färbungskomponente wäre beispielsweise eine Plausibilitätsprüfung denkbar, die warnt, wenn in einem Prozess zwei Elemente mit der gleichen Bezeichnung unterschiedliche Farben haben. Bei der Korrektur der Farben könnte gewarnt werden, wenn eine deutliche Verbesserung der Kategorie, also von rot auf grün, vorgenommen werden soll. Dies könnte den Nutzer bei zu häufigen Warnungen aber auch stören und sollte daher zumindest deaktivierbar sein. Generell sollten für derartige Funktionen Anwendertests durchgeführt werden, die aufzeigen, welche Funktionalitäten nützlich und welche eher störend wirken.

Ästhetik hilft, dem Anwender ein angenehmes Nutzungserlebnis zu bieten. Sie ist recht subjektiv und daher schwierig zu definieren. Für den Prototyp ist dieses Thema eher weniger relevant.

Barrierefreiheit hilft Menschen mit Behinderungen oder sonstigen Einschränkungen bei der Benutzung des Systems. Dieses Thema ist definitiv sehr wichtig, kann für den Prototyp aber zunächst weitestgehend vernachlässigt werden, da hier zunächst ohnehin nur eine sehr überschaubare Menge von Testern vorgesehen ist. Trotzdem sollte das Thema bei grundlegenden Designentscheidungen zumindest mit bedacht werden.

13.4.6. Zuverlässigkeit

Reife gibt an, wie viele Probleme beim Betrieb einer Software auftreten. Eine hohe Reife steht für wenig auftretende Probleme und wird in der Regel erst nach einer gewissen Betriebszeit erreicht. Hier kann für den Prototyp dementsprechend kein besonders guter Wert erwartet werden.

Verfügbarkeit gibt die Zeit an, die eine Software in einer bestimmten Zeitspanne verwendet werden konnte, in der es also nicht zu Ausfällen kam. Auch dieser Wert ist für den Prototyp unerheblich. Auch bei einem fertigen Produkt sind die Anforderungen hier nicht allzu hoch, da mit der Software keine zeitkritischen Prozesse bearbeitet werden. Nichtsdestotrotz sollte eine gewisse Verfügbarkeit sichergestellt werden, um Kunden zufriedenzustellen.

Fehlertoleranz bezieht sich in diesem Kontext darauf, dass es bei Fehleingaben sowie Fehlfunktionen des Betriebssystems oder anderer Systeme nicht zu Datenverlusten kommt. Das kann recht einfach durch ein regelmäßiges Autosave erreicht werden. Noch besser wäre eine automatische Versionierung der erstellten Prozesse.

Wiederherstellbarkeit betrachtet den Zustand nach einem Systemausfall bzw. den Aufwand, der zur Wiedererlangung der Arbeitsfähigkeit notwendig ist. Auch dieser Aspekt ist als eher weniger wichtig zu betrachten, da die Nutzung der Software nicht geschäftskritisch ist. Trotzdem ist natürlich auch hier die Kundenzufriedenheit ein Thema.

13.4.7. Übertragbarkeit

Anpassbarkeit bezieht sich auf Änderungen im Umfeld der Software, beispielsweise des Betriebssystems. Hier spielen einerseits neue Versionen, andererseits aber

auch ein kompletter Wechsel des Betriebssystems oder z.B. einzelner Hardwarekomponenten eine Rolle. Für den Prototyp ist diese Anforderung zunächst eher nebensächlich. Bei einem späteren Produkt sollte aber zumindest ein gewisser Grad an Anpassbarkeit gegeben sein.

Installierbarkeit meint einerseits die Einfachheit und Schnelligkeit der Installation, andererseits aber auch die (im besten Fall) Abwesenheit von Seiteneffekten. Für den Prototyp ist zumindest der Installationsvorgang recht unerheblich, da die Installation auf den Systemen der Tester hier vom Entwickler unterstützt werden kann. Für das fertige Produkt sollte aber definitiv eine einfache und schnelle Installation angestrebt werden. Seiteneffekte sollten schon im Prototyps-Stadium möglichst ausgeschlossen werden.

Austauschbarkeit betrifft die Frage, wie einfach das System bei Bedarf ausgetauscht werden kann. Da im Wesentlichen keine Abhängigkeiten zu anderen Systemen zu erwarten sind, sollte dies kein Problem darstellen – abgesehen davon, dass zumindest aktuell kein entsprechendes Alternativsystem zur Verfügung steht.

13.4.8. Geeignete Funktionalität

Geeignete Funktionalität beschreibt im Wesentlichen die Erfüllung der funktionalen Anforderungen, wie sie in Abschnitt 13.3 beschrieben werden. Generell ist die geeignete Funktionalität in der Regel als das wichtigste Qualitätsmerkmal anzusehen. Für den zu entwickelnden Prototyp kann das allerdings mit gewissen Einschränkungen betrachtet werden.

Vollständigkeit beschreibt den Anteil der umgesetzten spezifizierten Anwendungsfälle. Hier sollten zumindest die Anwendungsfälle mit der Priorität *hoch* umgesetzt werden. Das ergäbe hier eine Minimalanforderung von 75%.

Korrektheit bezeichnet die Fehlerfreiheit der implementierten Funktionalitäten. Diese ist für einen ersten Prototyp generell nicht allzu wichtig, da gefundene Fehler problemlos in einem späteren Schritt behoben werden können. Im fertigen Produkt hat diese aber natürlich höchste Relevanz. Interessant ist hier insbesondere die Korrektheit im Use Case 2: Einfärben (siehe Abschnitt 13.3.2). Einerseits kann hier keine hundertprozentige Korrektheit erwartet werden. Andererseits steigt der Nutzen der Software aber deutlich mit der Korrektheit dieses Anwendungsfalls, weshalb ein möglichst hoher Wert angestrebt werden sollte. Für konkrete Werte sollten mögliche Anwender befragt werden, ab welcher zu erwartenden Wahrscheinlichkeit der korrekten Einfärbung sie das Werkzeug nutzen würden. Für den

Prototyp ist ein sinnvoller erster Schritt, den Grad der Korrektheit der Einfärbung zu messen.

Angemessenheit bedeutet, dass jeder Nutzer genau die Funktionalitäten angeboten bekommt, die er auch braucht. Im konkreten Fall könnte das z.B. bedeuten, dass nur der Prozessmanager die Möglichkeit erhält, Prozesse zu erstellen, zu bearbeiten und einzufärben. Alle anderen Nutzergruppen hingegen könnten Prozesse nur betrachten. Für den ersten Prototyp kann das allerdings vernachlässigt werden, da hiermit zunächst nur die grundsätzliche Funktionalität gezeigt werden soll. Außerdem ist es fraglich, ob die Anwendungsfälle wirklich in jeder Organisation so umgesetzt werden, wie sie oben dargestellt werden. Möglicherweise sollen auch andere Anwender die Option haben, Prozesse zu bearbeiten, falls diese die nötige Kompetenz besitzen. Trotzdem muss die Möglichkeit eines entsprechenden Rechtesystems in spätere Designentscheidungen miteinbezogen werden.

13.5. Zusammenfassung und Priorisierung

Im Folgenden sollen die Ergebnisse der Anforderungsanalyse kurz zusammengefasst und die einzelnen Anforderungen priorisiert werden.

13.5.1. Funktionale Anforderungen

Die Analyse der funktionalen Anforderungen hat zwei Kernfunktionalitäten ergeben: Die Modellierung von Geschäftsprozessen und deren automatische Einfärbung. Die übrigen beiden Anwendungsfälle *Färbung korrigieren* und *Prozess betrachten/analysieren* können auf technischer Ebene im Grunde als Teilaspekte der Modellierung aufgefasst werden.

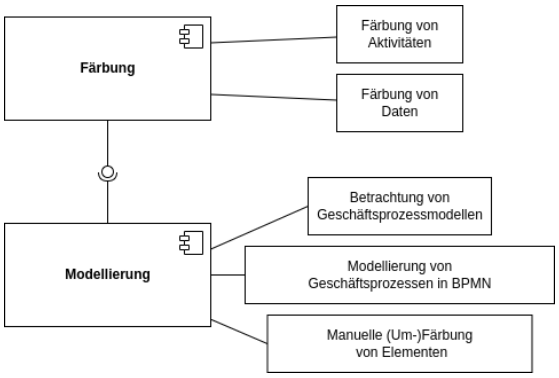


Abbildung 13.6.: Zu entwickelnde Komponenten mit Teilanforderungen

Abbildung 13.6 stellt die Anforderungen an die beiden Komponenten grafisch dar. Die Priorität der Färbungskomponente wurde im **Use Case 2** als *hoch* festgelegt. Die Priorität der Modellierungskomponente geht aus den Anwendungsfällen nicht klar hervor. Für die Modellierung an sich wurde in **Use Case 1** die Priorität *mittel* festgelegt, da es ausreichend viele Werkzeuge für diesen Zweck gibt. Allerdings haben **Use Case 3** und **Use Case 4** die Priorität *hoch* und es ist nicht klar, ob ein externes Modellierungswerkzeug die von der Färbungskomponente vergebenen Farben verarbeiten kann. Daher muss hier zunächst überprüft werden, ob alle Anwendungsfälle von einem bestehenden Werkzeug abgedeckt werden können.

13.5.2. Nichtfunktionale Anforderungen

Abbildung 13.7.: Priorisierung der Qualitätsmerkmale nach ISO/IEC 25010:2011

Die verschiedenen Kategorien nichtfunktionaler Anforderungen in Abbildung 13.7 sind im Gegensatz zu Abbildung 13.5 nach ihrer Priorität für den Prototyp sortiert. Die Prioritäten sind darin auch farblich markiert:

- Rot steht für eine hohe Priorität und somit für Anforderungen, die unbedingt umgesetzt werden müssen;
- Gelb steht für Anforderungen mit einer mittleren Priorität, die nach Möglichkeit umgesetzt werden sollten;
- Und Grün steht für Anforderungen mit niedriger Priorität, die für den Prototyp zunächst vernachlässigt werden können.

Die *geeignete Funktionalität* hat hierbei als einzige Kategorie eine hohe Priorität, weil der Prototyp gerade die Funktionalitäten umsetzen und demonstrieren soll.

Allerdings wird den einzelnen Merkmalen jeweils nur eine mittlere Priorität zugeordnet (siehe Abbildung 13.8), da bei einem Prototyp weder *Vollständigkeit* noch *Korrektheit* erwartet werden kann. Und auch die *Angemessenheit* hat nur eine mittlere

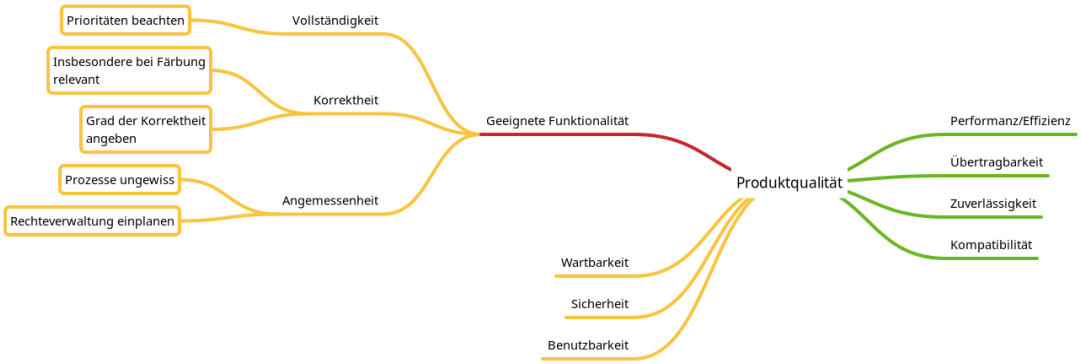


Abbildung 13.8.: Priorisierung der Qualitätsmerkmale im Bereich *Geeignete Funktionalität*

Priorität, weil hierfür Informationen fehlen, die erst im Rahmen der Evaluation des Prototyps erlangt werden können.

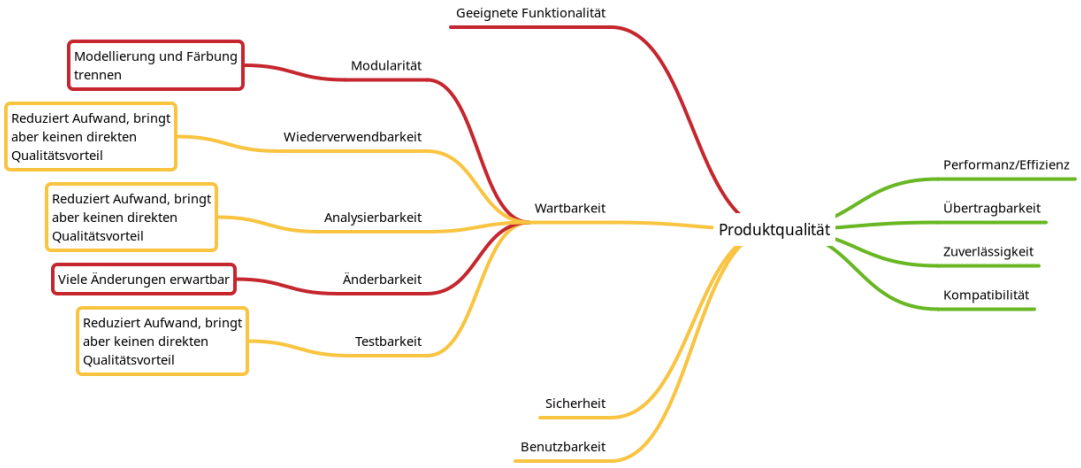


Abbildung 13.9.: Priorisierung der Qualitätsmerkmale im Bereich *Wartbarkeit*

Abbildung 13.9 zeigt die Prioritäten und zu betrachtenden Aspekte der einzelnen Merkmale der Kategorie *Wartbarkeit*. Generell haben hier alle Merkmale eine mindestens mittlere Priorität. Dies lässt sich damit begründen, dass der Prototyp weiterentwickelt werden soll und hierfür alle Merkmale recht wichtig sind. Eine besonders hohe Priorität hat die *Modularität*, da an mehreren Stellen klar geworden ist, dass die Software in zwei zentrale Komponenten aufgeteilt werden sollte. Ebenfalls eine hohe Priorität hat die *Änderbarkeit*, da diese für die Weiterentwicklungen besonders zentral ist.

Die Kategorie *Sicherheit* (siehe Abbildung 13.10) hat ebenfalls eine mittlere Priorität. Hier hat allerdings die Mehrheit der Merkmale eine eher geringe Priorität. Demgegen-

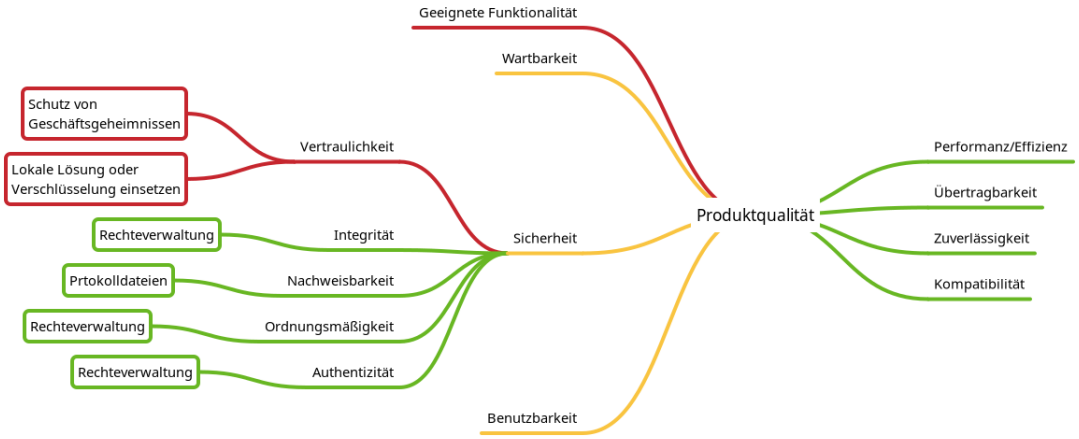


Abbildung 13.10.: Priorisierung der Qualitätsmerkmale im Bereich *Sicherheit*

über wird der *Vertraulichkeit* eine hohe Priorität zugeschrieben, da einerseits der Schutz der Geschäftsgeheimnisse potentieller Kunden enorm wichtig ist und andererseits dieser Aspekt schon früh in die Entwicklung miteinbezogen werden muss, da er unter Umständen Auswirkungen auf wichtige Designentscheidungen hat.

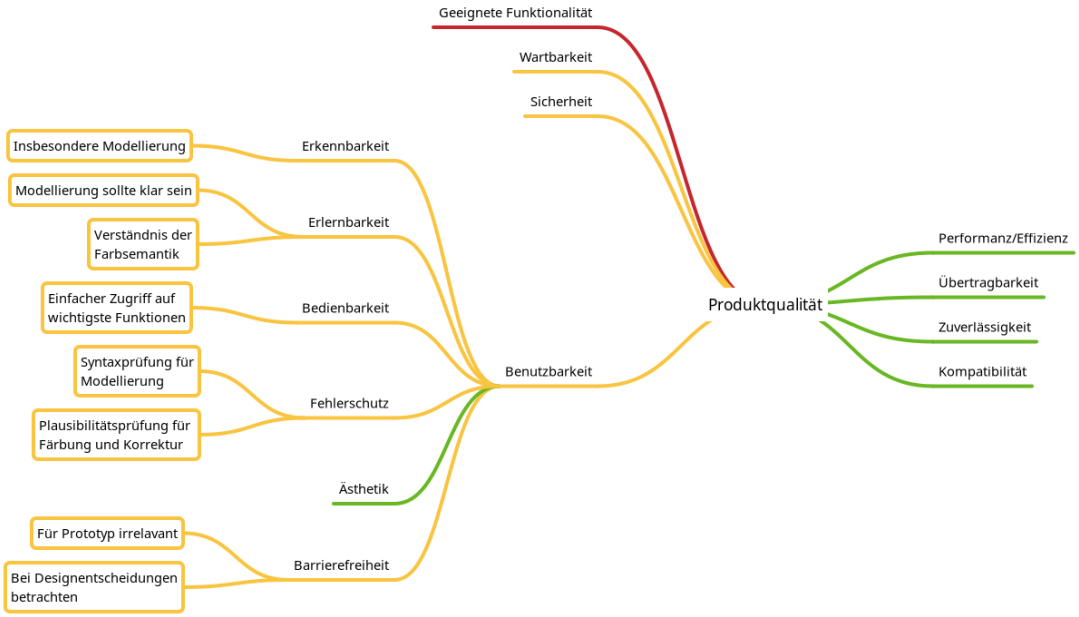


Abbildung 13.11.: Priorisierung der Qualitätsmerkmale im Bereich *Benutzbarkeit*

Die *Benutzbarkeit* hat ebenfalls eine mittlere Priorität, wie in Abbildung 13.11 zu sehen. Für die meisten Merkmale gilt hier, dass sie frühzeitig bedacht werden müssen, da eine späte Betrachtung zu großem Änderungsaufwand führen kann. Lediglich die

Ästhetik hat eine geringe Priorität, da das Design einerseits recht leicht auch später noch angepasst werden kann und Ästhetik andererseits abhängig vom Nutzer ist und das Design daher unter Umständen auch individuell angepasst werden sollte.

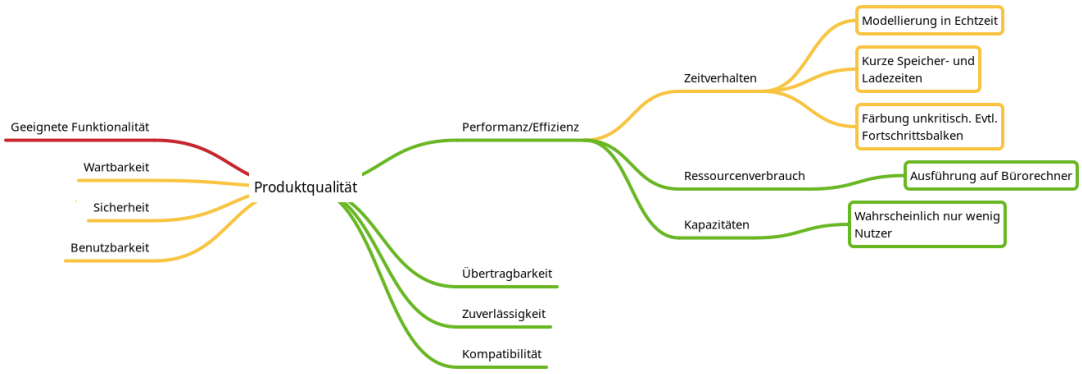


Abbildung 13.12.: Priorisierung der Qualitätsmerkmale im Bereich *Performanz Effizienz*

Der *Performanz und Effizienz* wird insgesamt nur eine geringe Priorität zugeordnet. Die einzelnen Merkmale lassen sich Abbildung 13.12 entnehmen. Einzig das *Zeitverhalten* hat hier eine mittlere Priorität, wobei diese sich eher auf das fertige Produkt als auf den Prototyp bezieht. Aber auch für die erste Evaluation ist ein angemessenes Zeitverhalten förderlich. Außerdem sollte bei deutlich zu schlechtem Zeitverhalten frühzeitig nach Gründen gesucht werden, da eventuell umfangreiche Änderungen nötig sind.

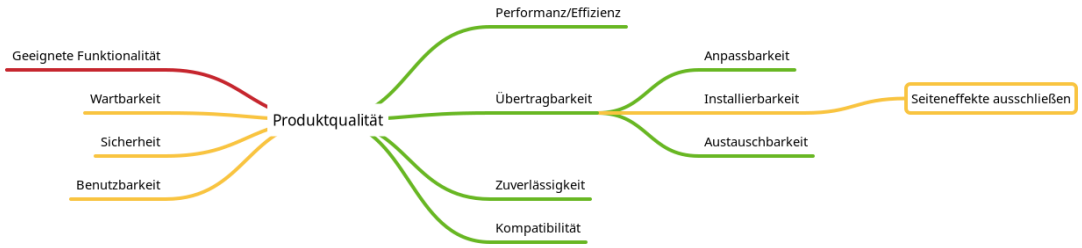


Abbildung 13.13.: Priorisierung der Qualitätsmerkmale im Bereich *Übertragbarkeit*

Auch die Kategorie *Übertragbarkeit*, dargestellt in Abbildung 13.13 hat insgesamt eine geringe Priorität. Allerdings hat die *Installierbarkeit* hier eine mittlere Priorität, da zumindest Seiteneffekte bei der Installation frühzeitig betrachtet werden sollten.

In der Kategorie *Zuverlässigkeit* haben alle Merkmale eine geringe Priorität (siehe Abbildung 13.14).

Auch die Kompatibilität (siehe Abbildung 13.15) hat eine geringe Priorität, was auch für beide Teilmerkmale gilt.

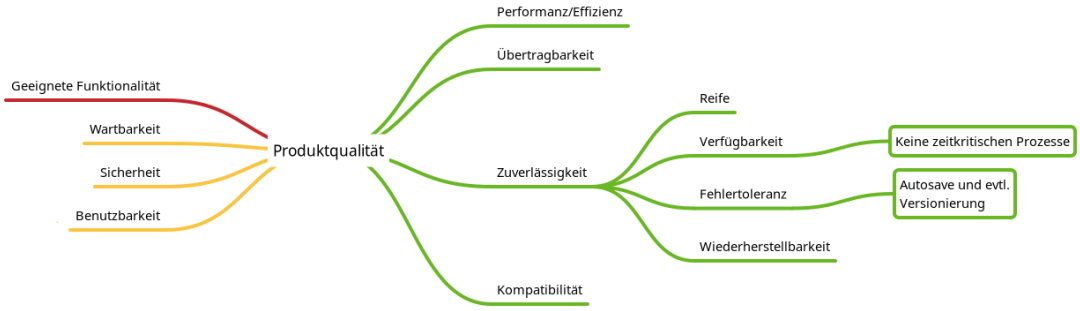


Abbildung 13.14.: Priorisierung der Qualitätsmerkmale im Bereich *Zuverlässigkeit*

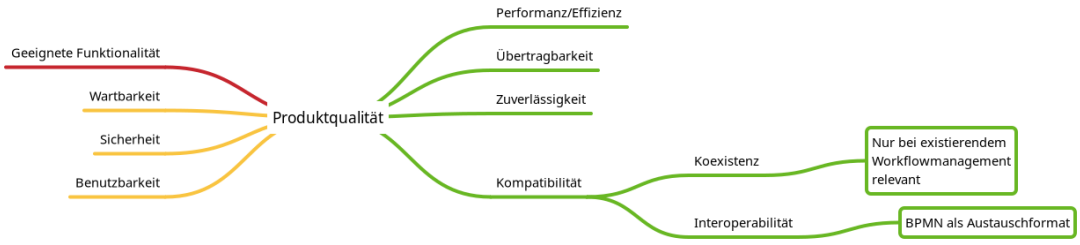


Abbildung 13.15.: Priorisierung der Qualitätsmerkmale im Bereich *Kompatibilität*

14. Realisierung

Dieses Kapitel zeigt die prototypische Umsetzung der in Kapitel 13 dargestellten Anforderungen in Form eines Plugins für ein bestehendes Modellierungswerkzeug. Hierfür wird zunächst das Basissystem beschrieben und dessen Auswahl begründet (siehe Abschnitt 14.1). Anschließend folgt der generelle Aufbau des Plugins in Abschnitt 14.2. Ein Kernelement der Realisierung ist die Erweiterung des BPMN-Standards um Attribute, welche die Datenschutz-Kategorien repräsentieren. Dies wird in Abschnitt 14.4 erläutert. Das andere Kernelement des Plugins stellt die eigentliche Kategorisierung dar. Hierfür werden in Abschnitt 14.3 drei verschiedene Ansätze vorgestellt, welche zu unterschiedlichen Graden fertiggestellt wurden. Abschnitt 14.5 fasst die Ergebnisse abschließend kurz zusammen, bevor diese im folgenden Kapitel 15 evaluiert werden.

14.1. Basissystem

Grundsätzlich gibt es bei der Realisierung einer Software immer die Möglichkeiten komplett neu anzufangen oder aber auf ein bestehendes System aufzusetzen und dieses zu erweitern.

Da für einen wesentlichen Teil des zu entwickelnden Systems, nämlich die Geschäftsprozessmodellierung mit BPMN, schon einige Systeme vorhanden sind, erscheint eine komplette Neuentwicklung nicht sinnvoll.

14.1.1. Bestehende Systeme

Bei der Betrachtung bestehender Systeme, die erweitert werden können, fällt der Blick zunächst auf den in [FWS11] veröffentlichten „Business Application Modeler (BAM)“. Dieses Modellierungswerkzeug verfolgt das Ziel, Compliance zu gewährleisten, indem Prozessmodelle algorithmisch mit Hilfe von Model Checking auf die Einhaltung temporaler Regeln geprüft werden[Ni15]. Die Regeln werden dabei grafisch modelliert. Das System wurde ursprünglich für (erweiterte) Ereignisgesteuerte Prozessketten (eEPK) entwickelt, aber auch für die Arbeit mit BPMN erweitert[St12].

Der oben beschriebene Forschungsprotoyp ist für seinen Zweck zwar durchaus nützlich, bringt aber Schwächen mit sich. Das Forschungssystem hat noch offene Punkte,

so dass nicht alle theoretisch möglichen Regeln durch das System verarbeitet werden können. Somit können nicht alle Bereiche des Datenschutzes überprüft werden. Dazu kommt der Umstand, dass BAM eine Eigenentwicklung auf Basis der integrierten Entwicklungsumgebung Eclipse[1] ist. Das ganze System ist daher stark auf die Anwendung durch Softwareentwickler ausgelegt. Diese sind aber in der Praxis nicht die bedeutendste Zielgruppe für das System und ebenfalls nicht für den hier beschriebenen Prototyp. Bei der hier relevanten Zielgruppe der Prozessmanager, Datenschutzbeauftragten und Unternehmensleitungen ist von einer geringen Akzeptanz eines solchen Systems auszugehen.

Daher werden im Folgenden einige auf dem Markt verfügbare Modellierungswerkzeuge betrachtet. Hierfür wurde ein Kriterienkatalog erstellt, um eine Auswahl zu treffen. Die Erfahrungen mit BAM gehen aber in diese Betrachtung und in die Entwicklung ein.

Als erstes Kriterium wurde natürlich geprüft, ob die Software überhaupt erweitert werden kann und darf. Der nächste betrachtete Aspekt ist auf Grund der geringen verfügbaren Mittel das Lizenzmodell. Die Software sollte nach Möglichkeit zumindest für Forschungszwecke kostenlos nutz- und erweiterbar sein. Außerdem sollte der BPMN-Standard möglichst umfassend und korrekt unterstützt werden.

Verfügbare Modellierungswerkzeuge wurden zunächst einer Liste[2] der *BPMN Model Interchange Working Group* (BPMN MIWG) der OMG entnommen. Dort werden alle Werkzeuge gelistet, die von der BPMN MIWG bzgl. ihrer Umsetzung des Standards getestet wurden. Einige gelistete Werkzeuge, die durch online Recherche nicht gefunden werden konnten (z.B. Camunda Eclipse Plugin, ModelFoundry, actiBPM), wurden nicht weiter betrachtet. Der Aspekt der BPMN-Standard-Konformität sollte bei allen gelisteten Werkzeugen ohnehin gegeben sein und wird daher nicht weiter überprüft.

Die Ergebnisse der Recherche bzgl. Erweiterbarkeit und Lizenzen sind Tabelle B.1 und Tabelle B.2 im Anhang zu entnehmen. Insgesamt sind sieben der geprüften Werkzeuge sicher erweiterbar. Bei drei weiteren Werkzeugen wurden Hinweise gefunden, die eine Erweiterbarkeit nahelegen, allerdings ohne, dass hierzu weitere Informationen angegeben sind. Diese Werkzeuge werden wegen des erhöhten Einarbeitungsaufwands nicht weiter betrachtet. Von den erweiterbaren Werkzeugen unterliegen die meisten einer Lizenz, welche für das Projekt in Frage kommt. Einzig der Case Agile Enterprise Explorer unterliegt keiner entsprechend freien Lizenz, weshalb dieser ebenfalls nicht weiter betrachtet wird. Für die weitere Auswahl kommen also noch sechs Werkzeuge in Frage:

- Camunda Modeler
- Case Agile BPMN View

[1] https://eclipseide.org/

[2] https://bpmn-miwg.github.io/bpmn-miwg-tools/

- Modelio
- Yaoqiang BPMN Editor
- Activity
- Bonita

Diese werden im Folgenden hinsichtlich ihrer Eignung untersucht. Hierfür werden die folgenden Kriterien betrachtet:

Dokumentation: Der Erweiterungsmechanismus sollte möglichst gut beschrieben sein, um einen einfachen und schnellen Einstieg zu ermöglichen.

Betriebssystem: Bevorzugt wird eine betriebssystemunabhängige Lösung, da dies auch eine Anforderung für das zu entwickelnde System ist (siehe 13.2.1).

Letzte Änderung: Außerdem sollte das Werkzeug möglichst aktuell und nicht veraltet sein, damit beispielsweise aktuelle Funktionalitäten der verwendeten Programmiersprachen verwendet werden können und ein geringeres Risiko für offene Schwachstellen besteht.

Tabelle 14.1.: Bewertung verfügbarer Modellierungswerkzeuge

Werkzeug	Dokumentation	OS	Letzte Änderung[3]
Camunda Modeler	+[4][5]	+	Feb 2023
Case Agile BPMN View	-[6]	-[7]	Sep 2019
Modelio	+[8]	+	Mär 2022
Yaoqiang BPMN Editor	-	+	Feb 2023
Activity	-	(+)[9]	Mär 2023
Bonita	+[10]	(+)[11]	Okt 2022

Die Bewertung der sechs in Frage kommenden Werkzeuge wird in Tabelle 14.1 zusammengefasst.

[3] Stand: 10.03.2023
[4] Dokumentation: https://docs.camunda.io/docs/components/modeler/desktop-modeler/plugins/
[5] Beispiele: https://github.com/camunda/camunda-modeler-plugins
[6] https://github.com/bzinchenko/bpmnview
[7] nur Windows
[8] https://github.com/ModelioOpenSource/Modelio/wiki/Module-Developer-Guide-Index
[9] Docker, AWS oder GKE
[10] https://documentation.bonitasoft.com/bonita/2022.2/software-extensibility/extensions-sdk
[11] Offiziell nur Microsoft Windows Server, Red Hat Enterprise, CentOS, Ubuntu

Eine ausreichend gute Dokumentation für den Erweiterungsmechanismus ist auf den Webseiten des Camunda Modelers, sowie von Modelio und Bonita auffindbar. Im Github-Repository von BPMN View sind einige Codebeispiele aufgeführt, die aber nicht wirklich als Dokumentation angesehen werden können. Zu Yaoqiang und Activity sind gar keine Informationen bzgl. Erweiterungen verfügbar.

Die Betriebssystemunabhängigkeit (Linux, macOS und Windows) ist beim Camunda Modeler, Modelio und Yaoqiang gegeben. BPMN View ist nur für Windows verfügbar. Activity soll eigentlich als Cloud-Dienst verwendet werden und ist für Amazon Web Services (AWS) und die Google Kubernetes Engine (GKE) verfügbar. Darüber hinaus kann aber auch Docker verwendet werden, was über diesen Umweg auch eine lokale betriebssystemunabhängige Installation möglich macht. Bonita unterstützt offiziell Microsoft Windows Server, Red Hat Enterprise, CentOS und Ubuntu.

Als sehr aktuell können der Camunda Modeler, Yaoqiang, und Activity angesehen werden. Hier wurden die letzten Änderungen jeweils im aktuellen Jahr 2023 vorgenommen. Auch Bonita mit Stand Oktober 2022 ist noch in Ordnung. Modelio wurde zuletzt im März 2022 überarbeitet. BPMN View kann mit einer letzten Aktualisierung im September 2019 definitiv als veraltet angesehen werden.

Aus den drei beschriebenen Kriterien ergibt sich also folgende Rangliste:

1. Camunda Modeler
2. Modelio
3. Bonita
4. Yaoqiang BPMN Editor
5. Activity
6. Case Agile BPMN View

Für diese Arbeit fällt die Wahl daher auf eine Erweiterung des Camunda Modelers, der im folgenden Abschnitt beschrieben wird.

14.1.2. Camunda Modeler

Der Camunda Modeler ist ein Produkt der Camunda Services GmbH mit Sitz in Berlin[Ca22]. Das Hauptprodukt des Unternehmens stellt die Camunda Platform, ein umfangreiches Workflowmanagementsystem dar.

Grundsätzlich gibt es zwei unterschiedliche Varianten des Camunda Modelers:

- Der *Web Modeler* ist fester Bestandteil der Camunda Platform und wird aus dieser heraus gestartet.

- Der *Desktop Modeler* hingegen ist eine eigenständige Desktop-Applikation.

Darüber hinaus existiert noch das kostenlose und ohne Anmeldung nutzbare webbasierte Modellierungswerkzeug bpmn.io[12], welches eine sehr ähnliche Oberfläche wie der Camunda Modeler bietet und auch im Wesentlichen auf dem gleichen Code basiert[Uh21].

Die weiteren Ausführungen beziehen sich auf den Desktop Modeler, im Folgenden nur *Modeler* genannt. Die Ergebnisse sollten aber auch relativ einfach auf den Web Modeler und bpmn.io anwendbar sein. Für die Implementierung wurde der Modeler in Version 5.5.1 verwendet. Erste Tests mit der neueren Version 5.8.0 haben ebenfalls keine Fehler ergeben.

Exkurs: Camunda Platform

Die Camunda Platform existiert seit April 2022 in zwei Varianten:

- Camunda Platform 7 und
- Camunda Platform 8

Camunda Platform 7 wurde bis dato einfach Camunda Platform oder Camunda 7 genannt und basiert auf dem früheren Produkt Camunda BPM, welches auf dem Quelltext des weiter oben betrachteten Systems Activity basiert, aber schon 2013 davon abgespalten wurde [Wi22]. Bis heute basiert Camunda Platform 7 auf der Activity-Engine.

Camunda Platform 8 hingegen ist eine Weiterentwicklung der bis 2022 existierenden „Camunda Cloud“, die im Gegensatz zu Camunda 7 auf der eigens entwickelten Zeebee-Process-Engine baisiert [SRZ22].

Die beiden Versionen sollen insgesamt fünf Jahre, also bis 2027 parallel betrieben und mit Updates versorgt werden [De22a].

Technisch unterscheiden die beiden Versionen sich beispielsweise durch die Skalierbarkeit der Process Engines. Zeebee war von Anfang an auf den Cloudbetrieb ausgelegt und verwendet daher keine relationale Datenbank, was eine deutlich bessere Skalierbarkeit zur Folge hat [SRZ22]. Darüber hinaus haben die beiden Versionen aber auch einige unterschiedliche Funktionalitäten, die hier nicht näher betrachtet werden sollen. Nähere Informationen finden sich in [De22a] und [SRZ22].

Auf Grund der Auswahlmöglichkeit im Modeler, für welche der beiden Plattformversionen ein Prozess entwickelt werden soll (siehe Abbildung 14.1), ist diese Unterscheidung relevant.

Da die beiden Versionen der Plattform unterschiedliche Funktionen haben, unterstützen sie nicht unbedingt Modelle, die für die jeweils andere Variante entwickelt wurden.

[12] `https://bpmn.io`

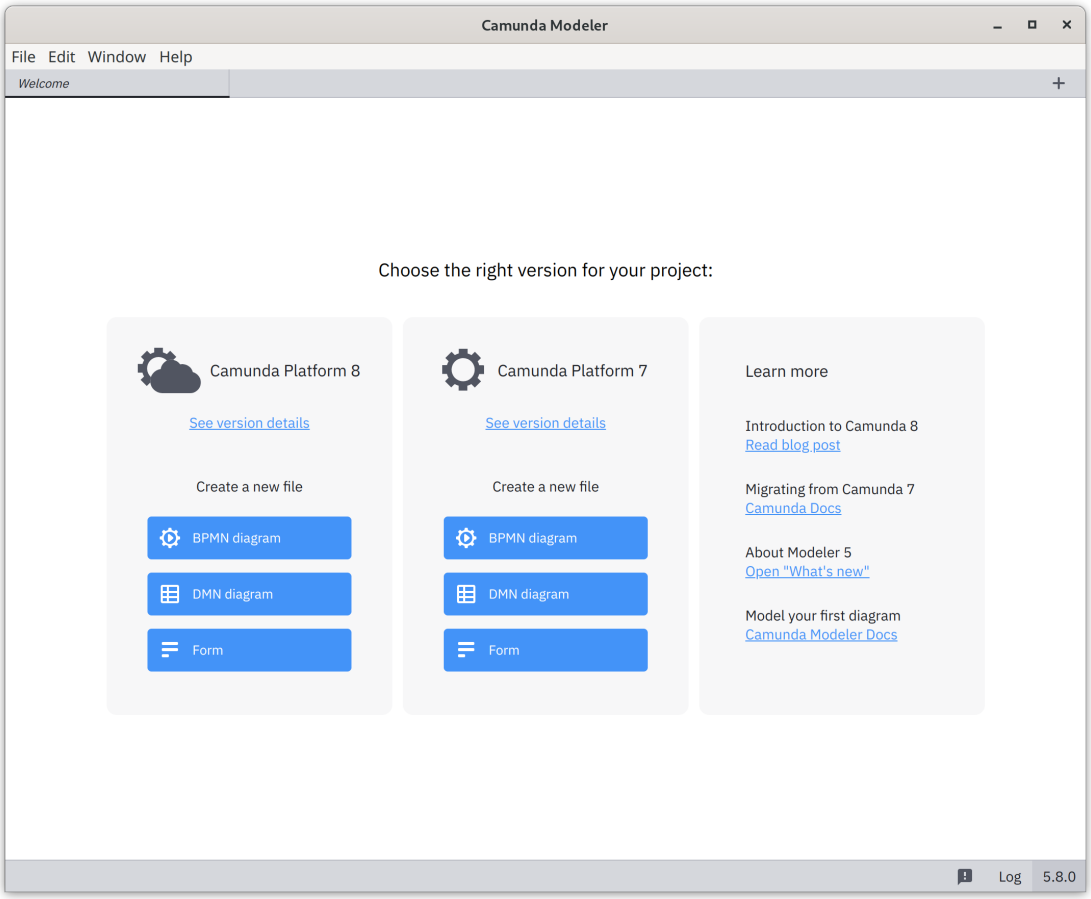


Abbildung 14.1.: Startseite des Camunda Modelers

Somit unterscheidet sich auch das Verhalten des Modelers abhängig von der Auswahl der Plattform zu Beginn in geringem Maße. Insbesondere erscheint aktuell[13] noch eine Fehlermeldung bei der Verwendung von abstrakten Aufgaben, wie in Abbildung 14.2 zu erkennen ist.

Abstrakte Aufgaben mögen zwar für die Automatisierung eines Prozesses problematisch sein, für die Modellierung, zu Zwecken der Veranschaulichung und Dokumentation, sind sie aber durchaus relevant. Das Problem kann zwar umgangen werden, indem auf die Nutzung der Version 8.2 umgeschaltet wird, diese liegt aber zur Zeit nur in einer Alpha-Version vor.

Insgesamt wirkt die Camunda Platform 8 noch nicht wirklich ausgereift. Daher wird im Folgenden im Modeler immer Camunda Platform 7 ausgewählt. Die Ergebnisse sollten sich aber künftig auch auf die neue Version übertragen lassen.

[13]Version 5.8.0 des Camunda Desktop Modelers

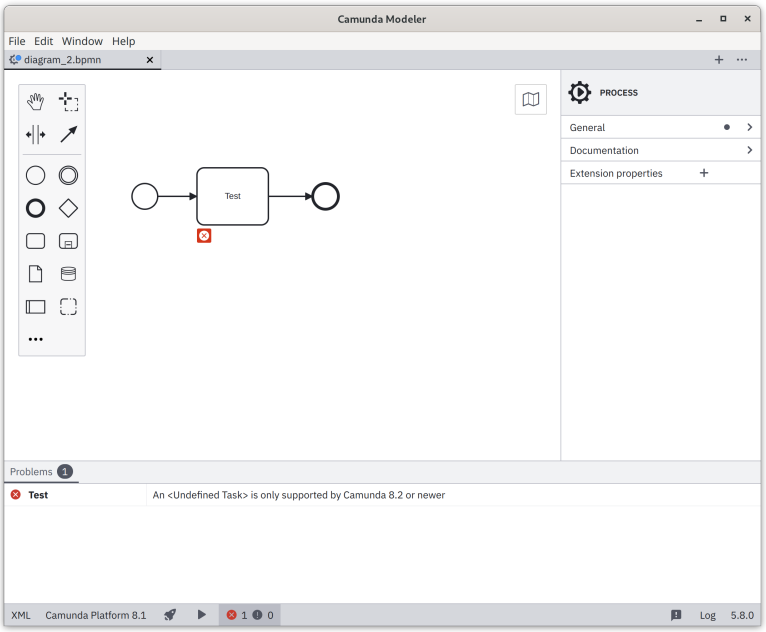


Abbildung 14.2.: Fehler bei Auswahl von Camunda Platform 8 und abstrakter Aufgabe

Funktionen/Benutzeroberfläche

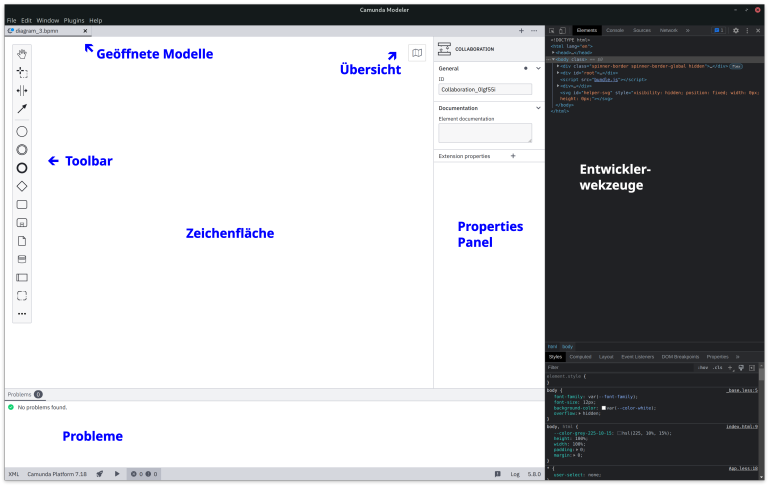


Abbildung 14.3.: Oberfläche des Modelers

Die Benutzeroberfläche des Modelers ist recht übersichtlich, wie der Screenshot in Abbildung 14.3 zeigt. Den größten Teil des Fensters macht die Zeichenfläche aus, auf der das Prozessmodell entsteht. Am linken Rand der Zeichenfläche findet sich eine Toolbar mit den verschiedenen Modellelementen und einigen allgemeinen Werkzeugen. In der rechten oberen Ecke der Zeichenfläche kann eine Übersichtskarte eingeblendet

werden, falls das Modell nicht auf die Zeichenfläche passt. Ganz oben im Fenster findet sich neben der Menüleiste eine weitere Leiste, in der geöffnete Modelle in Form von Tabs angezeigt werden. Im unteren Teil des Fensters befindet sich eine Übersicht, in der vorhandene Probleme im Modell angezeigt werden (siehe hierfür Abbildung 14.2). Darunter findet sich noch eine Leiste mit einigen Funktionsbuttons.

Im Menü können zwei Sidebars eingeblendet werden. Einerseits ist dies das „Properties Panel“, in welchem die Attribute der Modellelemente verändert werden können. Hier kann beispielsweise auch ein Skript für Skript-Tasks hinterlegt werden. Die andere Sidebar beinhaltet die Entwicklerwerkzeuge, wie sie auch in Webbrowsern verwendet werden können.

Für die eigentliche Modellierung werden mehrere Optionen unterstützt. Die Modellelemente aus der Toolbar können per Drag and Drop auf die Zeichenfläche gezogen werden. Alternativ kann auch ein Modellelement und anschließend die gewünschte Stelle der Zeichenfläche angeklickt werden. Zusätzlich können viele Elemente auch direkt von einem schon bestehenden Modellelement ergänzt werden. Hierbei werden dann auch automatisch die entsprechenden Konnektoren eingebaut.

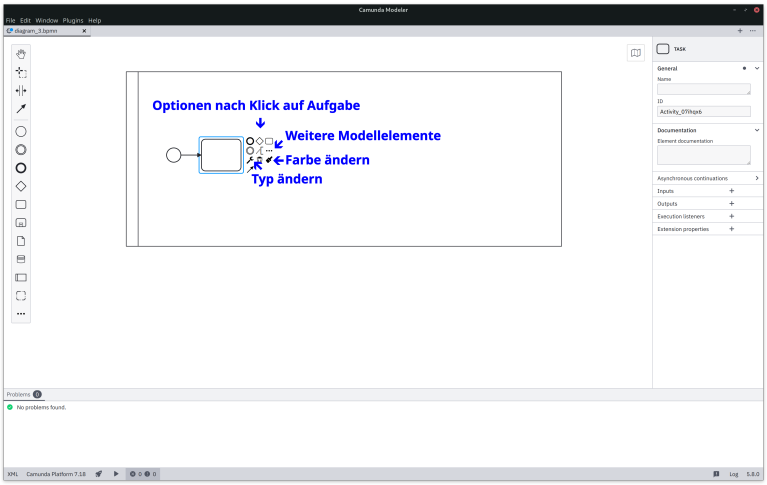


Abbildung 14.4.: Optionen für eine Aufgabe

In der Toolbar finden sich alle Modellelemente nur in einer recht abstrakten Form. Es existieren beispielsweise nur abstrakte Aufgaben und Ereignisse und auch nur exklusive Gateways. Um diese in die korrekte Form zu bringen, muss das jeweilige Element zunächst in der abstrakten Form auf die Zeichenfläche gesetzt werden. Dann kann dieses angeklickt werden, worauf sich ein Optionsmenü öffnet, in dem sich auch ein Button zum Bearbeiten findet. Das Optionsmenü ist in Abbildung 14.4 abgebildet. Hervorzuheben ist hier auch noch der Button der die Farbe des Modellelements verändert.

Die meisten Syntaxfehler werden schon bei der Modellierung ausgeschlossen, indem eine Falschmodellierung gar nicht erst ermöglicht wird. Es kann etwa, sobald ein Pool vorhanden ist, keine Aufgabe außerhalb dieses Pools auf die Zeichenfläche gesetzt werden. Auch die Konnektoren haben beispielsweise automatisch die korrekte Form.

Technische Grundlagen

Für das Verständnis des Erweiterungsmechanismus müssen zunächst einige technische Grundlagen zum Modeler erläutert werden. Die Informationen in diesem Abschnitt stammen überwiegend aus der offiziellen Camunda Dokumentation[14], sowie den jeweils angegebenen Websites der einzelnen Technologien.

Wie bereits in Abschnitt 14.1.2 erwähnt, existiert neben dem Modeler auch noch eine Webversion mit dem Namen bpmn.io. Beide Systeme haben die gleiche Basis. Hierfür wird das Framework *Electron*[15] verwendet, welches auf der verbreiteten JavaScript-Laufzeitumgebung *Node.js*[16] und dem Open Source Browser *Chromium*[17] basiert. Electron unterstützt so die Entwicklung plattformunabhängiger Anwendungen mit JavaScript, HTML und CSS.

Für die Modellierung der Prozessmodelle wurde die Bibliothek *bpmn-js*[18] entwickelt, welche wiederum *diagram-js*[19] als Grundlage verwendet. Die Bibliothek unterstützt hier nicht nur die grafische Darstellung, sondern kennt auch die Logik hinter BPMN, was beispielsweise für die Syntaxprüfung verwendet werden kann. Abbildung 14.5 stellt die Funktionen dar und veranschaulicht den Zusammenhang der Bibliotheken.

Die in Abschnitt 14.1.2 vorgestellte Benutzeroberfläche basiert neben bpmn-js auf der JavaScript-Bibliothek *React*[20].

Plugin Entwicklung

Camunda stellt als Startpunkt für die Entwicklung von Plugins ein Beispiel-Plugin zur Verfügung, welches alle nötigen Konfigurationen enthält. Der größte Teil dieses Beispiels kann für die Entwicklung eigener Erweiterungen weiterverwendet werden. Es müssen (je nach konkretem Einsatzzweck) nur wenige Dateien angepasst werden.

Die Verzeichnisstruktur des Beispiel-Plugins findet sich in Abbildung 14.6.

Relevant für die Anpassung des Beispiel-Plugins an die eigenen Bedürfnisse sind hier die beiden fettgedruckten Dateien **ExampleBpmnJsExtension.js**, in der die eigentliche

[14] https://docs.camunda.io/
[15] https://www.electronjs.org/de/
[16] https://nodejs.org/en
[17] https://www.chromium.org/chromium-projects/
[18] https://bpmn.io/toolkit/bpmn-js/
[19] https://github.com/bpmn-io/diagram-js/
[20] https://react.dev/

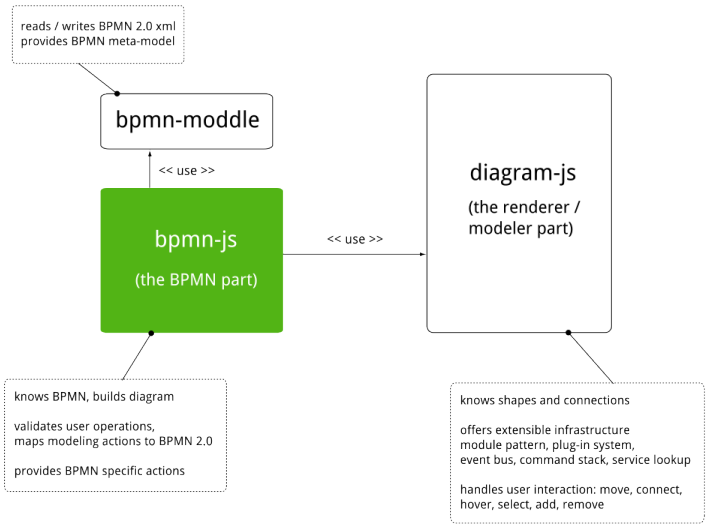


Abbildung 14.5.: Architektur von bpmn-js (aus [bp22])

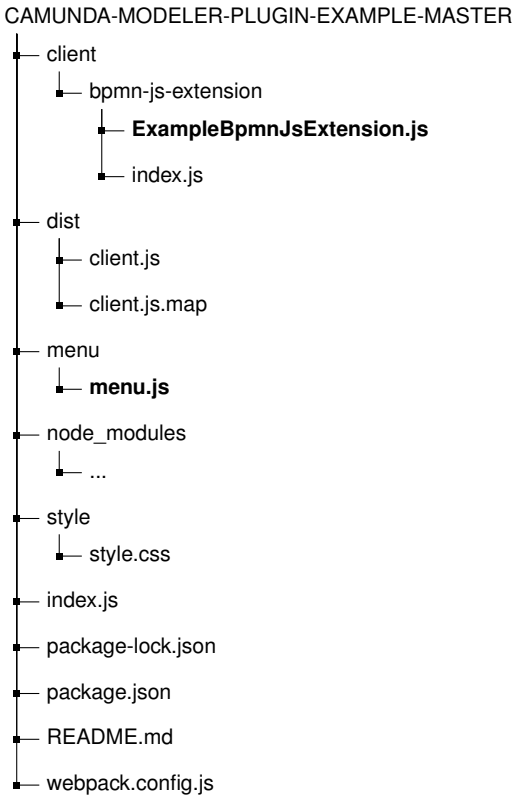


Abbildung 14.6.: Verzeichnisstruktur des Beispiel-Plugins

Funktionalität implementiert wird und **menu.js**, in welcher Menüeinträge definiert werden.

Wenn das fertig entwickelte Plugin zu Camunda hinzugefügt werden soll, muss dieses zunächst mit npm[21] gebaut werden. Anschließend wird das ganze Verzeichnis in das Verzeichnis des Modelers unter dem Pfad *resources/plugin* kopiert. Nach einem Neustart des Modelers ist das Plugin integriert und betriebsbereit.

14.2. Aufbau des Plugins

Gemäß den Erläuterungen in Abschnitt 14.1.2 wurde ein Plugin entwickelt, welches den Modeler um die verbleibenden funktionalen Anforderungen aus Abschnitt 13.3 ergänzt. Das ist im Wesentlichen die automatische Kategorisierung von Modellelementen.

Die Entwicklung des Plugins wird ausführlich in der im Rahmen der Dissertation betreuten Abschlussarbeit von Semir Velovic [Ve23] beschrieben. Dieser Abschnitt fasst die wichtigsten Entwicklungsschritte zusammen.

14.2.1. Menü-Einträge

Um das Plugin auszuführen, wurden Menüeinträge in den Modeler eingefügt. Hierfür müssen zwei Dateien verändert werden.

Einerseits ist das die Datei `menu.js` (siehe Listing C.1) im Verzeichnis `menu`. Hier wird der eigentliche Menüeintrag in den Modeler eingefügt. Dazu wird automatisch ein neues Menü *„Plugins“* erstellt, in welchem die einzelnen Menüpunkte erscheinen. Im Wesentlichen wird hier eine Bezeichnung definiert und mit einer Aktion verknüpft.

Die Aktion wiederum wird mit sämtlicher weiterer Funktionalität des Plugins in der Datei `DatenschutzPlugin.js` (siehe Listing C.2) im Verzeichnis `client/bpmn-js-extension` definiert.

Insgesamt wurden auf diese Weise vier Menüeinträge definiert und mit einer Funktionalität belegt:

1. *Aktivitäten überprüfen*
2. *Datenobjekte überprüfen*
3. *Datenobjekte aktualisieren*
4. *Zurücksetzen*

Die ersten beiden Menüpunkte stoßen hier jeweils die Kategorisierung der Aktivitäten bzw. Datenobjekte an. Grundsätzlich können diese beiden Menüpunkte auch zusammengefasst werden. Für die Entwicklung und erste Tests haben sich separate Einträge aber bewährt.

[21] https://www.npmjs.com/

Der Menüeintrag *„Datenobjekte aktualisieren“* dient dem Erweitern des Wörterbuchs (siehe Absatz 14.3.1). Für Aktivitäten gibt es keinen entsprechenden Eintrag, da hier ein anderes Verfahren zur Kategorisierung verwendet wird (siehe Abschnitt 14.3.2).

„Zurücksetzen“ letztlich entfernt alle Kategorisierungen und färbt das ganze Prozessmodell wieder weiß. Dieser Menüeintrag wurde auch hauptsächlich zu Testzwecken eingebaut. Ob diese Funktion auch im Praxiseinsatz nützlich ist, müssen weitere Tests zeigen.

14.2.2. Filterung der relevanten Modellobjekte

Eine zentrale Funktion des Plugins ist die Filterung relevanter Modellelemente. In Kapitel 8 wird dargelegt, dass generell nicht alle Elemente eines Prozessmodells relevant für den Datenschutz und dementsprechend auch nicht einzufärben sind. Daher müssen die relevanten Elemente zunächst aus dem Prozessmodell gefiltert und entsprechend ihres Typs in einzelnen Sammlungen zusammengefasst werden.

Hierfür kann auf die `elementRegistry` zurückgegriffen werden, in der alle Modellelemente mit den nötigen Attributen gespeichert sind. Insbesondere ist das eine ID, die Bezeichnung und auch der Typ. Listing 14.1 zeigt exemplarisch das Filtern von Aufgaben, die den Typ *„bpmn:Task“* haben. Im Prototyp wird bisher nur die Kategorisierung von Aufgaben und Datenobjekten umgesetzt. Daher werden auch nur diese gefiltert. Mit anderen Modellelementen kann aber analog verfahren werden.

```
1 const activities = elementRegistry.filter(function (element) {
2         if (element.type == "bpmn:Task") return element;
3       });
```

Listing 14.1: Filtern von Aufgaben

Einfärbung der Modellelemente

Wie bereits erwähnt, bietet der Modeler bereits die Möglichkeit, Modellelemente manuell einzufärben. Hierfür wird ein bestimmtes Farbschema verwendet: Alles, was in einem (farblosen) Standard-BPMN-Modell schwarz ist[22], wird jeweils in einer kräftigen Variante der gewählten Farbe[23] eingefärbt und die weißen Teile der Modellelemente werden in einer etwas blasseren Schattierung der jeweiligen Farbe eingefärbt. Beispiele finden sich etwa in Abbildung 8.2 und 8.3. Dieses Farbschema und auch die konkreten

[22] Das sind insbesondere die Randlinien und die Beschriftung des Elements sowie eventuelle Zusatzsymbole im Modellelement.

[23] Blau, Orange (Gelb), Grün, Rot und Violett

Farbwerte sollen auch bei der automatisierten Färbung beibehalten werden. Hierfür wurden für jede Farbkategorie die entsprechenden HEX-Farbcodes ermittelt. Mit der Anweisung aus Listing 14.2 wird die entsprechende Färbung (im Beispiel Rot) auf das jeweilige Modellelement angewandt.

```
commandStack.execute('element.setColor', {
    elements: [element],
    colors: {
        fill: "#ffcdd2",
        stroke: "#e53935"
});
```

Listing 14.2: Färben von Modellelementen

14.3. Kategorisierung

Kern des Plugins ist die Kategorisierung der einzelnen Modellelemente, also die semantische Analyse und anschließende Zuordnung einer Farbkategorie. Hierfür kommen verschiedene Ansätze in Frage, die im Folgenden erläutert werden. Die ersten beiden Ansätze basieren lediglich auf den Bezeichnungen der zu klassifizierenden Modellelemente und wurden im Prototyp umgesetzt. Der letzte Ansatz ist deutlich komplexer und wird nur theoretisch beschrieben.

14.3.1. Wörterbuch

Ein Ansatz der Kategorisierung ist ein mit der Zeit wachsendes Wörterbuch (siehe Abschnitt 12.1.1).

Hierfür müssen zunächst einige Prozesse manuell analysiert und kategorisiert werden. Die gewählte Kategorisierung wird durch das Plugin dann in dem Wörterbuch gespeichert. Beim Aufruf der automatischen Kategorisierungsfunktion werden die Bezeichnungen der Modellelemente mit dem Inhalt des Wörterbuchs abgeglichen. Wird eine Übereinstimmung gefunden, wird das Modellelement entsprechend kategorisiert und eingefärbt. Falls keine Übereinstimmung gefunden wird, erfolgt auch keine Kategorisierung und das Modellelement bleibt ungefärbt. Der Anwender kann anschließend die noch verbleibenden Modellelemente manuell einfärben und die Kategorisierung in dem Wörterbuch speichern. So werden mit der Zeit immer mehr Elemente automatisch gefärbt und folglich wird immer weniger manueller Aufwand nötig sein. Abbildung 14.7 zeigt den Ablauf in Form eines Prozessmodells.

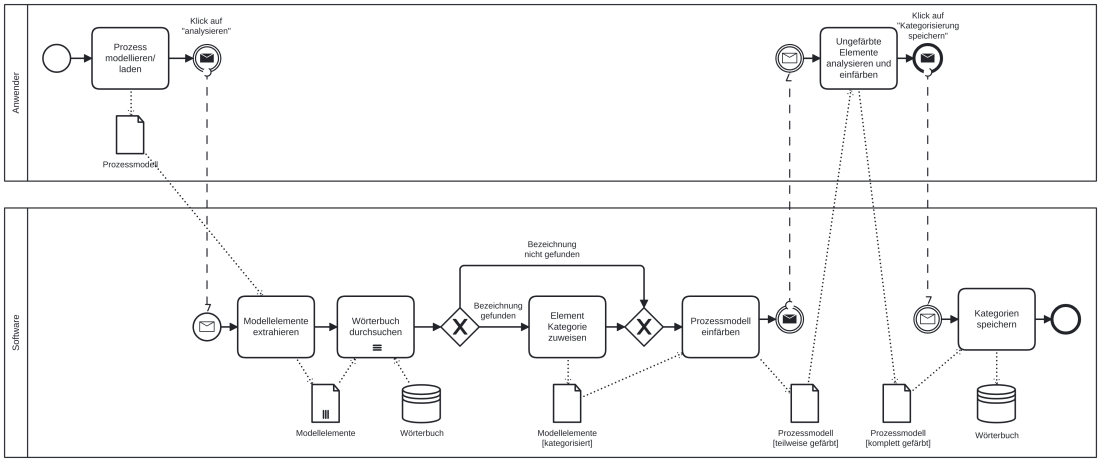


Abbildung 14.7.: Prozessmodell für die Kategorisierung mittels Wörterbuch

Umsetzung

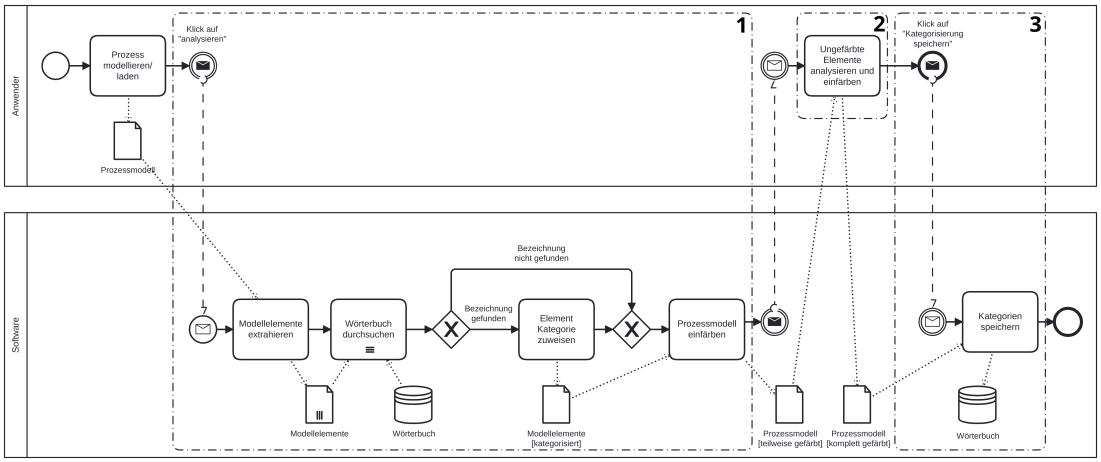


Abbildung 14.8.: Schritte im Prozessmodell für die Kategorisierung mittels Wörterbuch

Das beschriebene Vorgehen besteht aus drei wesentlichen Schritten, die in Abbildung 14.8 zu sehen sind und im Folgenden näher ausgeführt werden.

1. Der Abgleich der Modellelemente mit dem Wörterbuch und die entsprechende Kategorisierung und Einfärbung

2. Die manuelle Einfärbung der Elemente, die nicht automatisch kategorisiert werden konnten

3. Das Hinzufügen der manuellen Kategorisierung zum Wörterbuch

Dieses Verfahren wurde im Prototyp für die Kategorisierung der Datenobjekte implementiert. Eine Umsetzung für die anderen Modellelemente ist aber unproblematisch.

Im Folgenden wird zunächst das Wörterbuch an sich und anschließend die Implementierung der oben genannten drei Ablaufschritte erläutert.

Struktur des Wörterbuchs Die Struktur des Wörterbuchs ist recht einfach. Hier muss im Wesentlichen jeweils ein String, der die Bezeichnung des jeweiligen Elements enthält, mit je einem Wert für eine der drei Kategorien gemappt werden. Der Datentyp für die Kategorie kann sogar flexibel gewählt werden. Es sind beispielsweise Ganzzahlen mit den Werten 1-3 denkbar oder auch Strings, welche die Farben repräsentieren.

Für die Umsetzung kommen daher auch mehrere Optionen in Frage. Einerseits könnte natürlich eine Datenbank aufbaut werden. Hierbei ist der Aufwand allerdings in Anbetracht der geringen Anforderungen vergleichsweise groß. Alternativ könnte beispielsweise eine CSV-Datei genutzt werden.

Zum Test wurde, um den Aufwand vergleichsweise gering zu halten, eine Online-Datenbank[24] verwendet, auf die über eine REST-Schnittstelle zugegriffen werden kann.

1. Abgleich mit dem Wörterbuch Der erste Schritt im Prozess aus Abbildung 14.8 ist der Abgleich mit dem Wörterbuch. Dieser wird bei der ersten Nutzung natürlich noch kein Ergebnis liefern, da das Wörterbuch in diesem Moment noch leer ist. An dieser Stelle wird aber der Fall betrachtet, dass schon einige initiale Daten im Wörterbuch eingetragen sind.

Das Plugin wird wie in Abschnitt 14.2.1 durch einen Menüeintrag aufgerufen. Dann wird die erstellte Liste der Datenobjekte (siehe Abschnitt 14.2.2) sequenziell mit dem Wörterbuch verglichen.

Um Tippfehler und ähnliches als Fehlerquelle größtenteils auszuschließen, wird hierfür jeweils eine Levenshtein-Distanz berechnet (siehe Abschnitt 5.2.1) und nur Einträge mit einer Ähnlichkeit von mindestens 90% betrachtet. Hier muss abgewogen werden, welcher Schwellenwert am sinnvollsten zum Einsatz kommen sollte. Bei Verwendung eines geringeren Schwellenwerts werden tendenziell mehr Modellelemente eingefärbt, was grundsätzlich natürlich wünschenswert ist. Allerdings ist die Fehlerquote hier auch entsprechend höher. Beides betrifft verschiedene nichtfunktionale Anforderungen. Bei der Fehlerquote handelt es sich um die Korrektheit. Dieser wird in Abschnitt 13.4 eine mittlere Priorität zugeschrieben. Die Anzahl der kategorisierten Modellelemente lässt sich je nach Betrachtungswinkel entweder der Bedienbarkeit oder dem Zeitverhalten zuordnen. Beides hat ebenfalls eine mittlere Priorität. Es wird in der Definition der nichtfunktionalen Anforderungen aber auch darauf hingewiesen, dass die höchste Priorität

[24] https://restdb.io/

die Präsentation der Funktionalitäten ist. Die Korrektheit ist diesem Aspekt zuzuordnen. Deshalb wird zunächst ein recht hoher Schwellenwert verwendet.

Wenn hier mehrere Einträge in Frage kommen, wird jeweils der mit der größten Übereinstimmung gewählt. Schwieriger ist die Entscheidung für den Fall, dass mehrere Einträge mit einer exakt gleichen Übereinstimmung, aber unterschiedlichen Klassen, vorhanden sind. Die jeweils unkritischere Klasse sollte in keinem Fall gewählt werden, da das den Nutzer unter Umständen fälschlicherweise in Sicherheit wiegt. Eine plausible Möglichkeit ist aber die Verwendung der jeweils kritischeren Klasse. In den Interviews aus Abschnitt 10.1 wurde allerdings besprochen, dass es sinnvoller sein könnte, die entsprechenden Modellelemente einfach unkategorisiert zu lassen. Hier kommt es auf den konkreten Einsatzzweck des Systems an. Wenn die Hauptzielgruppe Datenschutzbeauftragte sind, ist es sicherlich sinnvoll, in diesen Fällen keine Kategorisierung vorzunehmen, sondern diese dem Nutzer zu überlassen. Bei Anwendern mit eher weniger Fachwissen im Bereich Datenschutz bietet die kritische Kategorisierung aber den Vorteil, dass dann eher der korrekte Umgang mit dem entsprechenden Datenobjekt geprüft wird. Daher wird dieser Ansatz umgesetzt. Falls gar keine Übereinstimmung gefunden wird, wird ein entsprechender Fehlerwert zurückgegeben.

2. Manuelle Kategorisierung Falls nicht für alle Datenobjekte eine Übereinstimmung im Wörterbuch gefunden wird, müssen die verbleibenden Objekte manuell kategorisiert werden.

Da der Camunda Modeler ohnehin die Färbung der Modellelemente ermöglicht, muss hierfür aber nichts weiter implementiert werden.

3. Erweiterung des Wörterbuchs Die Grundidee des Ansatzes ist es, dass das Wörterbuch im Laufe der Zeit ohne großen zusätzlichen Aufwand wächst. Daher kann die manuelle Kategorisierung durch einen weiteren Menüpunkt (siehe Abschnitt 14.2.1) direkt in das Wörterbuch übertragen werden. Hier ist zu überlegen, ob eine automatische Speicherung, etwa nach jeder Farbänderung, sinnvoller wäre. Das hätte den Vorteil, dass der Nutzer die Speicherung nicht vergessen könnte und generell eine Aktion weniger nötig wäre. Dies wirkt sich positiv auf die nichtfunktionalen Anforderungen „Bedienbarkeit“ und „Fehlerschutz“ aus. Allerdings besteht hier die Gefahr, dass eine Fehlerhafte Kategorie gespeichert wird. Das kann passieren, wenn der Nutzer sich während der Kategorisierung zunächst unsicher ist und erst die eine Kategorie, später aber noch eine andere Kategorie wählt. Aber auch ein bloßes „Verklicken“ könnte zu derartigen Fehlern führen. Daher wird auf die aktive Speicherung durch den Nutzer zurückgegriffen.

Bei Auswahl dieses Menüpunkts werden wieder alle im Prozessmodell vorhandenen Datenobjekte mit dem Wörterbuch abgeglichen. Wenn für ein Datenobjekt kein Eintrag vorhanden ist, wird ein neuer Eintrag zum Wörterbuch hinzugefügt.

Hier stellt sich die Frage, wie verfahren wird, wenn ein bestimmter Begriff schon im Wörterbuch vorhanden ist, aber vom Modellierer eine andere Kategorie vergeben wird. Grundsätzlich gibt es hier drei Optionen:

1. Ein Update der Kategorie,
2. das Ignorieren der Änderung und
3. die Speicherung beider Kategorien in zwei Einträgen.

Da zwei Datenobjekte in unterschiedlichen Kontexten durchaus unterschiedlich bewertet werden können, implementiert der aktuelle Prototyp die dritte Variante.

Bewertung

Insgesamt bringt dieser Ansatz zur Kategorisierung zwar eine gute Korrektheit, aber das Problem eines relativ hohen initialen Aufwands für die Anwender mit sich. Ein möglicher Optimierungsansatz hierfür ist ein zentrales Wörterbuch, welches alle Nutzer der Software, unabhängig von ihrer Organisation, verwenden können. Hierbei würden zumindest Nutzer aus Organisationen, welche die Software erst vergleichsweise spät einführen, von dem Vorteil eines bereits bestehenden umfangreichen Wörterbuchs profitieren, sodass wahrscheinlich auch schon bei der ersten Nutzung einige Modellelemente eingefärbt werden können. Das ist allerdings stark abhängig von den verwendeten Bezeichnungen. Diese werden höchstwahrscheinlich zumindest branchen-, wenn nicht gar unternehmensabhängig, sein.

14.3.2. Machine Learning

Ein weiterer Ansatz für die Klassifizierung der Modellelemente nutzt Verfahren aus den Bereichen ML und NLP. Teile hiervon werden ausführlicher in der Abschlussarbeit von Florim Peci beschrieben, welche im Rahmen dieser Dissertation entstanden ist (siehe[Pe22]). Dieser wurde für den Prototyp exemplarisch für die Kategorisierung der Aktivitäten verwendet. Daher beziehen sich die folgenden Ausführungen auch in erster Linie auf die Aktivitäten. Es wurde aber auch ein Modell für Datenobjekte erstellt, welches auf Grund der wenigen Trainingsdaten aber nicht sinnvoll zum Einsatz kommen kann.

Generell handelt es sich bei der vorliegenden Problemstellung um ein Klassifizierungsproblem. Hierfür eignet sich gut das überwachte Lernen (siehe Abschnitt 5.1.1).

Der Ansatz besteht aus mehren Schritten, die in Abbildung 14.9 in Form eines Prozessmodells dargestellt sind und im Folgenden näher beschrieben werden.

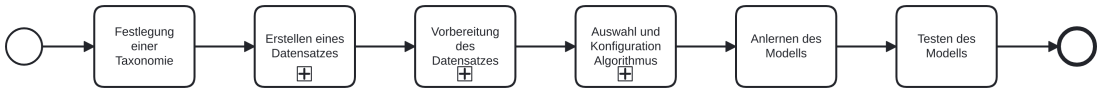


Abbildung 14.9.: Ablauf der Klassifizierung mit Machine Learning

Festlegung einer Taxonomie für die Klassen

Für die Klassifizierung müssen zunächst die nötigen Zielklassen definiert werden. Hierfür kann die Taxonomie aus Abschnitt 12.2.3 verwendet werden. Auf dieses Vorgehen beziehen sich auch die folgenden Ausführungen. Allerdings kann diese Taxonomie nicht direkt auf die drei Farbkategorien übertragen werden. Daher wurde ein weiteres Modell angelernt, bei welchem nur die drei Farben als Klassen verwendet wurden. Dieses kommt auch im Prototyp zum Einsatz.

Erstellen eines Datensatzes

Anschließend muss ein Datensatz erstellt werden, mit welchem das Modell angelernt werden kann. Grundlage hierfür bilden Prozessmodelle der CAU, wie sie exemplarisch auch in Abschnitt 7.3 verwendet werden. Insgesamt stehen 339 Prozesse (inklusive Teilprozessen) zur Verfügung. Diese beinhalten zusammen 3812 Aufgaben und 198 Datenobjekte.

Filterung Für die Erstellung des Datensatzes werden zunächst die jeweils relevanten Objekte aus den XML-Dateien der Prozesse extrahiert. Als Auswahlkriterium dient hier die Bezeichnung des XML-Tags (für Aufgaben `bpmn2:task`). Für jede Aufgabe wird jeweils die Bezeichnung, also der Inhalt der Aufgabe, gespeichert.

Bereinigung Der so erstellte Datensatz enthält einige Einträge, die bei der Weiterverarbeitung zu Problemen führen können und muss daher bereinigt werden. Hierbei werden Duplikate und fehlerhafte (z.B. leere) Einträge entfernt. Nach der Bereinigung sind noch 2349 Einträge vorhanden.

Klassifizierung Um ein Dataframe für das Anlernen des Algorithmus zu erhalten, muss nun eine manuelle Klassifizierung erfolgen. Hierfür wurde allen Aufgaben eine Klasse der Taxonomie zugeordnet und in einer CSV-Datei gespeichert.

Erstellung des Dataframe Mit dem so vorbereiteten Datensatz kann ein Dataframe in Python erstellt werden. Hierfür wird das Werkzeug *Pandas*[25] verwendet. Vorab muss der Datensatz noch weiter bereinigt werden. Hierfür werden die folgenden Schritte ausgeführt (siehe auch Abschnitt 5.2.2):

- Transformation in Kleinschreibung
- Entfernen von Punktzeichen
- Lemmatisierung
- Entfernen von Stoppwörtern

Abschließend müssen die so überarbeiteten Zeichenketten in einen Vektor überführt werden. Hierfür wird die TF-IDF-Methode (siehe [Ko19]) verwendet.

Außerdem werden die natürlichsprachlichen Bezeichnungen der Klassen als Ganzzahlen codiert.

Aufteilen des Dataframe Als letzter Vorbereitungsschritt muss das fertige Dataframe noch in einen Trainings- und einen Testdatensatz aufgeteilt werden. Auf Grund der vergleichsweise geringen Datenmenge, werden 85% der Einträge als Trainingsdatensatz verwendet, um ein möglichst valides Ergebnis zu erreichen.

Auswahl eines Algorithmus

Für das Anlernen des Modells muss ein Algorithmus ausgewählt werden. Zu Testzwecken werden zwei verschiedene Algorithmen verwendet und die Ergebnisse miteinander verglichen. Dies ist einerseits K-nearest neighbor (KNN) und andererseits Support Vector Machines (SVM).

Festlegung der Hyperparameter

Nach der Festlegung der verwendeten Algorithmen können Hyperparameter ausgewählt werden um das Training zu optimieren. Hierbei werden die Parameter an Hand der Grid Search Methode ausgewählt.

Für den KNN-Algorithmus muss die Variable k bestimmt werden. Hierfür wurden einige Werte getestet, indem des Modell probeweise mit ihnen angelernt wurde. Das beste Ergebnis wurde mit $k = 1$ erzielt.

Beim SVM-Algorithmus muss einerseits der Regularisierungsparameter festgelegt werden, der eine „Strafe“ für Fehlklassifikationen vergibt. Nach mehreren Tests kann der

[25] https://pandas.pydata.org/

bestmögliche Regularisierungsparameter als *10* bestimmt werden. Andererseits muss noch eine Kernelfunktion festgelegt werden. Hier wurde das beste Ergebnis mit der *Gaussian Radial Basis* erzielt. Außerdem wird noch der Gamma-Parameter festgelegt, der den Einfluss eines einzelnen Datenpunkts steuert. Hier kann ein Wert von *0,1* als Optimum für den vorliegenden Fall festgestellt werden.

Anlernen und Testen des Modells

Mit der beschriebenen Konfiguration und dem Dataframe wurde ein Modell angelernt. Das beste Ergebnis hat die SVM-Methode erzielt. Hier konnte eine Testgenauigkeit von 69,405% ermittelt werden. Diese ist zwar nicht optimal, reicht für weitere Tests aber aus.

Integration in das Plugin

Um die Klassifizierung mit dem ML-Modell durchzuführen, wurde dieses auf einen externen Server geladen. Aus dem Modeler-Plugin wird mit Hilfe eines POST-Requests ein JSON-Objekt mit der Liste der zu kategorisierenden Aufgaben an die entsprechende Schnittstelle des Servers gesendet. Serverseitig werden die Aufgaben dann mit dem ML-Modell klassifiziert und anschließend wieder ein JSON-Objekt zurückgegeben. Dieses enthält eine Liste von Paaren aus der ID der Aufgabe und der jeweiligen Farbkategorie.

Die JSON-Antwort wird vom Plugin dann genutzt, um die Aufgaben entsprechend einzufärben (siehe Abschnitt 14.2.2).

Bewertung

Der große Vorteil des beschriebenen Vorgehens ist der geringe Aufwand für den Anwender. Es existieren aber auch Schwächen, die in den folgenden Abschnitten erläutert werden.

Datenbasis Wie oben erläutert, basiert die automatische Klassifizierung der Prozessmodelle auf einem Machine Learning Ansatz. Ein Grundproblem bei derartigen Verfahren ist, dass die Qualität der Ergebnisse immer direkt von der Qualität und Quantität der Trainingsdaten abhängt. Beides ist im untersuchten Themenfeld problematisch.

Einerseits wird eine verhältnismäßig große Menge an Trainingsdaten benötigt. Hierbei ist zu beachten, dass für das Training des Algorithmus nur Prozessmodelle in der Sprache und Notation verwendet werden können, in der auch die später zu klassifizierenden Daten vorliegen. Für diese Arbeit beschränken sich die möglichen Trainingsdaten also auf deutschsprachige BPMN-Modelle. Im Gegensatz zu Texten oder Bildern, werden Geschäftsprozessmodelle aber selten veröffentlicht. So werden im deutschsprachigen

Bereich in erster Linie Prozessmodelle veröffentlicht, die zu Ausbildungs- oder Demonstrationszwecken erstellt wurden. Diese sind aber häufig recht kurz und abstrakt gehalten und unterscheiden sich dementsprechend von tatsächlich genutzten Prozessmodellen aus Unternehmen. Letztere wiederum sind schwierig zu erhalten, da zum einen nur vergleichsweise wenige Unternehmen überhaupt ihre Geschäftsprozesse formalisieren. Und auch wenn BPMN als Standard angesehen werden kann, ist der Anteil der Unternehmen, die Prozesse in BPMN modellieren, entsprechend noch geringer. Zum anderen wollen aber natürlich auch viele Unternehmen ihre Geschäftsprozessmodelle externen Personen nicht preisgeben, u.a. da diese mitunter Geschäftsgeheimnisse enthalten können.

Daher lag für diese Arbeit nur ein recht geringer Datensatz vor, der für das Training verwendet werden konnte. Darüber hinaus stammt dieser Datensatz nur aus einer Quelle, dem Prozessmanagement der CAU. Das hat wiederum Auswirkungen auf die Qualität der Daten.

Da die verwendeten Daten größtenteils aus einer Organisation stammen, ist beispielsweise die Domäne relativ ähnlich. Zwar liegen Prozessmodelle zu recht unterschiedlichen Themen (z.B. Personalverwaltung, Erstellung von Studienausweisen, Beschaffung) vor, jedoch können alle Prozesse dem Bereich der öffentlichen Verwaltung zugeordnet werden. Außerdem arbeiten an der CAU nur relativ wenig Menschen an der Modellierung der Prozesse. All dies kann sich auf die verwendete Sprache innerhalb der Prozessmodelle auswirken. So lassen sich in den Modellen beispielsweise einige Fachbegriffe und Abkürzungen finden, die in anderen Organisationen - insbesondere aus anderen Branchen - so nicht zu finden sind. Das erschwert natürlich die Klassifizierung von Prozessen aus anderen Organisationen, in denen eben andere Begrifflichkeiten genutzt werden.

Außerdem ist bei der manuellen Klassifikation aufgefallen, dass die Modellelemente sich relativ ungleichmäßig auf die Klassen verteilen. So können etwa 458 Aufgaben der Klasse *„Kommunizieren oder Weitergeben“* zugeordnet werden, aber nur drei Aufgaben der Klasse *„Einschränken“*. Gerade für die Klassen mit sehr wenig Elementen ist die Aussagekraft des Modells sehr begrenzt.

Prozessmodelle Während der Recherche für diese Arbeit ist aufgefallen, dass viele Prozessmodelle nicht korrekt sind. Insbesondere sind viele Bezeichnungen nicht sprechend. So wurden in Modellen beispielsweise Datenobjekte gefunden, die schlicht als „Datenobjekt“ bezeichnet werden. Das ist für die Auswertung der Datenschutzrelevanz offensichtlich ein Problem, da maximal aus dem Kontext ersichtlich werden kann, um welchen Inhalt es sich hier handelt. Außerdem wird in vielen Prozessmodellen mit sehr speziellen Abkürzungen oder Fachbegriffen gearbeitet, die nur für Domänenexperten oder Unternehmensangehörige zu verstehen sind. Hier gestaltet sich das Training des

Algorithmus als sehr schwierig und folglich können auch keine korrekten Ergebnisse bei der Klassifizierung dieser Begriffe erwartet werden.

14.3.3. Ontologie/Reasoning

In Abschnitt 12.3 wird erläutert, warum ein komplexerer Ansatz zur Kategorisierung betrachtet werden sollte, der nicht nur die Bezeichnungen der Modellelemente, sondern noch weitere Aspekte einbezieht.

Hierfür bietet sich das Reasoning über eine Ontologie an. Ontologien, als ursprünglich philosophisches Konzept, finden sich heute in verschiedensten Bereichen. Im Kontext der Informatik wird eine Ontologie als „explizite Spezifikation einer Konzeptualisierung“ [Gr93] definiert. Gemeint ist hiermit einerseits die Definition einer Terminologie eines bestimmten Wissensbereichs, andererseits aber auch der Beziehungen untereinander sowie weiteren Regeln [He05]. Ziele des Ontologie-Entwurfs sind allen voran die einfachere Kommunikation von Wissen - sowohl zwischen Menschen, als auch zwischen Maschinen - und die Organisation dieses Wissens. Darüber hinaus kann eine korrekt definierte Ontologie aber auch für das automatische Schließen weiteren Wissens verwendet werden [GL02; He05]. Dieses Schließen wird auch als Reasoning bezeichnet.

Klassische Ontologien bilden absolute Zusammenhänge ab: Ein Tisch ist ein Möbelstück und er hat Beine und ein Mensch (der im übrigen auch Beine hat) kann ihn verwenden um ein Buch auf ihm abzulegen. Der Tisch, das Möbelstück, die Beine, der Mensch und das Buch und der Vorgang des Ablegens wären in diesem Beispiel Konzepte, die zusammen mit ihren Beziehungen in einer Ontologie definiert werden. Allerdings sind nicht alle Sachverhalte so klar. Selbst in diesem einfachen Beispiel könnte es ja durchaus sein, dass der Tisch keine Beine hat, sondern an der Wand oder der Decke befestigt ist und auch Menschen ohne Beine kommen durchaus vor. Daher wurden sogenannte probabilistische Ontologien eingeführt, die Wahrscheinlichkeiten für einen gewissen Zusammenhang einbeziehen, um die Realität korrekter abbilden zu können. Die Wahrscheinlichkeiten erschweren aber natürlich das automatische Schließen [Pe20; Ri15].

Im Folgenden werden einige grundlegende Gedanken zur Erstellung einer solchen probabilistischen Ontologie beschrieben, die alle nötigen Konzepte abbildet, um automatisch auf die Datenschutz-Kategorie eines BPMN-Modellelements zu schließen.

Grundsätzlich ist es im Bereich der Ontologien immer sinnvoll, bereits bestehende Ansätze zu betrachten und diese zu erweitern. Es finden sich auch bereits einige Ontologien, die sich sowohl mit dem Thema Datenschutz als auch mit der Prozessmodellierung, genauer mit BPMN, befassen. Zwei passende Ausgangspunkte können die

BPMN 2.0 Based Ontology for Business Process Representation (BBO) [26] , sowie die Data Protection Ontology von Bartolini et al. [27] darstellen.

Diese beiden Ontologien können um nicht relevante Konzepte reduziert und anschließend kombiniert werden. Außerdem müssen noch einige weitere Konzepte hinzugefügt werden.

Nach dem Aufbau der Ontologie kann mittels Reasoning auf neue Konzepte geschlossen werden. Im vorliegenden Fall ist dies die Kategorisierung von Begriffen in die drei Farbkategorien. Zu beachten ist hier, dass nicht in allen Fällen sicher geschlossen werden kann, da oft mehrere Faktoren betrachtet werden müssen, die sich unter Umständen widersprechen. Daher muss es sich um eine probabilistische Ontologie handeln. Das bedeutet, dass neben der Definition der Konzepte jeweils Wahrscheinlichkeiten angegeben werden müssen, die für das Reasoning gelten. So kann etwa mit einer Wahrscheinlichkeit von x% gelten, dass eine Aufgabe in einem Pool, der einen Betroffenen darstellt, grün markiert wird. Andererseits kann aber mit Wahrscheinlichkeit von y% gelten, dass die gleiche Aufgabe gelb oder rot zu färben ist, weil es sich um eine sendende Aufgabe handelt. Hier müssen durch viele Tests entsprechend sinnvolle Wahrscheinlichkeiten ermittelt werden.

Der Ontologie-Ansatz ist sehr komplex. Insbesondere für die Definition der Wahrscheinlichkeiten sind viele Tests nötig, die wiederum viele Testdaten verlangen. Es hat sich aber bereits bei der Umsetzung des ML-Ansatzes gezeigt, dass nicht hinreichend viele Daten vorliegen. Daher wird der Ansatz nicht weiter verfolgt.

14.4. Erweiterung des BPMN-Standards

Neben den eigentlichen Kernfunktionalitäten, also insbesondere der Färbung der Modellelemente, muss das eingefärbte Prozessmodell auch gespeichert werden können. Wie bereits in Abschnitt 3.1 erläutert, werden BPMN-Modelle in einem speziellen XML-Format gespeichert, welches auch erweiterbar ist. Dieser Abschnitt beschreibt eine entsprechende Erweiterung, welches auch die Semantik der Färbung grundlegend abbildet.

Für die Speicherung der reinen Färbung existiert bereits eine Erweiterung mit dem Titel *BPMN in Color*[28], welche unter anderem auch von Camunda implementiert wird. Hier werden zusätzliche Attribute eingeführt, mit denen Rahmen- und Hintergrundfarben in Form von Hexadezimal-Farbcodes definiert werden können.

[26] `https://www.irit.fr/recherches/MELODI/ontologies/BBO/index-en.html`
[27] `https://www.w3.org/community/dpvcg/wiki/Data_Protection_Ontology_by_Bartolini_et._al`
[28] https://github.com/bpmn-miwg/bpmn-in-color

Wie in Abschnitt 3.1.6 beschrieben, können Erweiterungen des BPMN-Standards auf verschiedene Arten definiert werden. An dieser Stelle wird die Erweiterung mit einem entsprechenden XML Schema umgesetzt, da dies im vorliegenden Fall die einfachste Variante darstellt. Im Verlauf dieses Abschnitts werden die notwendigen Erweiterungen in dieser Repräsentation dargestellt und erläutert. Grundlage bietet die Schema-Datei *Semantic.xsd* (siehe Abschnitt 3.1.5).

Vorweg können alle Erweiterungen aber auch schon Abbildung 14.10 entnommen werden.

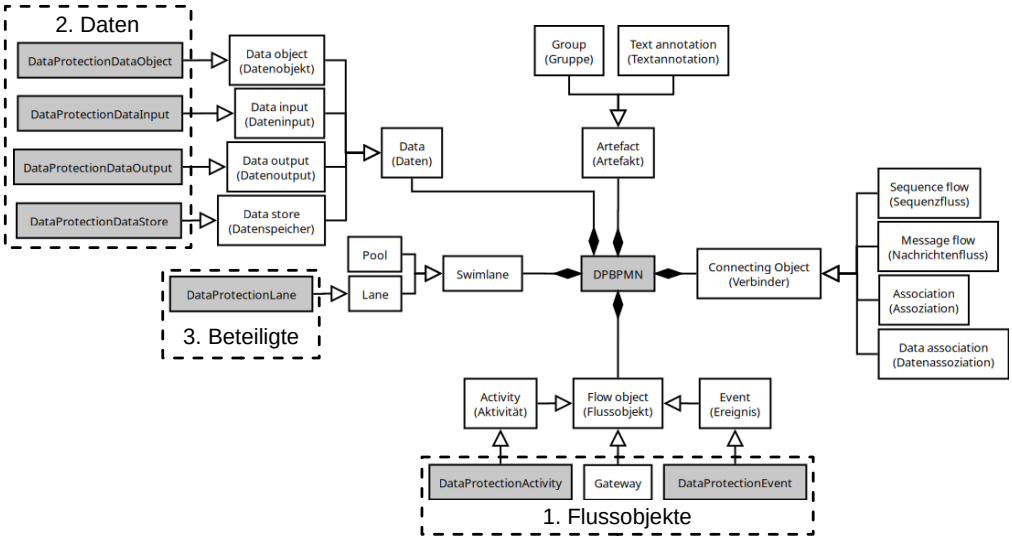


Abbildung 14.10.: Erweitertes BPMN-Metamodell

Die Erweiterungen können dabei in drei Gruppen zusammengefasst werden. Die beiden Klassen aus dem Bereich „Flussobjekte“ und die vier Klassen aus dem Bereich „Daten“ werden jeweils um ein Attribut namens `dataProtectionLevel` erweitert, welches in beiden Fällen die Werte *green*, *amber* und *red* (in Anlehnung an das TLP aus Abschnitt 6.3) annehmen kann.

Darüber hinaus wird die `Lane` um ein Attribut `subject`(Rechtssubjekt) erweitert, welches die Werte *dataSubject* (Betroffener), *controller* (Verantwortlicher), *processor* (Auftragsverarbeiter) und *thirdParty* (Dritter) annehmen kann. Obwohl in 2.1.5 noch ein weiteres Rechtssubjekt als relevant identifiziert wurde, so muss dieses - der Empfänger - hier nicht aufgenommen werden, da dieser Status einerseits nur temporär vorliegen kann und andererseits aus dem Kontext jederzeit automatisiert entnommen werden kann, ob ein Beteiligter in diesem bestimmten Moment ein Empfänger ist (siehe auch Abschnitt 12.3).

Grundlage: Datentypen

Für die Erweiterung der verschiedenen Klassen werden zunächst die beiden Datentypen für die oben beschriebenen Attribute benötigt. Da es sich in beiden Fällen um Aufzählungstypen handelt, ist das Vorgehen identisch.

```
1 <xsd:simpleType name="tDataProtectionLevel">
2     <xsd:restriction base="xs:string">
3         <xsd:enumeration value="green"/>
4         <xsd:enumeration value="amber"/>
5         <xsd:enumeration value="yellow"/>
6     </xsd:restriction>
7 </xsd:simpleType>
```

Listing 14.3: XML-Datentyp für das dataProtectionLevel

Listing 14.3 zeigt die Definition des Datentyps für die `dataProtectionLevel`. XML Schema sieht für Aufzählungstypen die Restriktion eines anderen Standarddatentyps vor. In diesem Fall wird der Datentyp `String` auf die Werte *green*, *amber* und *red* eingegrenzt.

```
1 <xsd:simpleType name="tSubject">
2     <xsd:restriction base="xs:string">
3         <xsd:enumeration value="dataSubject"/>
4         <xsd:enumeration value="controller"/>
5         <xsd:enumeration value="processor"/>
6         <xsd:enumeration value="thirdParty"/>
7     </xsd:restriction>
8 </xsd:simpleType>
```

Listing 14.4: XML-Datentyp für das subject

Analog zum `dataProtectionLevel` wird in Listing 14.4 das `subject`, also Rechtssubjekt, ebenfalls als Restriktion eines *Strings* definiert.

1. Flussobjekte

Auf Basis des `dataProtectionLevel` kann eine Schemadefinition für eine `dataProtectionActivity`, also eine BPMN-Aktivität mit einem zusätzlichen Attribut, welches das `dataProtectionLevel` abgibt, erstellt werden (siehe auch Abbildung 14.10).

Wie in Listing 14.5 zu sehen, wird hierfür einfach die entsprechende Klasse für eine Aktivität um ein Attribut mit dem oben erstellten Datentyp *tDataProtectionLevel* erweitert.

Analog dazu kann auch eine Erweiterung für Ereignisse definiert werden. Listing 14.6 zeigt die Definition des `dataProtectionEvent`.

```
1 <xsd:element name="dataProtectionActivity" type="
    tDataProtectionActivity"/>
2 <xsd:complexType name="tDataProtectionActivity">
3     <xsd:complexContent>
4         <xsd:extension base="tActivity"/>
5         <xsd:attribute name="dataProtectionLevel" type="
              tDataProtectionLevel"/>
6     </xsd:complexContent>
7 </xsd:complexType>
```

Listing 14.5: XML-Repräsentation der dataProtectionActivity

```
1 <xsd:element name="dataProtectionEvent" type="
    tDataProtectionEvent" substitutionGroup="flowElement"/>
2 <xsd:complexType name="tDataProtectionEvent">
3     <xsd:complexContent>
4         <xsd:extension base="tEvent"/>
5         <xsd:attribute name="dataProtectionLevel" type="
              tDataProtectionLevel"/>
6     </xsd:complexContent>
7 </xsd:complexType>
```

Listing 14.6: XML-Repräsentation des dataProtectionEvent

2. Daten

Auch die Notationselemente, die die verschiedenen Arten von Datenobjekten abbilden, werden auf vergleichbare Art erstellt. Zu beachten ist hier allerdings, dass nicht entsprechend der Darstellung in Abbildung 14.10 nur eine Oberklasse, sondern alle Kindklassen einzeln erweitert werden müssen. Das beruht auf der Tatsache, dass in der BPMN-Schemadefinition überhaupt keine Oberklasse für die verschiedenen Datenelemente besteht, sondern diese alle vom allgemeinen Typ *„BaseElement“* abgeleitet werden. Von diesem Typ werden aber auch einige völlig andere Elemente abgeleitet, sodass eine allgemeine Erweiterung dieses Typs hier keinen Sinn ergibt.

Somit wird also in Listing 14.7 ein *dataProtectionDataObject* von der Klasse `tDataObjectReference` abgeleitet, indem - analog zum oben beschriebenen Vorgehen (siehe Abschnitt 14.4) - ein Attribut `dataProtectionLevel` hinzugefügt wird. Zu beachten ist hier, dass im BPMN-Schema sowohl die Klasse `tDataObject` als auch die Klasse `tDataObjectReference` existieren. Letztere stellt im Wesentlichen eine Referenz auf ein Datenobjekt dar, die um einen Status erweitert wird. Es können also zu einem `DataObject` mehrere Elemente der Klasse `DataObjectReference` existieren. In den Modell-Dateien werden alle modellierten Datenobjekte als `tDataObjectReference` gespeichert. Die Objekte vom `tDataObject` existieren zwar in der Datei, werden aber

```
<xsd:element name="dataProtectionDataObject" type="
    tDataProtectionDataObject" substitutionGroup="flowElement"/>
<xsd:complexType name="tdataProtectionDataObject">
    <xsd:complexContent>
        <xsd:extension base="tDataObjectReference">
            <xsd:attribute name="dataProtectionLevel" type="
                tDataProtectionLevel"/>
        </xsd:extension>
    </xsd:complexContent>
</xsd:complexType>
```

Listing 14.7: XML-Repräsentation des dataProtectionDataObject

```
<xsd:element name="dataProtectionDataInput" type="
    tDataProtectionDataInput"/>
<xsd:complexType name="tDataProtectionDataInput">
    <xsd:complexContent>
        <xsd:extension base="tDataInput">
            <xsd:attribute name="dataProtectionLevel" type="
                tDataProtectionLevel"/>
        </xsd:extension>
    </xsd:complexContent>
</xsd:complexType>
```

Listing 14.8: XML-Repräsentation des DataProtectionDataInput

```
<xsd:element name="dataProtectionDataOutput" type="
    tDataProtectionDataOutput" />
<xsd:complexType name="tDataProtectionDataOutput">
    <xsd:complexContent>
        <xsd:extension base="tDataOutput">
            <xsd:attribute name="dataProtectionLevel" type="
                tDataProtectionLevel"/>
        </xsd:extension>
    </xsd:complexContent>
</xsd:complexType>
```

Listing 14.9: XML-Repräsentation des dataProtectionDataOutput

im Modell nicht dargestellt. Daher sind sie für die Färbung unerheblich und die Klasse muss nicht erweitert werden.

Analog zum *„dataProtectionDataObject“* werden auch Klassen für einen *„dataProtectionDataInput“* (Listing 14.8), einen *„dataProtectionDataOutput“* (Listing 14.9) und den *„dataProtectionDataStore“* (Listing 14.10) von den entsprechenden Standardklassen abgeleitet.

```
<xsd:element name="dataProtectionDataStore" type="
    tDataProtectionDataStore" substitutionGroup="rootElement"/>
<xsd:complexType name="tDataProtectionDataStore">
    <xsd:complexContent>
        <xsd:extension base="tDataStore">
            <xsd:attribute name="dataProtectionLevel" type="
                tDataProtectionLevel"/>
        </xsd:extension>
    </xsd:complexContent>
</xsd:complexType>
```

Listing 14.10: XML-Repräsentation des dataProtectionDataStore

3. Beteiligte

Letztlich wird noch eine Erweiterung für die Abbildung der beteiligten Rechtssubjekte eingeführt. Hierfür wird die Klasse `tLane` erweitert, die eine Swimlane darstellt. Pools werden in der XML-Repräsentation als Menge von Lanes definiert und müssen daher nicht separat erweitert werden. Listing 14.11 zeigt die Erweiterung um den selbst definierten Datentyp `subject` zu einer `dataProtectionLane`.

```
<xsd:element name="dataProtectionLane" type="tDataProtectionLane"
    />
<xsd:complexType name="tDataProtectionLane">
    <xsd:complexContent>
        <xsd:extension base="tLane">
            <xsd:attribute name="subject" type="tSubject"/>
        </xsd:extension>
    </xsd:complexContent>
</xsd:complexType>
```

Listing 14.11: XML-Repräsentation einer Datenschutzswimlane

14.5. Zusammenfassung

Die prototypische Implementierung baut auf die zuvor beschriebenen Anforderungen auf. Es wird nach der Diskussion verschiedener Alternativen der Camunda Modeler als sehr verbreitetes, bestehendes BPMN-Modellierungssystem als zu erweiternde Basis verwendet. Damit wird die Akzeptanz des Prototyps erhöht. Die Erweiterung des Camunda Modelers erfolgt in Form eines Plugins. Der zentrale Aspekt des Plugins ist die automatische Kategorisierung und Einfärbung der Modellierungselemente. In der prototypischen Realisierung wird besonders darauf geachtet, dass die Erweiterung gegenüber dem BPMN-Standard konform ist.

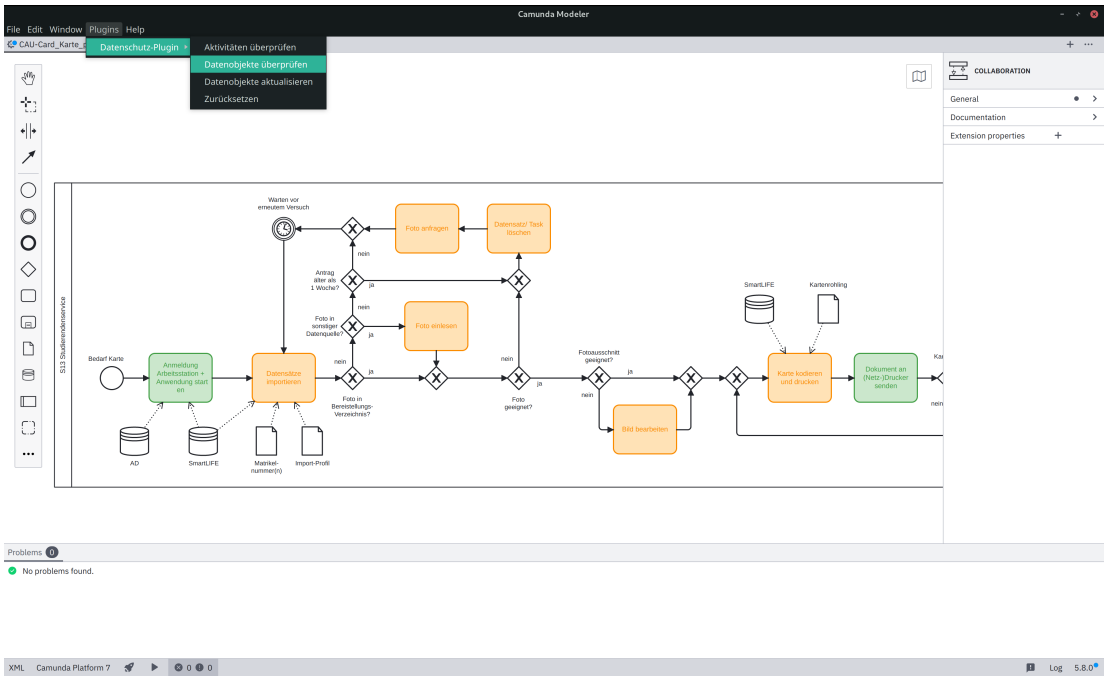


Abbildung 14.11.: Bildschirmfoto des Camunda Desktop Modelers mit integriertem Plugin

Der Camunda Desktop Modeler mit dem Plugin wird in Abbildung 14.11 dargestellt. Im angezeigten Prozessmodell wurden die Aufgaben bereits eingefärbt. Aktuell ausgewählt ist die Funktion zum Kategorisieren der Datenobjekte und -speicher.

15. Evaluation des Prototyps

In diesem Kapitel erfolgt eine grundlegende Evaluation des in Kapitel 14 beschriebenen Prototyps auf Basis der Anforderungen aus Kapitel 13. Hierfür werden gemäß der Gliederung der Anforderungsspezifikation zunächst die funktionalen und anschließend die nichtfunktionalen Anforderungen betrachtet.

15.1. Funktionale Anforderungen

Die funktionalen Anforderungen gliedern sich in fünf Anwendungsfälle, zu welchen jeweils überprüft werden muss, ob diese entsprechend der Spezifikation umgesetzt sind.

15.1.1. Use Case 1: Modellieren

Der erste Anwendungsfall betrifft die Modellierung der Prozesse. Da für den Prototyp ein recht verbreitetes Modellierungswerkzeug als Basis verwendet wird, kann diese Anforderung als erfüllt betrachtet werden. Der verwendete Camunda Modeler wird auch regelmäßig durch die BPMN MIWG (siehe Abschnitt 14.1.1) auf seine Implementierung des BPMN-Standards getestet und erreicht hier sehr gute Ergebnisse. Entsprechend ist von externer Seite belegt, dass BPMN-Modelle dem Standard entsprechend modelliert werden können.

Eigene Tests haben allerdings auch einige kleinere Schwächen ergeben. Einerseits ist die Modellierungsumgebung abhängig von der verwendeten Camunda Plattform. Bei Nutzung der neuen Plattform bestehen einige Einschränkungen, die mit einem großen Fokus auf die Automatisierung der Prozesse begründet werden kann (siehe auch Abschnitt 14.1.2). Das Problem kann aber einfach durch die Wahl der alten Plattform behoben werden. Außerdem sind auch Änderungen geplant, die eine flexiblere Nutzung des Modelers auch mit der neuen Plattform ermöglichen sollen. Außerdem haben die Tests ergeben, dass einige speziellere Modellelemente mit dem Camunda Modeler nicht in die Prozessmodelle eingefügt werden können. Das betrifft z.B. Dateninputs und -outputs. Auf diese Elemente kann aber durchaus auch verzichtet werden, sodass die Einschränkung hier sehr gering ausfällt.

Insgesamt kann der Anwendungsfall definitiv als hinreichend gut umgesetzt betrachtet werden.

15.1.2. Use Case 2: Analysieren/Einfärben

Der zweite Anwendungsfall bildet die Kernfunktionalität des entwickelten Plugins ab. Generell wird dieser auch abgedeckt. Die automatische Kategorisierung hat allerdings einige Schwächen.

Es wurden zwei verschiedene Ansätze umgesetzt. Der Wörterbuch-Ansatz, der für die Datenobjekte verwendet wird, liefert zwar überwiegend korrekte Ergebnisse, erfordert aber einen gewissen Initialisierungsaufwand. Man kann daher kritisieren, dass bei der Erstbenutzung des Prototyps der Anwendungsfall noch nicht umgesetzt wird. Dieses Problem ist im Rahmen eines Prototyps aber zu vernachlässigen. Sollte der Ansatz in einem späteren Produktivsystem zum Einsatz kommen, sollte vorab eine Datenbank für häufig vorkommende Bezeichnung angelegt werden, um den Aufwand für den Anwender zu reduzieren.

Eine andere mögliche Optimierung kann die Änderung des verwendeten Schwellenwerts der Levenshtein-Distanz sein. Ein geringerer Schwellenwert erhöht hier im Allgemeinen die Anzahl der kategorisierten Elemente. In [Ve23] wird aber ein Test mit verschiedenen Schwellenwerten (50%, 75% und 90%) beschrieben, der ergibt, dass die Ergebnisse bei allen drei Varianten ohnehin relativ ähnlich sind. Insbesondere die Ergebnisse für die Schwellenwerte 75% und 90% gleichen sich mit zunehmendem Umfang des Wörterbuchs aneinander an. Für eine Weiterentwicklung des Prototyps bieten sich weitere Tests an, um den bestmöglichen Schwellenwert zu ermitteln.

Der ML-Ansatz, welcher für die Aufgaben verwendet wird, führt insgesamt zu eher schlechten Ergebnissen. Vergleichsweise viele Elemente werden falsch kategorisiert. Dabei ist die Implementierung an sich aber nicht das Problem, sondern die viel zu geringe Menge an Testdaten. Auch hier ist das Ergebnis für den Prototyp als solches ausreichend. Die geforderte Funktionalität ist durchaus vorhanden. Für einen Produktiveinsatz müsste das ML-Modell mit einem deutlich größeren Datensatz angelernt werden.

Außerdem können bislang nur Aufgaben und Datenobjekte, sowie -speicher eingefärbt werden. Hier muss noch eine Erweiterung um Ereignisse erfolgen. Die Implementierung an sich ist hier kein großer Aufwand. Allerdings sind Ereignisse grundsätzlich eher mit Aufgaben als mit Datenobjekten vergleichbar, weshalb sich hier auch der ML-Ansatz anbietet. Im verwendeten Trainingsdatensatz waren hierzu aber kaum Daten vorhanden, weshalb die Erstellung eines eigenen Modells nicht sinnvoll war.

15.1.3. Use Case 3: Färbung korrigieren

Die manuelle Korrektur der Färbung wird wieder direkt vom verwendeten Camunda Modeler unterstützt. Alle relevanten Modellelemente können durch den Anwender

problemlos in den betrachteten Farben Rot, Gelb[1] und Grün eingefärbt werden und die Färbung kann auch jederzeit verändert werden.

In einer späteren Entwicklungsiteration könnten im entsprechenden Menüpunkt noch die jeweiligen Bezeichnungen der Farben so angepasst werden, dass daraus direkt die Semantik klar wird. So könnte Rot etwa mit „kritisch“ bezeichnet werden. Ob dies wirklich sinnvoll ist, muss aber in ausführlichen Anwendertests geprüft werden.

15.1.4. Use Case 4: Prozess betrachten/analysieren

Der Anwendungsfall „Prozess betrachten/analysieren“ stellt technisch keine besonders großen Anforderungen. In der Modellierungsansicht kann der Prozess selbstverständlich auch betrachtet und (manuell) analysiert werden. Das implementierte Plugin kann die Analyse und spätere Optimierung des Prozesses aber unterstützen, wie in Kapitel 10 dargelegt wird.

Für eine spätere Version des Systems kann diskutiert werden, ob eine schreibgeschützte Ansicht eine sinnvolle Ergänzung darstellt, damit alle Änderungen von der korrekten Stelle vorgenommen werden und es nicht versehentlich zu fehlerhaften Änderungen kommt. Dies kann alternativ aber auch durch ein entsprechendes Speichermanagement im Unternehmen sichergestellt werden.

15.1.5. Use Case 5: Prozess bearbeiten

Das Bearbeiten des Prozesses ist technisch vergleichbar mit dem initialen Modellieren. Auch hier sind die Funktionen bereits durch die Basisversion des Camunda Modelers gegeben. Auch hier gelten aber natürlich die Einschränkungen aus Abschnitt 15.1.1.

15.1.6. Zusammenfassung

Insgesamt können die funktionalen Anforderungen als erfüllt betrachtet werden. Viele Aspekte werden hier schon durch eine sinnvolle Wahl des Basissystems erreicht. Aber auch der entwickelte Prototyp kann alle zusätzlich benötigten Funktionen abbilden

[1] Offiziell wird die Farbe im Modeler als „Orange“ bezeichnet und geht auch eher in diese Richtung. Auch bei Verkehrsampeln wird ist die mittlere Farbe, die im Allgemeinen als „Gelb“ bezeichnet wird, aber eher ein Orange. Und auch im TLP wird die entsprechende Farbe als „Amber“, also Bernstein, bezeichnet. Daher kann diese Färbung als hinreichend gut verständlich betrachtet werden.

15.2. Nichtfunktionale Anforderungen

Abbildung 13.7 stellt noch einmal die Priorisierung der nichtfunktionalen Anforderungen dar, wie sie in Abschnitt 13.5.2 zusammengefasst wurden.

Im Folgenden wird zu jedem aufgeführten Bereich kurz dargelegt, inwiefern die Anforderungen vom erstellten Prototyp erfüllt werden.

Abbildung 15.1.: Priorisierung der Qualitätsmerkmale nach ISO/IEC 25010:2011

Zur Erinnerung wird hier erneut die in der Abbildung verwendete Farbnotation definiert:

- Rot entspricht einer hohen Priorität;
- Gelb einer mittleren Priorität;
- Grün einer niedrigen Priorität.

15.2.1. Geeignete Funktionalität

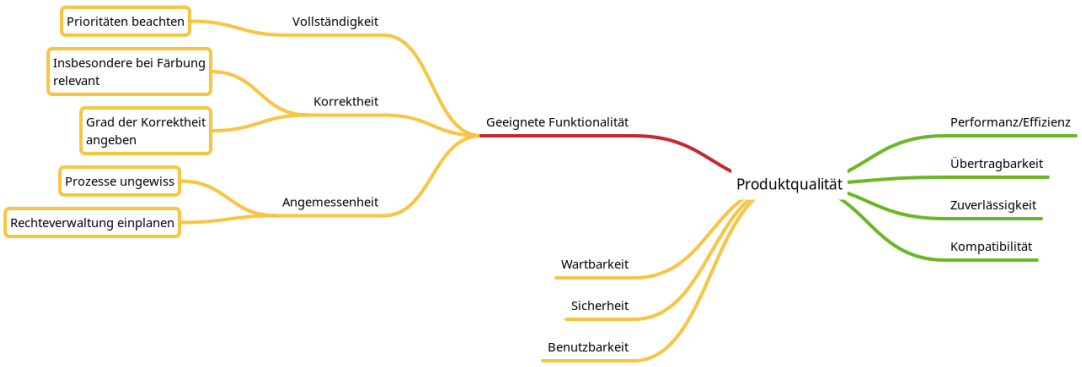


Abbildung 15.2.: Priorisierung der Qualitätsmerkmale im Bereich *„Geeignete Funktionalität“*

Hauptziel des Prototyps war die Demonstration der Funktionalität, weshalb dieser Aspekt als einziger eine hohe Priorität hat. Das Ziel wurde erreicht, wie die Evaluation der funktionalen Anforderungen in Abschnitt 15.1 zeigt.

In Bezug auf die Vollständigkeit wird hier Anforderung von 75% sogar übertroffen. Natürlich gibt es aber noch einige mögliche Erweiterungen. Von Korrektheit kann nicht gesprochen werden, da die automatische Kategorisierung relativ häufig falsche Ergebnisse produziert. Das wurde aber schon in der Anforderungsanalyse bedacht. Das Ziel, den Grad der Korrektheit anzugeben, konnte nicht erreicht werden, da insbesondere für den Wörterbuch-Ansatz dieser über die Zeit betrachtet sehr variabel ist. Daher ist die Information wenig aussagekräftig, weshalb auf eine Angabe verzichtet wird. Die Angemessenheit wurde bei der Realisierung bislang nicht betrachtet, da die konkreten Anforderung für den Prototyp hier auch noch nicht gegeben ist. Einen ersten Ansatz in diese Richtung stellt der in Abschnitt 15.1.4 vorgeschlagene Schreibschutz dar. Dieser könnte mit einem Rechtesystem kombiniert werden, um die Angemessenheit sicherzustellen.

15.2.2. Wartbarkeit

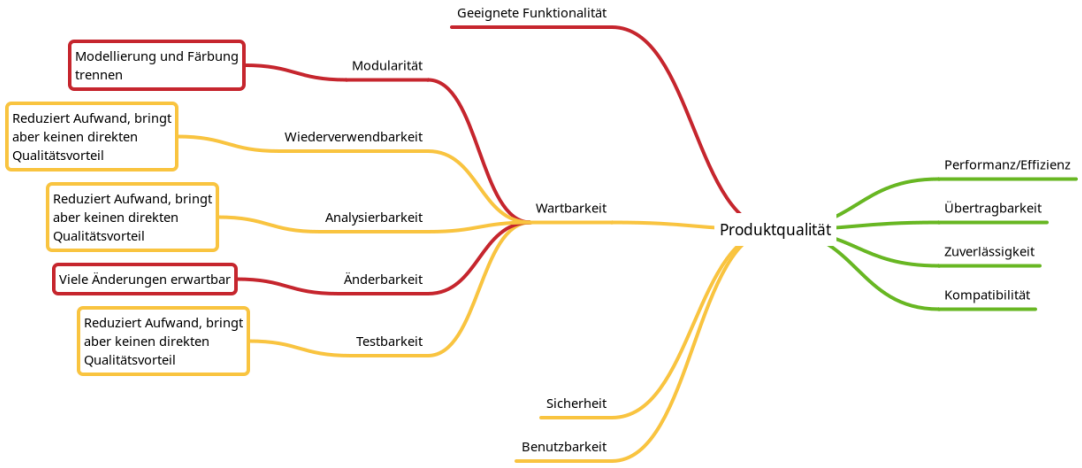


Abbildung 15.3.: Priorisierung der Qualitätsmerkmale im Bereich *„Wartbarkeit“*

Die Wartbarkeit des Systems ist noch ausbaufähig, grundsätzlich aber gegeben.

Die Modularität ist bereits wegen des Plugin-Ansatzes gegeben. So kann das Basissystem völlig unabhängig vom erstellten Plugin verändert oder weiter erweitert werden. Auch der Aufbau des Plugins ist durch die vorgegebene Struktur bereits modular. Die Kategorisierung der verschiedenen Modellelemente könnte optional weiter getrennt

werden. Bei einer späteren Kategorisierung aller Modellelemente mit dem gleichen Ansatz, kann dies aber auch zu Problemen führen.

Auch für die Wiederverwendbarkeit ist der Plugin-Ansatz als positiv zu bewerten. Darüber hinaus wurde das ML-Modell komplett unabhängig implementiert und ist nur durch eine Schnittstelle mit dem Plugin verbunden. Dieser Teil kann also problemlos auch von völlig anderen Systemen verwendet werden.

Die Analysierbarkeit ist noch ausbaufähig. Das liegt schon daran, dass nicht alle Aspekte des Camunda Modelers gut dokumentiert sind. Viele vorhandene Beispiele erleichtern aber den Einstieg und ermöglichen eine vergleichsweise schnelle Einarbeitung. Auch der eigene Code kann noch besser dokumentiert werden. Es liegt aber zumindest eine ausführliche Dokumentation des Plugins in [Ve23] und des ML-Modells in [Pe22] vor.

Änderbarkeit ist vor allem auch durch das Plugin-Prinzip gegeben. Vorteilhaft ist hier auch, dass bei Änderungen kein hoher Aufwand nötig ist, um das veränderte Plugin in den Camunda Modeler einzufügen.

Die Testbarkeit ist noch nicht optimal. Auf Grund vieler zu prüfender Bestandteile, ist eine Umgebung für umfassende Tests nur mit hohem Aufwand zu erreichen. Für weitere Iterationen der Entwicklung sollte das ein Ziel sein.

15.2.3. Sicherheit

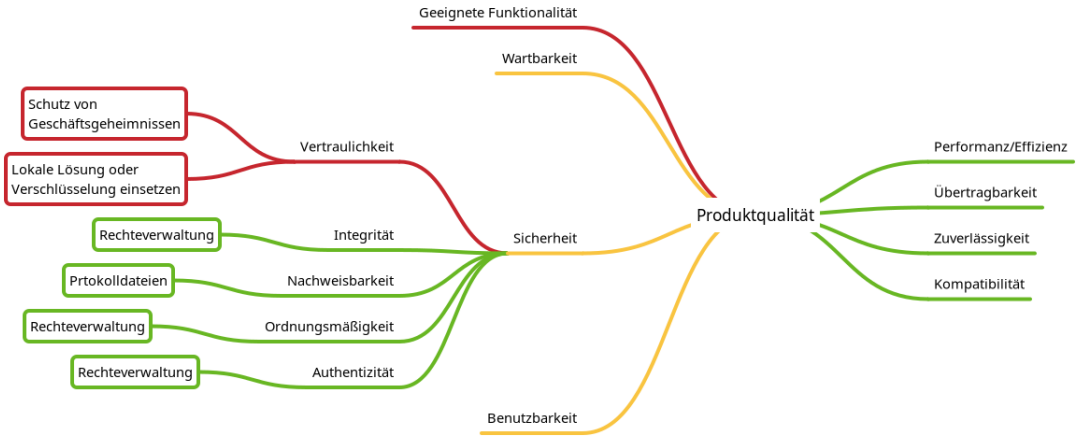


Abbildung 15.4.: Priorisierung der Qualitätsmerkmale im Bereich *„Sicherheit“*

Die Anforderungen an die Sicherheit sind für den Prototyp eher gering und können daher erfüllt werden.

Für die Vertraulichkeit ist es vorteilhaft, dass hier der Hauptteil der Anwendung lokal betrieben wird. Hier ist der Anwender selbst für die Absicherung seiner Infrastruktur

zuständig. Einzig die ML-Komponente ist auf einen externen Server ausgelagert. Das wurde aber lediglich aus Testgründen so designt. Es ist ebenso ein lokaler Betrieb möglich. Außerdem wird hier nur eine Liste mit Aufgaben-Bezeichnungen übergeben. Die davon ausgehende Gefahr, die Vertraulichkeit zu verletzen ist als eher gering einzuschätzen.

Da im Camunda Modeler kein Rechtesystem enthalten ist, muss hier für die Sicherstellung der Integrität auf schon vorhandene Systeme im Unternehmen zurückgegriffen werden. Wegen der geringen Priorität ist das aber auch unkritisch. Selbiges gilt auch für die Ordnungsmäßigkeit und die Authentizität.

15.2.4. Benutzbarkeit

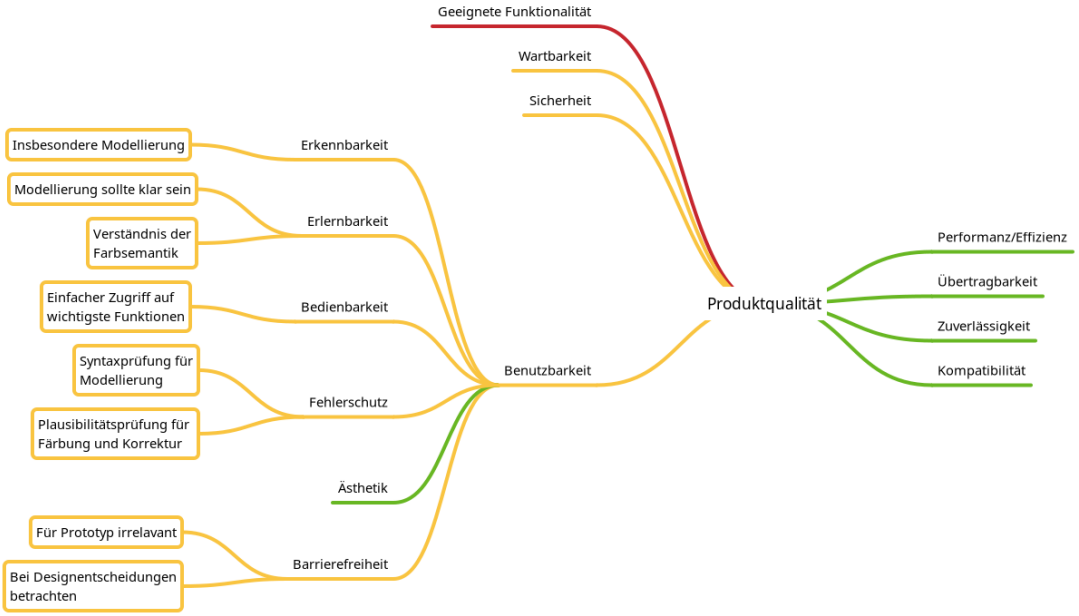


Abbildung 15.5.: Priorisierung der Qualitätsmerkmale im Bereich *„Benutzbarkeit“*

Erste Tests haben eine sehr gute Benutzbarkeit des Prototyps ergeben.

Eine gute Erkennbarkeit ist auf verschiedenen Ebenen gegeben: Einerseits wird der verbreitete Standard BPMN verwendet. Andererseits wird auf ein relativ verbreitetes Modellierungswerkzeug zurückgegriffen. Der Aufbau des Systems entspricht auch gängigen Konventionen, sodass auch bei der Erstbenutzung ein gewisser Wiedererkennungswert vorhanden ist. Abgesehen davon, kann auch der Farbkodierung als wichtiger Aspekt des Prototyps eine gute Erkennbarkeit zugeschrieben werden, wie in Kapitel 10 erläutert wird.

Die gute Erkennbarkeit bedingt auch eine einfache Erlernbarkeit. Auch die übersichtliche Menüführung trägt hierzu bei. Es sollte noch eine umfassende Dokumentation erstellt werden um den Einstieg weiter zu erleichtern. Auch die Umsetzung der in Kapitel 11 vorgeschlagenen Erweiterungen kann hier eine Unterstützung bieten.

Die Bedienbarkeit kann im Camunda Modeler als sehr gut bewertet werden. Bei der Entwicklung des Plugins wurde das Design weiter eingehalten um die gute Bedienbarkeit beizubehalten. Eine Verbesserung könnte die Erstellung von Buttons direkt auf der Zeichenfläche an Stelle der Menüeinträge darstellen.

Für die Modellierungskomponente ist ein Fehlerschutz in Form einer Syntaxüberprüfung gegeben. Eine Plausibilitätsprüfung für die Färbung sollte zu Testzwecken noch implementiert werden.

Auch wenn die Ästhetik ohne umfangreiche Anwendertests schlecht zu beurteilen ist, kann wegen der Verwendung eines etablierten Systems von einer hinreichend guten Ästhetik ausgegangen werden.

Barrierefreiheit ist aktuell noch nicht gegeben, da bislang nur eine reine Einfärbung der Modellelemente implementiert wurde und diese für farbfehlsichtige Menschen schlecht zu erkennen ist. Die Realisierung der Erweiterungen aus Kapitel 11 kann aber auch hier unterstützen.

15.2.5. Performanz und Effizienz

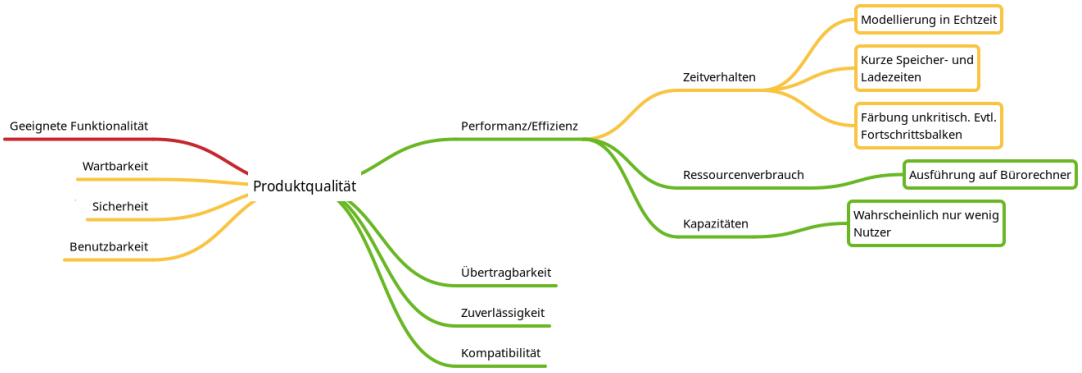


Abbildung 15.6.: Priorisierung der Qualitätsmerkmale im Bereich *„Performanz und Effizienz“*

Das Zeitverhalten des Prototyps kann insgesamt als gut betrachtet werden. Die Modellierung erfolgt nahezu in Echtzeit. Die automatisch Kategorisierung über das Wörterbuch läuft ebenfalls ohne erkennbare Verzögerung ab. Die Kategorisierung über das ML benötigt etwas mehr Zeit. Aber auch hier kann ein durchschnittlich großes Modell in weniger als zehn Sekunden kategorisiert werden. Die Wartezeit für den

Nutzer bis zur Beendigung des Prozesses kann daher als vertretbar betrachtet werden. Sinnvoll ist hier die Erweiterung um eine Rückmeldung, dass die Kategorisierung gerade läuft. Aktuell kann nach der Auswahl des entsprechenden Menüeintrags der Eindruck entstehen, es passiere nichts.

Der Ressourcenverbrauch wurde nicht gezielt getestet. Bei der Ausführung auf verschiedenen Rechnern mit unterschiedlicher Rechenleistung kam es aber zu keinen Problemen.

Der aktuelle Prototyp ist auf die Einzelnutzung ausgelegt. Grundsätzlich kann das System in verschiedenen Instanzen auf beliebig vielen Rechnern zeitgleich verwendet werden. Über einen gemeinsamen Speicher kann dann auch auf die gleichen Prozesse zugegriffen werden. Der Prototyp selbst implementiert aber keine Sicherheitsmechanismen, die etwa vor einer gleichzeitigen Bearbeitung der gleichen Datei schützen.

15.2.6. Übertragbarkeit

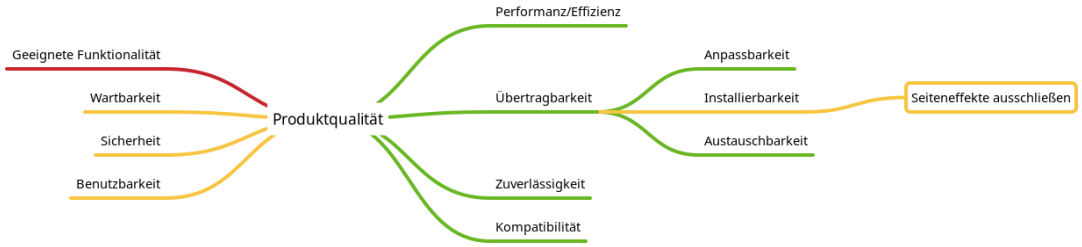


Abbildung 15.7.: Priorisierung der Qualitätsmerkmale im Bereich *„Übertragbarkeit“*

Im Bereich Übertragbarkeit sollte insbesondere die Installierbarkeit betrachtet werden. Diese ist beim Prototyp aber unkritisch, da die Software nach dem Download ohne Installation ausführbar ist.

Auch die Anpassbarkeit des Systemumfelds und die Austauschbarkeit ist gegeben.

15.2.7. Zuverlässigkeit

Die Zuverlässigkeit wird generell als eher nebensächlich beurteilt. Letztlich kann diese zum aktuellen Zeitpunkt auch noch nicht sinnvoll bewertet werden.

Zumindest die Fehlertoleranz als auch die Wiederherstellbarkeit können aber positiv bewertet werden. Hier wird davon ausgegangen, dass keine nennenswerten Probleme auftreten.

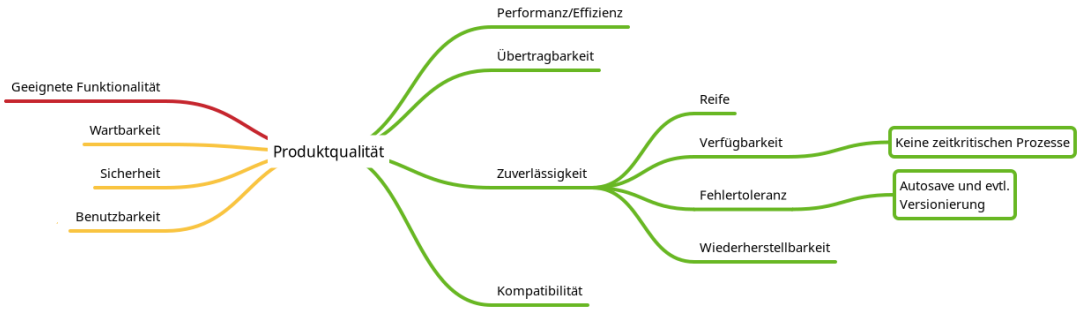


Abbildung 15.8.: Priorisierung der Qualitätsmerkmale im Bereich *„Zuverlässigkeit“*

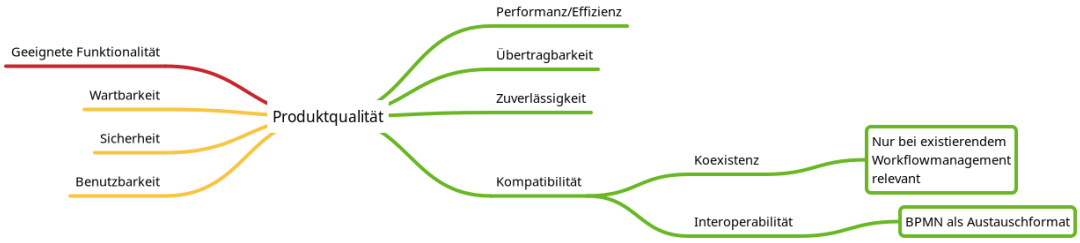


Abbildung 15.9.: Priorisierung der Qualitätsmerkmale im Bereich *„Kompatibilität“*

15.2.8. Kompatibilität

Auch die Kompatibilität ist eher unkritisch. Bezüglich der Koexistenz sollte es keine Einschränkungen geben. Zur Interoperabilität bestehen keine besonderen Anforderungen. Positiv hervorgehoben werden kann aber, dass Camunda auch eine Workflowengine bereitstellt und somit eine relativ einfache Integration einer Prozessautomatisierung möglich ist.

15.3. Zusammenfassung

Insgesamt erfüllt der Prototyp die wichtigsten funktionalen und nichtfunktionalen Anforderungen, die in Kapitel 13 definiert wurden. Das Ziel des Prototyps, ein Instrument für weitergehende Evaluationen des eigentlichen Konzepts zu erschaffen, ist erreicht worden. Es können allerdings noch Optimierungen und Erweiterungen umgesetzt werden. Insbesondere betrifft das die automatische Kategorisierung der Modellelemente, die zwar funktioniert, aber noch Potential für Verbesserungen bietet. Außerdem ist die Integration der Erweiterungen des Ansatzes, die in Kapitel 11 erläutert werden, eine sinnvolle Ergänzung.

Teil IV.

Fazit

16. Zusammenfassung und Diskussion

Datenschutz ist ein Thema mit großer Relevanz. Dies gilt sowohl für den Bürger, dessen Schutz das Ziel ist, als auch für Unternehmen, Behörden und andere Organisationen, die sich an die entsprechende Regeln halten müssen, um keine Strafen zu riskieren. Diese Arbeit zeigt einen Weg auf, das Bewusstsein für den Datenschutz zu erhöhen und die vorliegenden Regeln einzuhalten, um alle Beteiligten vor negativen Folgen zu schützen.

Hierfür werden nach der Motivation des Themas und einer Einführung der Methodik in Kapitel 1 zunächst in Teil I einige wichtige Grundlagen erläutert. Anschließend wird in Teil II ein Konzept dargestellt, welches auf der farblichen Kennzeichnung von Datenschutzaspekten in Geschäftsprozessmodellen basiert. In Kapitel 6 wird hierfür zunächst der aktuelle Forschungsstand betrachtet und Konzepte relevanter Arbeiten aus verschiedenen Bereichen verglichen. Zur Veranschaulichung werden in Kapitel 7 drei exemplarische Prozesse eingeführt, die aus Bereichen stammen, in denen der Datenschutz eine besonders wichtige Rolle spielt. Diese sind das Personalwesen, das Gesundheitswesen und die öffentliche Verwaltung. In Kapitel 8 wird dann das eigentliche Konzept erläutert. Dieses bewertet und kategorisiert die einzelnen Elemente eines Geschäftsprozessmodells bezüglich ihrer Kritikalität für den Datenschutz anhand verschiedener Kriterien. Die Bewertung wird mit den Farben direkt im Prozessmodell visualisiert. Das Konzept wird in Kapitel 9 anschließend auf die zuvor eingeführten Beispiele angewandt. Kapitel 10 beschreibt das Vorgehen und die Ergebnisse einer zweistufigen Evaluation des Konzepts. Hier wird zunächst eine Expertenbefragung und anschließend ein Vergleichsexperiment betrachtet. Beide Studien stützen den Nutzen des Konzepts. In Kapitel 11 werden noch einige mögliche Erweiterungen des Konzepts vorgeschlagen, bevor in Kapitel 12 die technische Betrachtung beginnt. Hier werden zunächst theoretisch verschiedene Ansätze der automatischen Analyse und Kategorisierung der Prozesselemente diskutiert. In Teil III folgt dann die Umsetzung des Konzepts in einem ersten Prototyp. Hierfür werden einleitend in Kapitel 13 die funktionalen und nichtfunktionalen Anforderungen an ein solches System definiert. Anschließend wird die eigentliche Realisierung in Form eines Plugins für den Camunda Modeler beschrieben. Ein besonders wichtiger Aspekt ist auch hier die automatische Kategorisierung. Abschließend wird der Prototyp in Kapitel 15 mit Blick auf die Spezifikation evaluiert.

16.1. Beiträge der Arbeit

Die Arbeit verfolgt die Design Science Methode. Ziel dieser Methode ist die Erstellung von Artefakten, z.B. Methoden und Modellen. Im Rahmen dieser Dissertation wurden drei Artefakte erstellt, welche auch als Beiträge der Arbeit zur Wissenschaft zu betrachten sind.

Den wichtigsten Beitrag stellt hier das entwickelte Konzept dar, also das Modell zur Repräsentation von Datenschutzaspekten in Geschäftsprozessmodellen.

Außerdem wurde ein Prototyp entwickelt, der für die weitere Forschung verwendet werden kann. Letztlich liefert die Arbeit aber auch den Beitrag einer Evaluationsmethodik. Das zweistufige Verfahren mit einer qualitativen und einer quantitativen Komponente hat sich hierbei bewährt. In Kapitel 10 werden aufgetretene Probleme bei der Evaluation und entsprechende Lösungsansätze diskutiert, sodass die Methodik künftig auch für andere Ansätze angewendet werden kann.

16.2. Einsatzmöglichkeiten

Das vorgestellte Konzept kann in Verbindung mit dem Prototyp zu verschiedenen Zwecken eingesetzt werden. Allen voran steht hier eine Unterstützung bei der allgemeinen Risikoeinschätzung für Datenschutzaspekte in Geschäftsprozessen. Diese wiederum kann einerseits für die Verringerung der Risiken, z.B. durch Prozessoptimierungen eingesetzt werden. Dieser Aspekt kann dem in Art. 25 DSGVO definierten Prinzip *„Datenschutz durch Technikgestaltung“* zugeordnet werden.

Andererseits kann die Risikoabschätzung aber auch für die Erfüllung der Dokumentationspflichten hilfreich sein, nämlich zur Abwägung, ob eine Datenschutzfolgeabschätzung notwendig ist. Die Erfüllung der Dokumentationspflichten kann aber auch noch anders unterstützt werden. So sind die aufbereiteten Prozesse etwa eine gute Grundlage für die Erstellung eines Verzeichnisses von Verarbeitungstätigkeiten.

Insgesamt kann das Konzept zu einem höheren Datenschutzbewusstsein aller beteiligten Personen führen, welches nicht nur die Compliance eines Unternehmens oder einer sonstigen Organisation erhöht, sondern vor allem die Rechte betroffener Personen schützt.

16.3. Limitationen

Auch wenn das vorgestellte Konzept einen großen Nutzen hat und in einer ersten Evaluation überzeugen konnte, gibt es einige Limitationen.

Zwei Aspekte betreffen hier grundlegende Einschränkungen, die bewusst vorgenommen wurden, um das Problem für diese Arbeit hinreichend übersichtlich zu gestalten. Einerseits ist das die Einschränkung auf den deutschen Rechtsraum. Die Grundlage bei der juristischen Betrachtung bildet zwar die DSGVO, weshalb sich die meisten Aspekte auf andere EU- und EWR-Mitgliedsstaaten übertragen lassen sollten, die nationale Gesetzgebung muss hier aber auch betrachtet werden. In Ländern außerhalb des Geltungsbereichs der DSGVO herrscht teilweise eine völlig andere Rechtslage, weshalb hier wahrscheinlich einige Anpassungen nötig wären, bevor das Konzept international anwendbar wäre.

Der andere Aspekt ist die Modellierungsnotation. Hier wurde bewusst BPMN als sehr verbreitete Notation gewählt. Es existieren aber noch weitere Notationen, die ebenfalls häufig verwendet werden. Grundsätzlich sollte sich das Konzept auf andere Notationen anpassen lassen. Teilweise wird hier aber ohnehin mit Farben gearbeitet (etwa bei eEPK). Hier müsste eine entsprechende Anpassung der Visualisierung des Konzepts geschehen, da die zusätzliche Verwendung der Ampel-Notation hier zu Verwirrung führen kann.

Neben diesen bewussten Entscheidungen ist aber noch ein weiterer Aspekt relevant: Die Barrierefreiheit. Das beschriebene Konzept ist für farbfehlsichtige Menschen schwierig nutzbar. Hier können aber die Erweiterungen aus Kapitel 11 unterstützen. Alternativ (oder zusätzlich) könnten auch einfach Nummern für die einzelnen Kategorien eingefügt werden, die in den Modellelementen auftauchen. Hier geht zwar der Vorteil der grafischen Darstellung verloren, Menschen mit Einschränkung können das Konzept aber immerhin nutzen. Das Problem der Farben besteht im Übrigen unter bestimmten Umständen für alle Nutzer. Ein Beispiel hierfür ist der Ausdruck eines Prozessmodells in Graustufen. Gerade in der öffentlichen Verwaltung wird auch heute noch viel mit Ausdrucken gearbeitet, sodass dieses Szenario nicht unwahrscheinlich ist. Und auch in der Wissenschaft zeigt sich dieses Problem. So wurde etwa der Konferenzband, in welchem die Grundlagen des Konzepts ursprünglich veröffentlicht wurden, auch in Graustufen gedruckt, weshalb sich auch hier ergänzend eine Nummerierung findet[WSG21].

Wie die Ausführungen dieses Abschnitts zeigen, lassen sich alle Limitationen beheben. Weitere Details dazu finden sich im folgenden Kapitel 17.

17. Ausblick

Die zukünftigen Arbeiten am vorgestellten Ansatz lassen sich in drei Bereiche gliedern. Einerseits sollte der bestehende Prototyp und auch das Konzept an sich weiter optimiert werden. Andererseits sind noch verschiedene Erweiterungen denkbar. Darüber hinaus sollte eine zusätzliche Evaluation erfolgen.

17.1. Optimierung

Die Optimierung des Prototyps betrifft in erster Linie die automatische Kategorisierung. Das Wörterbuch-Verfahren bedeutet einen hohen initialen Aufwand. Hier wäre ein umfassendes vorkonfiguriertes Wörterbuch ein Lösungsansatz. Aber auch der ML-Ansatz kann deutlich optimiert werden. Hierfür sind insbesondere deutlich mehr Trainingsdaten nötig. Eine weitere Unterstützung kann hier die Betrachtung verschiedener Muster bieten, wie sie auch in den in Kapitel 6 beschriebenen Design Pattern verwendet werden. Alternativ kann die Kategorisierung mit dem in Abschnitt 14.3.3 kurz eingeführten Ontologie-Ansatz angegangen werden.

Abgesehen von der automatischen Kategorisierung kann auch die grundlegende Systematik der Kategorisierung weiter betrachtet werden. Weitere Ansätze der Kategorie-Definitionen sollten eruiert und getestet werden.

17.2. Erweiterungen

Das beschriebene Konzept und der Prototyp könnten in verschiedene Richtungen erweitert werden.

17.2.1. Andere Modellierungsnotationen

So wäre beispielsweise eine Anwendung auf andere Modellierungsnotationen, wie z.B. eEPK denkbar. Hierbei sind allerdings einigen Fällen konzeptionelle Überlegungen notwendig. So werden in eEPK etwa ohnehin alle Notationselemente eingefärbt. Hier stellt sich einerseits die Frage ob von den Prozesselementen für die Anwendung des Färbungsansatzes zunächst ihre üblichen Farbe entfernt werden sollte, was allerdings

eventuell die Aussagekraft des Prozessmodells verringert. Andererseits könnte die übliche Farbe zusätzlich zur Ampel-Färbung beibehalten werden. Das könnte allerdings insgesamt sehr unübersichtlich und verwirrend wirken. Neben diesen konzeptionellen Überlegungen stellt auch die Implementierung eine Herausforderung dar. Hier könnte entweder auf eine andere Grundarchitektur zurückgegriffen werden, da Camunda bislang nur die BPMN-Modellierung unterstützt. Oder alternativ könnte Camunda um die entsprechend andere Notation erweitert werden. Zum anderen müsste aber auch der Algorithmus zur automatischen Färbung an die veränderten Gegebenheiten angepasst werden, wenn beispielsweise einige Notationselemente, auf die für die Klassifizierung zurückgegriffen wird, nicht existieren oder anders realisiert sind.

17.2.2. Andere Rechtsräume

Eine andere mögliche Erweiterung wäre die Anpassung an andere Rechtsräume. Im Rahmen dieser Arbeit wurde im Wesentlichen die Situation in Deutschland betrachtet. Da hier ein Großteil der Rechtslage auf der EU-DSGVO basiert, lassen sich die Ergebnisse höchstwahrscheinlich weitestgehend auf das EU-Ausland übertragen. In anderen Ländern muss dies aber nicht der Fall sein. Eventuell sind auch hier recht gravierende Anpassungen notwendig, da die Grundlagen für eine Datenverarbeitung möglicherweise ganz anders geregelt sind und somit eine Unterscheidung zwischen Aktivitäten, die eine Einwilligung erfordern und solchen, die dies nicht tun, unter Umständen nicht sinnvoll ist.

17.2.3. Zusatzinformationen

Generell hat die Färbung der Prozesse sich als sinnvolles Mittel zur Visualisierung von Datenschutzaspekten herausgestellt (siehe Kapitel 10). Allerdings sind auch zusätzliche Informationen wünschenswert, die im besten Fall auch mit graphischen Elementen dargestellt werden können. In Kapitel 11 werden einige Möglichkeiten dargestellt. Diese sollten technisch umgesetzt und anschließend evaluiert werden.

Eine weitere, bislang noch nicht betrachtete Idee ist ein Hinweis auf mögliche Prozessoptimierungen, um den Datenschutz zu verbessern. Beispielsweise könnten Hinweise erfolgen, wenn eine Verarbeitungstätigkeit ausgeführt wird, für die eine Einwilligung notwendig ist, diese zuvor aber nicht eingeholt wurde. Hier könnte zunächst abgefragt werden, ob diese Tätigkeit wirklich notwendig ist. Wenn das der Fall ist, kann mit Hilfe von Design Pattern eine mögliche Umsetzung vorgeschlagen werden.

17.2.4. Technische Erweiterungen des Prototyps

Der implementierte Prototyp wurde als Plugin für den Camunda Desktop Modeler umgesetzt. Wegen der gemeinsamen Basis sollte es relativ einfach möglich sein, das Plugin auf den Camunda Web Modeler und bpmn.io zu portieren (siehe Abschnitt 14.1.2). Das könnte die Nutzung weiter vereinfachen und mehr potentielle Nutzer bzw. Tester ansprechen.

17.2.5. Weitere Erweiterungsmöglichkeiten

Die Evaluation des Konzepts hat gezeigt, dass ein großes Problem der Mangel an für die Forschung zur Verfügung stehender Geschäftsprozessmodelle darstellt Ein Lösungsansatz hierfür wäre die automatische Generierung von Prozessmodellen aus textuellen Prozessbeschreibungen. Hierfür gibt es auch bereits erste Ansätze (siehe z.B. [FMP11]). Denkbar ist eine Integration einer solchen Generierungskomponente in den Prototyp.

17.3. Weitere Evaluation

Das beschriebene Konzept inklusive der Erweiterungen aus Kapitel 11 kann weiterführend evaluiert werden. Der erstellte Prototyp bietet hierfür ein gutes Hilfsmittel, welches hierbei auch seinerseits weitergehend getestet werden kann.

Für eine zukünftige Evaluation bietet es sich an, das zweistufige Verfahren aus Kapitel 10 erneut anzuwenden, wobei die dortigen Ausführungen zu den Schwächen der Methodik beachtet werden sollten.

Bei der Expertenbefragung können einerseits die gleichen Gesprächspartner erneut befragt werden. Der inzwischen bestehende Prototyp und die Erweiterungen werden hier wahrscheinlich zu neuen Erkenntnissen führen. Aber natürlich bietet sich auch die Befragung weiterer Experten mit unterschiedlichem fachlichen Hintergrund an. Auch die Befragung von Experten aus dem innereuropäischem Ausland kann weitere interessante Aspekte aufdecken. Ebenso kann das Vergleichsexperiment mit einer größeren Menge an Teilnehmern aus verschiedenen Bereichen an Aussagekraft gewinnen.

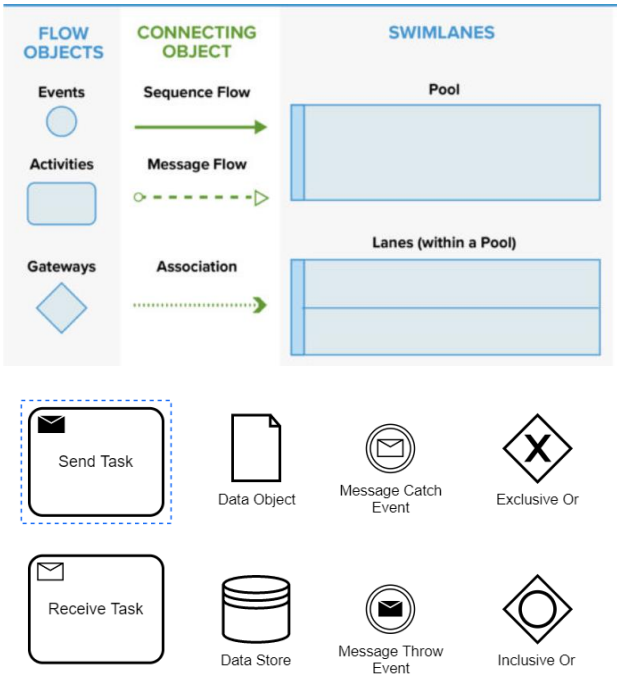
Eine zusätzlicher Aspekt, der untersucht werden kann, ist die fragestellungsabhängige Änderung des Konzepts. In Kapitel 8 werden verschiedene Optionen zur Definition der Farbkategorien definiert. Unter Umständen macht es Sinn, diese Definition je nach konkreter Fragestellung oder Anwender dynamisch anzupassen. Hierfür können verschiedene Ansätze ausgearbeitet und mit Experten diskutiert werden. Wenn hier ein positives Feedback erreicht wird, bietet sich auch ein erneutes Vergleichsexperiment an.

A. Experiment

Exercise: Privacy in Process Models

Exercise: Find as many problems of privacy as you can find in the models in 10 minutes. Everyone gets one model with colors and another model without colors.

Business Process Model Notation (BPMN):



Colors in the Models:

Activities:

- Green: No Privacy Concern
- Yellow: Data processing is allowed without consent
- Red: Consent is required for data processing

Documents:

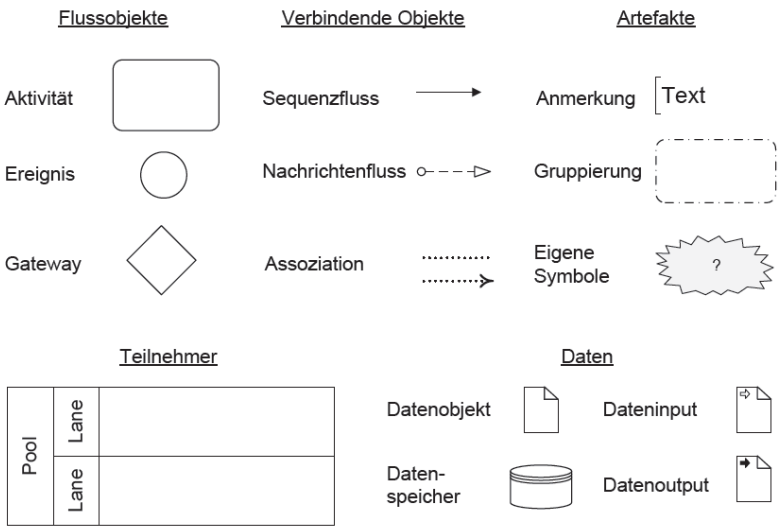
- Green: without personal data
- Yellow: with personal data
- Red: with critical personal data (Art.9 GDPR)

Abbildung A.1.: Aufgabenblatt Englisch

Aufgabe: Datenschutz in Prozessmodellen

Aufgabe: Kennzeichne alle datenschutzrelevanten Elemente im Prozessmodell. Dafür sind 10 Minuten Zeit. Jeder bekommt ein eingefärbtes Modell und eins ohne Farben.

Business Process Model Notation (BPMN):



Farben in den Modellen:

Aktivitäten:

- Grün: Nicht datenschutzrelevant
- Gelb: Datenverarbeitung ist ohne Einwilligung erlaubt
- Rot: Einwilligung ist notwendig zur Datenverarbeitung

Dokumente:

- Grün: ohne personenbezogene Daten
- Gelb: mit personenbezogenen Daten
- Rot: mit kritischen personenbezogenen Daten (Art.9 DSGVO)

Abbildung A.2.: Aufgabenblatt Deutsch

B. Modellierungswerkzeuge

Tabelle B.1.: Auswahl verfügbarer Modellierungswerkzeuge, Teil 1

Werkzeug	**Erweiterbar**	**Lizenz/Kosten**
Camunda Modeler[1]	+	MIT[2]
ARIS[3]	–	proprietär/kostenlose Testversion
ARIS Express[4]	–	proprietär/kostenlos
Cardanit[5]	–	proprietär/kostenlose Testversion
Adonis[6] [7]	–	proprietär/kostenlose Testversion
Trisotech Workflow Modeler[8]	–	proprietär/kostenlose Testversion
MID Innovator[9]	–	proprietär/kostenlose Testversion
Omnitracker BPMN[10]	–	proprietär/kostenlose Testversion
Viadee BPMN Modeler[11]	–	proprietär/kostenlose Version
bpanda[12]	–	proprietär/kostenlose Testversion
KnowProcess[13]	–	proprietär
SAP Signavio[14]	–	proprietär/kostenlose Academic Edition
W4 BPMN+[15]	–	proprietär/kostenlose Testversion
Case Agile Enterprise Explorer[16]	+	proprietär

[1] https://camunda.com/de/download/modeler/
[2] https://github.com/camunda/camunda-modeler/blob/develop/LICENSE
[3] https://www.softwareag.com/de_de/platform/aris/process-design.html
[4] https://www.ariscommunity.com/aris-express
[5] https://www.cardanit.com/
[6] https://www.boc-group.com/de/adonis-14-0-bringt-ihr-gpm-weiter/
[7] https://www.adonis-community.com/
[8] https://www.trisotech.com/digital-modeling-suite/
[9] https://innovator.de/geschaeftsprozessmodellierung/
[10] https://www.omnitracker.com/de/produkte/bpmn/
[11] https://www.viadee.de/en/solutions/business-process-management/bpmn-modeler
[12] https://bpanda.com/
[13] https://knowprocess.com/
[14] https://www.signavio.com/de/prozessmodellierung-mit-bpmn-2-0/
[15] https://ecosystem.itesoft.com/product/w4-bpmn-evaluation
[16] http://enterprise-explorer.com
[17] https://caseagile.com/products/bpmn-view/

Tabelle B.2.: Auswahl verfügbarer Modellierungswerkzeuge, Teil 2

Werkzeug	**Erweiterbar**	**Lizenz/Kosten**
Case Agile BPMN View[17]	+	MIT[18]
Sparx Enterprise Architect[19]	?	proprietär/kostenlose Testversion
Vizi Modeler[20]	–	proprietär/kostenlose Testversion
Modelio[21]	+	GPL3.0[22]
Yaoqiang BPMN Editor [23] [24]	+	GPL3.0
GenMyModel[25]	?	SaaS[26]/kostenlose Version
Activity[27]	+	Apache 2.0[28]
IBM Blueworks Live[29]	–	proprietär/kostenlose Testversion
Oracle BPMN[30]	–	proprietär/?
BIC Process Design[31]	–	proprietär/kostenlose Testversion
Bonita[32]	+	GPLv2
Intellior Aeneis[33]	–	proprietär/kostenlose Testversion
ibo Prometheus[34]	–	proprietär/?
iGrafx[35]	–	proprietär/kostenlose Testversion
Visual Paradigm[36]	–	proprietär/ab 99$
Bizagi Modeler[37]	?	proprietär/kostenlose Version
IBM Process Designer[38]	–	proprietär

[18] https://github.com/bzinchenko/bpmnview/blob/master/LICENSE
[19] https://www.sparxsystems.de/
[20] https://www.itp-commerce.com/de/uebersicht-vizi-bpm-suite/vizi-modeler/
[21] https://github.com/ModelioOpenSource/Modelio
[22] https://github.com/ModelioOpenSource/Modelio/blob/master/LICENSE
[23] https://bpmn.sourceforge.net/
[24] https://sourceforge.net/projects/bpmn/
[25] https://www.genmymodel.com
[26] Software as a Service
[27] https://www.activiti.org/
[28] https://github.com/Activiti/Activiti/blob/develop/LICENSE.txt
[29] https://www.ibm.com/de-de/products/blueworkslive
[30] https://www.oracle.com/middleware/technologies/bpm.html
[31] https://www.gbtec.com/de/software/bic-process-design/
[32] https://www.bonitasoft.com
[33] https://www.intellior.ag/software/
[34] https://www.ibo.de/software/prozessmanagement-ibo-prometheus
[35] https://www.igrafx.com/de/produkte/prozessmodellierung-journey-map/
[36] https://www.visual-paradigm.com
[37] https://www.bizagi.com/de/plattform/modeler
[38] https://www.ibm.com/docs/en/baw/20.x?topic=applications-process-designer

C. Quelltexte

C.1. Menüeinträge

```
1  module.exports = function (electronApp, menuState) {
2    return [{
3      label: 'Aktivitäten überprüfen',
4      accelerator: 'CommandOrControl+',
5      enabled: () => menuState.bpmn && menuState.platform === '
           platform',
6      action: () => electronApp.emit('menu:action', '
           checkActivities')
7    }, {
8      label: 'Datenobjekte überprüfen',
9      accelerator: 'CommandOrControl+',
10     enabled: () => menuState.bpmn && menuState.platform === '
           platform',
11     action: () => electronApp.emit('menu:action', '
           checkDataObjects')
12   },{
13     label: 'Datenobjekte aktualisieren',
14     accelerator: 'CommandOrControl+',
15     enabled: () => menuState.bpmn && menuState.platform === '
           platform',
16     action: () => electronApp.emit('menu:action', '
           updateDataObjects')
17   },{
18     label: 'Zurücksetzen',
19     accelerator: 'CommandOrControl+',
20     enabled: () => menuState.bpmn && menuState.platform === '
           platform',
21     action: () => electronApp.emit('menu:action', 'reset')
22   }
23   ]
24 }
```

Listing C.1: Plugin/menu/menu.js

C.2. Plugin-Funktionalität

```
1  'use strict';
2  let tempSim = 0;
3
4  export default function DatenschutzPlugin(commandStack,
       editorActions, elementRegistry, eventBus) {
5    this.commandStack = commandStack;
6
7
8    editorActions.register({
9      updateDataObjects: function () {
10       const dataObjects = elementRegistry.filter(function (
             element) {
11         if (element.type == "bpmn:DataObjectReference" || element
               .type == "bpmn:DataStoreReference") return element;
12       });
13       console.log(dataObjects);
14       let tempData = [];
15       for (let i = 0; i < dataObjects.length; i++) {
16         let klasse = undefined;
17         let color = dataObjects[i].di.fill;
18         if (color == "#ffcdd2") {
19           klasse = "rot"
20         } else if (color == "#ffe0b2") {
21           klasse = "gelb"
22         } else if (color == "#c8e6c9") {
23           klasse = "grün"
24         }
25         if (klasse != undefined) {
26           tempData = tempData.concat([[dataObjects[i].
                 businessObject.name, klasse]])
27         }
28       }
29       const formattedList = tempData.map(item => {
30         return {
31           Beschreibung: item[0],
32           Klasse: item[1]
33         };
34       });
35       var xhr = new XMLHttpRequest();
36       xhr.withCredentials = true;
37       xhr.addEventListener("readystatechange", function () {
38         if (this.readyState === 4) {
39           console.log(this.responseText);
40         }
```

```
41        });
42        xhr.open("POST", "https://datenobjekte-ff3f.restdb.io/rest/
              datenobjekte");
43        xhr.setRequestHeader("content-type", "application/json");
44        xhr.setRequestHeader("x-apikey", "6419fd0422634c74fb00afe0"
              );
45        xhr.setRequestHeader("cache-control", "no-cache");
46        xhr.send(JSON.stringify(formattedList));
47      }
48    })
49
50    editorActions.register({
51      checkDataObjects: function () {
52        let db = []
53        const dataObjects = elementRegistry.filter(function (
              element) {
54          if (element.type == "bpmn:DataObjectReference" || element
                .type == "bpmn:DataStoreReference") return element;
55        });
56
57        var data = null;
58        var xhr = new XMLHttpRequest();
59        xhr.withCredentials = false;
60        xhr.addEventListener("readystatechange", function () {
61          if (this.readyState === 4) {
62            let jsonRes = JSON.parse(xhr.responseText);
63            for (let i = 0; i < jsonRes.length; i++) {
64              db = db.concat([[jsonRes[i]._id, jsonRes[i].
                    Beschreibung, jsonRes[i].Klasse]])
65            }
66            console.log(db);
67            for (let i = 0; i < dataObjects.length; i++) {
68              let oldSim = 0;
69              let dataObject = dataObjects[i].di.bpmnElement.name;
70              for (let j = 0; j < db.length; j++) {
71                if (stringsMatch(dataObject.toUpperCase(), db[j
                      ][1].toUpperCase())) {
72                  let color = "#FFFFFF";
73                  if (tempSim > oldSim) {
74                    oldSim = 0 + tempSim;
75                    if (db[j][2] == "rot") {
76                      color = "#FF0000";
77                    } else if (db[j][2] == "grün") {
78                      color = "#00FF00";
79                    } else if (db[j][2] == "gelb") {
80                      color = "#FFFF77";
81                    }
```

```
82                      colorElement(dataObjects[i], color,
                            commandStack)
83                  }
84
85                }
86              }
87            }
88          }
89        });
90        xhr.open("GET", "https://datenobjekte-ff3f.restdb.io/rest/
              datenobjekte");
91        xhr.setRequestHeader("content-type", "application/json");
92        xhr.setRequestHeader("x-apikey", "6419fd0422634c74fb00afe0"
              );
93        xhr.setRequestHeader("cache-control", "no-cache");
94        xhr.send(data);
95      }
96    })
97
98    editorActions.register({
99      checkActivities: function () {
100       const allElements = elementRegistry.filter(function (
              element) {
101         return element;
102       });
103       console.log(allElements);
104       const activities = elementRegistry.filter(function (element
              ) {
105         if (element.type == "bpmn:Task") return element;
106       });
107       console.log(activities);
108       let jsonObject = { "model": [] };
109       let activityArr = [];
110       let newJsonString = undefined;
111
112       for (let i = 0; i < activities.length; i++) {
113         let task = activities[i].businessObject;
114         if (task.name) {
115           let tempArr = [{ "language": "BPMN", "id": task.id, "
                  type": "activity", "content": task.name }];
116           activityArr = activityArr.concat(tempArr);
117         }
118       }
119
120       jsonObject.model = activityArr;
121
122       newJsonString = JSON.stringify(jsonObject);
```

```
123         console.log(newJsonString);
124         var xhr = new XMLHttpRequest();
125         var url = "https://backend-klassifizierung.stackocean.com/
                privacymarker";
126         xhr.open("POST", url, true);
127         xhr.setRequestHeader("Content-Type", "application/json");
128         xhr.onreadystatechange = function () {
129           if (xhr.readyState === 4 && xhr.status === 200) {
130             console.log(xhr.responseText);
131             let jsonRes = JSON.parse(xhr.responseText);
132             let mark = jsonRes.mark;
133             checkPrivacy(activities, mark, commandStack);
134           }
135         };
136         xhr.send(newJsonString);
137
138
139       }
140     });
141
142     editorActions.register({
143       reset: function () {
144         const elements = elementRegistry.filter(function (element)
                {
145           return element;
146         });
147         elements.forEach(element => {
148           colorElement(element, "#FFFFFF", commandStack);
149         });
150       }
151     });
152
153 }
154
155 function colorElement(element, setcolor, commandStack) {
156     let cFill = "#FFFFFF";
157     let cStroke = "#000000";
158     let newColor = String(setcolor)
159     if (newColor == "#FF0000") {
160       cFill = "#ffcdd2"
161       cStroke = "#e53935"
162     } else if (newColor == "#00FF00") {
163       cFill = "#c8e6c9"
164       cStroke = "#43a047"
165     } else if (newColor == "#FFFF77") {
166       cFill = "#ffe0b2"
167       cStroke = "#fb8c00"
```

```
  } else {
    cFill = newColor
  }

  commandStack.execute('element.setColor', {
    elements: [element],
    colors: {
      fill: cFill,
      stroke: cStroke
    }
  });
}

function checkPrivacy(elements, mark, commandStack) {
  for (let i = 0; i < elements.length; i++) {
    let element = elements[i];
    for (let j = 0; j < mark.length; j++) {
      if (mark[j].id == element.id) {
        colorElement(element, mark[j].color, commandStack)
        break;
      }

    }
  }

}

function stringsMatch(str1, str2, threshold = 0.75) {
  const len1 = str1.length;
  const len2 = str2.length;
  const distance = [];

  for (let i = 0; i <= len1; i++) {
    distance[i] = [];
    for (let j = 0; j <= len2; j++) {
      distance[i][j] = 0;
    }
  }

  for (let i = 1; i <= len1; i++) {
    distance[i][0] = i;
  }
  for (let j = 1; j <= len2; j++) {
    distance[0][j] = j;
  }

  for (let j = 1; j <= len2; j++) {
```

```
215       for (let i = 1; i <= len1; i++) {
216         const cost = str1[i - 1] === str2[j - 1] ? 0 : 1;
217         distance[i][j] = Math.min(
218           distance[i - 1][j] + 1,
219           distance[i][j - 1] + 1,
220           distance[i - 1][j - 1] + cost
221         );
222       }
223     }
224
225     const similarity = 1 - distance[len1][len2] / Math.max(len1,
          len2);
226     tempSim = 1 - distance[len1][len2] / Math.max(len1, len2);
227     return similarity >= threshold;
228 }
229
230 DatenschutzPlugin.$inject = [
231     'commandStack',
232     'editorActions',
233     'elementRegistry',
234     'eventBus'
235 ];
```

Listing C.2: Plugin/menu/menu.js

Literatur

[Ag19] Agostinelli, S.; Maggi, F. M.; Marrella, A.; Sapio, F.: Achieving GDPR Compliance of BPMN Process Models. In: Achieving GDPR Compliance of BPMN Process Models. CAiSE. Bd. 350. Lecture Notes in Business Information Processing, Springer, Cham, Mai 2019, URL: https://www.researchgate.net/publication/333312868_Achieving_GDPR_Compliance_of_BPMN_Process_Models.

[AK22] AK Technik der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder, Hrsg.: Das Standard-Datenschutzmodell. Eine Methode zur Datenschutzberatung und -prüfung auf der Basis einheitlicher Gewährleistungsziele. Version 3. 2022.

[Al09] Allweyer, T.: Prozessmodellierung: Kluft zwischen Praxis und Forschung, Kurze Prozesse, 6. Nov. 2009, URL: https://www.kurze-prozesse.de/2009/11/06/prozessmodellierung-kluft-zwischen-praxis-und-forschung/, Stand: 15. 04. 2023.

[Al17] Albers, M.: Informationelle Selbstbestimmung als vielschichtiges Bündel von Rechtsbindungen und Rechtspositionen. In (Friedewald, M.; Lamla, J.; Roßnagel, A., Hrsg.): Informationelle Selbstbestimmung im digitalen Wandel. DuD-Fachbeiträge, Springer, Wiesbaden, S. 11–35, 2017.

[AMA13] Altuhhova, O.; Matulevičius, R.; Ahmed, N.: An Extension of Business Process Model and Notation for Security Risk Management: International Journal of Information System Modeling and Design 4/4, S. 93–113, 1. Okt. 2013, URL: https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/ijismd.2013100105.

[Ba19] Badakhshan, P.; Conboy, K.; Grisold, T.; vom Brocke, J.: Agile Business Process Management: A Systematic Literature Review and an Integrated Framework. Business Process Management Journal 26/6, S. 1505–1523, 18. Nov. 2019, URL: https://www.emerald.com/insight/content/doi/10.1108/BPMJ-12-2018-0347/full/html.

[BAF07] Becker, J.; Algermissen, L.; Falk, T.: Prozessorientierte Verwaltungsmodernisierung: Prozessmanagement im Zeitalter von E-Government und New Public Management. Springer, Berlin Heidelberg, 2007.

[BAR09] Becker, J.; Algermissen, L.; Räckers, M.: Prozessmodellierung als Schlüssel zur Umsetzung Der EU-DLR: Modellierung und Management von Verwaltungsprozessen auf Basis der EU-DLR mit der PICTURE-Methode. In (Schliesky, U., Hrsg.): Die Umsetzung Der EU-Dienstleistungsrichtlinie in Der Deutschen Verwaltung. Teil II: Verfahren, Prozesse, IT-Umsetzung. Kiel, 2009.

[BCM19] Bartolini, C.; Calabró, A.; Marchetti, E.: Enhancing Business Process Modelling with Data Protection Compliance: An Ontology-based Proposal: In: Proceedings of the 5th International Conference on Information Systems Security and Privacy. 5th International Conference on Information Systems Security and Privacy. SCITEPRESS - Science and Technology Publications, Prague, Czech Republic, S. 421–428, 2019, URL: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0007392304210428.

[Be07a] Becker, J.; Algermissen, L.; Pfeiffer, D.; Räckers, M.: Aufbau eines Verwaltungsübergreifenden Prozessregisters für öffentliche Verwaltungen mit der PICTURE-Methode. In: 10 Jahre IRIS: Bilanz Und Ausblick. Tagungsband Des 10. Internationalen Rechtsinformatik Symposions IRIS 2007. 2007.

[Be07b] Becker, J.; Algermissen, L.; Pfeiffer, D.; Räckers, M.: Bausteinbasierte Modellierung von Prozesslandschaften mit der PICTURE-Methode am Beispiel der Universitätsverwaltung Münster. WIRTSCHAFTSINFORMATIK 49/4, S. 267–279, 1. Aug. 2007, URL: https://doi.org/10.1007/s11576-007-0063-0, Stand: 02. 04. 2023.

[Be20] Berliner Beauftragte für Datenschutz und Informationsfreiheit: Hinweise für Berliner Verantwortliche zu Anbietern von Videokonferenz-Diensten, 3. Juli 2020.

[Be83] Bertin, J.: Semiology of Graphics: Diagrams, Networks, Maps. UMI Research Press, 1983.

[BF20] Besik, S.; Freytag, J.-C.: Managing Consent in Workflows under GDPR. In: ZEUS. Central-European Workshop on Services and their Composition. 18. Juni 2020.

[BHM20] vom Brocke, J.; Hevner, A.; Maedche, A.: Introduction to Design Science Research. In: Design Science Research. Cases. Progress in IS (PROIS), Springer, S. 1–13, 1. Sep. 2020.

[Bi23] Bitkom e. V.: Privacy Icons. Konzeptionierung zur Erstellung sinnvoller Icons in Umsetzung von Art. 12 DS-GVO und Begleitdokument zum Bitkom Icon Set, 2023, URL: https://www.bitkom.org/Themen/Datenschutz-Sicherheit/Privacy-Icons, Stand: 27. 04. 2023.

[BM15] Bartolini, C.; Muthuri, R.: Reconciling Data Protection Rights and Obligations: An Ontology of the Forthcoming EU Regulation. In: Proceedings of the Workshop on Language and Semantic Technology for Legal Domain (LST4LD). Recent Advances in Natural Language Processing (RANLP). 10. Sep. 2015.

[BMS15] Bartolini, C.; Muthuri, R.; Santos, C.: Using Ontologies to Model Data Protection Requirements in Workflows. In. JSAI - isAI Workshops. 17. Nov. 2015.

[BMS18] Barton, T.; Müller, C.; Seel, C., Hrsg.: Digitalisierung in Unternehmen: Von den theoretischen Ansätzen zur praktischen Umsetzung. Springer Fachmedien, Wiesbaden, 2018.

[bp22] bpmn.io: Bpmn-Js Walkthrough | Toolkits, bpmn.io, 2022, URL: https://bpmn.io/toolkit/bpmn-js/walkthrough/, Stand: 28. 04. 2023.

[Br13] Brucker, A. D.: Integrating Security Aspects into Business Process Models. it – Information Technology 55/6, S. 239–246, 1. Dez. 2013, URL: https://www.degruyter.com/document/doi/10.1524/itit.2013.2004/html.

[Bu83] Bundesverfassungsgericht: Volkszählung ("Volkszählungsurteil"), 15. Dez. 1983, URL: https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/1983/12/rs19831215_1bvr020983.html.

[Bu85] Bunge, M.: Epistemology & Methodology III: Philosophy of Science and Technology Part I: Formal and Physical Sciences. D. Reidel Publishing Company, 1985.

[Ca22] Camunda Services GmbH: Über Uns, Camunda, 2022, URL: https://camunda.com/de/about/, Stand: 11. 03. 2023.

[Ch14] Cherdantseva, Y.: Secure*BPMN - a Graphical Extension for BPMN 2.0 Based on a Reference Model of Information Assurance & Security, Cardiff University, Dez. 2014.

[CHR12] Cherdantseva, Y.; Hilton, J.; Rana, O.: Towards SecureBPMN - Aligning BPMN with the Information Assurance and Security Domain. In (Mendling, J.; Weidlich, M., Hrsg.): Business Process Model and Notation. Lecture Notes in Business Information Processing, Springer, Berlin, Heidelberg, S. 107–115, 2012.

[CKO92] Curtis, B.; Kellner, M. I.; Over, J.: Process Modeling. Communications of the ACM 35/9, S. 75–90, 1. Sep. 1992, URL: https://dl.acm.org/doi/10.1145/130994.130998.

[Da17] Datenschutzgruppe nach Artikel 29: WP 248 Rev. 01: Leitlinien zur Datenschutz-Folgenabschätzung (DSFA) und Beantwortung der Frage, ob eine Verarbeitung im Sinne der Verordnung 2016/679 „wahrscheinlich ein hohes Risiko mit sich bringt“, 4. Okt. 2017, URL: https://www.datenschutzkonferenz-online.de/media/wp/20171004_wp248_rev01.pdf, Stand: 16. 04. 2023.

[DA18] DAS EUROPÄISCHE PARLAMENT UND DER RAT DER EUROPÄISCHEN UNION: VERORDNUNG (EU) 2016/679 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung, 25. Mai 2018, URL: https://dejure.org/gesetze/DSGVO.

[De18] Der Bayerische Landesbeauftragte für den Datenschutz: Die Datenschutz-Grundverodnung. Ein Überblick, 25. Mai 2018, URL: https://www.datenschutz-bayern.de/datenschutzreform2018/ueberblick.pdf.

[De22a] Deehan, N.: Camunda Platform 8 for Camunda Platform 7 Users - What You Need to Know, Camunda, 12. Apr. 2022, URL: https://camunda.com/blog/2022/04/camunda-platform-8-for-camunda-platform-7-users-what-you-need-to-know/, Stand: 12. 03. 2023.

[De22b] Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit, Hrsg.: Datenschutz-Grundverordnung – Bundesdatenschutzgesetz. Texte und Erläuterungen. 2022.

[Di18] Die Landesbeauftragte für Datenschutz Schleswig-Holstein: Liste von Verarbeitungsvorgängen nach Art. 35 Abs. 4 DS-GVO für die gemäß Art. 35 Abs. 1 DS-GVO eine Datenschutz-Folgenabschätzung von Verantwortlichen durchzuführen ist. 25. Mai 2018, URL: https://www.datenschutzzentrum.de/uploads/datenschutzfolgenabschaetzung/20180525_LfD-SH_DSFA_Muss-Liste_V1.0.pdf, Stand: 16. 04. 2023.

[Di22] Dieterich, C.: Geschäftsführer haften für Datenschutzverstöße (DSGVO), RTS Steuerberatungsgesellschaft GmbH & Co. KG, 18. März 2022, URL: https://www.rtskg.de/service/news/beitrag/geschaeftsfuehrer-haften-fuer-datenschutzverstoesse-dsgvo.html, Stand: 02. 02. 2023.

[DP21] Doyé, T.; Prexl, A.: Digitalisierung in der öffentlichen Verwaltung – Notwendigkeit, Nutzen und Möglichkeiten: Fallbeispiel aus dem Raum Ingolstadt. In (Lehmann, L.; Engelhardt, D.; Wilke, W., Hrsg.): Kompetenzen für die digitale Transformation 2020. Springer, Berlin, Heidelberg, S. 83–89, 2021.

[DS22] DSK: Geschäftsordnung der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder (Datenschutzkonferenz), 21. Sep. 2022, URL: https://www.datenschutzkonferenz-online.de/media/dsk/Geschaeftsordnung_DSK_09-2022.pdf.

[Du21] Dumas, M.; La Rosa, M.; Mendling, J.; Reijers, H. A.: Grundlagen des Geschäftsprozessmanagements. Übersetzt von Thomas Grisold, Steven Groß, Jan Mendling, Bastian Wurm. Springer Berlin Heidelberg, Berlin, Heidelberg, 2021.

[EG18] Estdale, J.; Georgiadou, E.: Applying the ISO/IEC 25010 Quality Models to Software Product. In (Larrucea, X.; Santamaria, I.; O'Connor, R. V.; Messnarz, R., Hrsg.): Systems, Software and Services Process Improvement. Bd. 896, Springer International Publishing, Bilbao, S. 492–503, 2018, URL: http://link.springer.com/10.1007/978-3-319-97925-0_42, Stand: 18. 02. 2023.

[El07] Elliot, A. J.; Maier, M. A.; Moller, A. C.; Friedman, R.; Meinhardt, J.: Color and Psychological Functioning: The Effect of Red on Performance Attainment. Journal of Experimental Psychology: General 136/1, S. 154–168, 2007, URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0096-3445.136.1.154.

[EM07] Elliot, A. J.; Maier, M. A.: Color and Psychological Functioning. Current Directions in Psychological Science 16/5, S. 250–254, Okt. 2007, URL: http://journals.sagepub.com/doi/10.1111/j.1467-8721.2007.00514.x.

[EMS19] Engelbrecht, K.; Müller, S.; Stief, M.: Auftragsverarbeitung. Orientierungshilfe, hrsg. von Der Bayerische Landesbeauftragte für den Datenschutz, 1. Apr. 2019.

[Ep15] Epure, E.; Martín-Rodilla, P.; Hug, C.; Deneckère, R.; Salinesi, C.: Automatic Process Model Discovery from Textual Methodologies: An Archaeology Case Study. Proceedings - International Conference on Research Challenges in Information Science 2015/, S. 19–30, 19. Juni 2015.

[FI22] FIRST: Traffic Light Protocol, Aug. 2022, URL: https://www.first.org/tlp/docs/tlp-a4.pdf.

[Fl18] Fleischmann, A.; Oppl, S.; Schmidt, W.; Stary, C.: Ganzheitliche Digitalisierung von Prozessen: Perspektivenwechsel – Design Thinking – Wertegeleitete Interaktion. Springer Nature, 2018.

[FMP11] Friedrich, F.; Mendling, J.; Puhlmann, F.: Process Model Generation from Natural Language Text. In (Mouratidis, H.; Rolland, C., Hrsg.): Advanced Information Systems Engineering. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, S. 482–496, 2011.

[FR19a] Freund, J.; Rücker, B.: Praxishandbuch BPMN 2.0: Mit Einführung in DMN. Carl Hanser Verlag GmbH & Co. KG, München, 2019.

[FR19b] Freund, J.; Rücker, B.: Real-Life BPMN: Using BPMN and DMN to Analyze, Improve, and Automate Processes in Your Company. Camunda, Berlin, 2019.

[FWS11] Feja, S.; Witt, S.; Speck, A.: BAM: A Requirements Validation and Verification Framework for Business Process Models. In: 2011 11th International Conference on Quality Software. 2011 11th International Conference on Quality Software. S. 186–191, Juli 2011.

[GBC16] Goodfellow, I.; Bengio, Y.; Courville, A.: Deep Learning. MIT Press, 2016.

[GCC17] Gonçalves, A.; Correia, A.; Cavique, L.: Data Protection Risk Modeling into Business Process Analysis. In (Gervasi, O.; Murgante, B.; Misra, S. et al., Hrsg.): Computational Science and Its Applications – ICCSA 2017. Bd. 10404, Springer International Publishing, Cham, S. 667–676, 2017.

[GKC07] Ghose, A.; Koliadis, G.; Chueng, A.: Process Discovery from Model and Text Artefacts. In: 2007 IEEE Congress on Services (Services 2007). 2007 IEEE Congress on Services (Services 2007). S. 167–174, Juli 2007.

[GL02] Gruninger, M.; Lee, J.: Ontology - Applications and Design. Communications of the ACM 45/2, S. 39–41, 1. Feb. 2002, URL: https://doi.org/10.1145/503124.503146.

[GL13] Göpfert, J.; Lindenbach, H.: Geschäftsprozessmodellierung mit BPMN 2.0: business process model and notation. Oldenbourg Verlag, München, 2013.

[Gl20] Glinz, M.; van Loenhoud, H.; Staal, S.; Bühne, S.: Handbuch für das CPRE Foundation Level nach dem IREB-Standard. International Requirements Enginieering Board, 2020.

[GMS06] Governatori, G.; Milosevic, Z.; Sadiq, S.: Compliance Checking between Business Processes and Business Contracts. 2006.

[GNV17] Geiger, M.; Neugebauer, P.; Vorndran, A.: Automatic Standard Compliance Assessment of BPMN 2.0 Process Models. 2017.

[Go10] Goethe, J. W.: Zur Farbenlehre. Stuttgart, 1810.

[Go23] Goppelt, C.: Evaluation eines Ansatzes zur Darstellung des Datenschutzes in der Prozessmodellierung - Masterarbeit, März 2023.

[Go42] Goldstein, K.: Some Experimental Observations Concerning the Influence of Colors in the Function of the Organism. Occupational Therapy & Rehabilitation, American Journal of Physical Medicine & Rehabilitation 21/3, S. 147–151, 1942, URL: https://journals.lww.com/ajpmr/Citation/1942/06000/SOME_EXPERIMENTAL_OBSERVATIONS_CONCERNING_THE.2.aspx.

[Gr93] Gruber, T. R.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5/2, S. 199–220, Juni 1993, URL: https://linkinghub.elsevier.com/retrieve/pii/S1042814383710083.

[GS23] Gardini, M. A.; Sommer, G.: Digitalisierung und Digital Leadership im Tourismus – Im Gespräch mit Top-Führungskräften aus der Tourismusindustrie. In (Gardini, M. A.; Sommer, G., Hrsg.): Digital Leadership im Tourismus: Digitalisierung und Künstliche Intelligenz als Wettbewerbsfaktoren der Zukunft. Springer Fachmedien, Wiesbaden, S. 43–74, 2023.

[GW04] Gemino, A.; Wand, Y.: A Framework for Empirical Evaluation of Conceptual Modeling Techniques. Requir. Eng. 9/, S. 248–260, 1. Nov. 2004.

[GW22a] Goppelt, C.; Windrich, M.: Interview mit Georg Rasch Am 21.09.22, 21. Sep. 2022.

[GW22b] Goppelt, C.; Windrich, M.: Interview mit Isabelle Puttrus Am 15.08.22, 15. Aug. 2022.

[GW22c] Goppelt, C.; Windrich, M.: Interview mit Ricarda Radden Am 19.09.22, 19. Sep. 2022.

[GW22d] Goppelt, C.; Windrich, M.: Interview nit Stella Thoben Am 08.09.22, 8. Sep. 2022.

[He04] Hevner, A. R.; March, S. T.; Park, J.; Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28/1, S. 75–105, 2004, URL: https://www.jstor.org/stable/25148625.

[He05] Hesse, W.: Ontologie(n), Gesellschaft für Informatik: Informatiklexikon, 28. Juli 2005, URL: https://gi.de/informatiklexikon/ontologien, Stand: 08. 11. 2022.

[He18] Heldt, C.; Amelung, V. E.; Mühlbacher, A.; Krauth, C.: Definition: Compliance, https://wirtschaftslexikon.gabler.de/definition/compliance-27721, 10. Sep. 2018, URL: https://wirtschaftslexikon.gabler.de/definition/compliance-27721/version-333143, Stand: 10. 04. 2023.

[He22] Herrmann, A.: Grundlagen der Anforderungsanalyse: Standardkonformes Requirements Engineering. Springer Fachmedien, Wiesbaden, 2022.

[Hi02] Hitchman, S.: The Details of Conceptual Modelling Notations Are Important - A Comparison of Relationship Normative Language. Communications of the Association for Information Systems 9/1, 17. Sep. 2002, URL: https://aisel.aisnet.org/cais/vol9/iss1/10.

[Hi23] Hilge, P.: Grafische Markierung von Datenschutzaspekten in Geschäftsprozessmodellen - Masterarbeit in Bearbeitung unter Betreuung der Autorin, Stand 03.05.2023.

[HS20] Herbst, D. G.; Schildhauer, T.: Public Relations und Digitalisierung. Herbert von Halem Verlag, 2020.

[In09] Indulska, M.; Green, P.; Recker, J.; Rosemann, M.: Business Process Modeling: Perceived Benefits. In (Laender, A. H. F.; Castano, S.; Dayal, U.; Casati, F.; de Oliveira, J. P. M., Hrsg.): Conceptual Modeling - ER 2009. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, S. 458–471, 2009.

[In23a] InterCard GmbH Kartensysteme: Prospekt Validieren & Mehr, 2023.

[In23b] Intersoft Consuling: Alle Landesdatenschutzgesetze (LDSG) nach Bundesland, 2023, URL: https://dsgvo-gesetz.de/ldsg/, Stand: 22. 08. 2022.

[Je09] Jericho Forum: COA Paper Information Classification, Jan. 2009.

[JH74] Jacobs, K. W.; Hustmyer, F. E.: Effects of Four Psychological Primary Colors on GSR, Heart Rate and Respiration Rate. Perceptual and Motor Skills 38/3, S. 763–766, Juni 1974, URL: http://journals.sagepub.com/doi/10.2466/pms.1974.38.3.763.

[KB21] Khyani, D.; B S, S.: An Interpretation of Lemmatization and Stemming in Natural Language Processing. Shanghai Ligong Daxue Xuebao/Journal of University of Shanghai for Science and Technology 22/, S. 350–357, 7. Jan. 2021.

[KM21] Kummer, T.-F.; Mendling, J.: The Effect of Risk Representation Using Colors and Symbols in Business Process Models on Operational Risk Management Performance. Journal of the Association for Information Systems/22, S. 47, 2021.

[Ko19] Kowsari; Jafari Meimandi; Heidarysafa et al.: Text Classification Algorithms: A Survey. Information 10/4, S. 150, 23. Apr. 2019, URL: https://www.mdpi.com/2078-2489/10/4/150.

[Le20] Leifels, D. A.: Mangel an Digitalkompetenzen bremst Digitalisierung des Mittelstands – Ausweg Weiterbildung? 2020.

[LS18] Lackes, R.; Siepermann, M.: Definition: Künstliche Intelligenz (KI). In: Gabler Wirtschaftslexikon. Springer Fachmedien Wiesbaden, 19. Feb. 2018, URL: https://wirtschaftslexikon.gabler.de/definition/kuenstliche-intelligenz-ki-40285/version-263673, Stand: 15. 03. 2023.

[Ma21] Matzka, S.: Künstliche Intelligenz in den Ingenieurwissenschaften: Maschinelles Lernen verstehen und bewerten. Springer Fachmedien, Wiesbaden, 2021.

[Mc99] McShane, C.: The Origins and Globalization of Traffic Control Signals. Journal of Urban History 25/3, S. 379–404, März 1999, URL: http://journals.sagepub.com/doi/10.1177/009614429902500304.

[Me22] Meuche, T.: Dilemmata und Wege zur Digitalisierung der öffentlichen Verwaltung. Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO) 53/1, S. 99–108, 1. März 2022.

[Mo10] Moody, D.: The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. Software Engineering, IEEE Transactions on 35/, S. 756–779, 1. Jan. 2010.

[Mö20] Mönikes, J.: Weitergabe von Patientendaten: Abkehr von der „Einwilligeritis" in der Arztpraxis? - DSB Ratgeber, 7. Dez. 2020, URL: https://www.dsb-ratgeber.de/artikel/Datenschutz-Weitergabe-Patientendaten.html, Stand: 20. 03. 2023.

[MSB11] Mülle, J.; Stackelberg, S. von; Böhm, K.: A Security Language for BPMN Process Models. Karlsruher Institut für Technologie (KIT), 2011.

[Na01] Navarro, G.: A Guided Tour to Approximate String Matching. ACM Computing Surveys 33/1, S. 31–88, März 2001, URL: https://dl.acm.org/doi/10.1145/375360.375365.

[Ne22] Neuerer, D.: DIHK-Analyse: Viele Unternehmen sehen sich durch „datenschutzrechtliche Hemmnisse" behindert, 22. Feb. 2022, URL: https://www.handelsblatt.com/politik/deutschland/dihk-analyse-viele-unternehmen-sehen-sich-durch-datenschutzrechtliche-hemmnisse-behindert/28089652.html, Stand: 15. 04. 2023.

[Ni15] Niemand, S.; Feja, S.; Witt, S.; Speck, A.: On Improving the Maintainability of Compliance Rules for Business Processes. In (Abramowicz, W., Hrsg.): Business Information Systems. Bd. 208, Springer International Publishing, Cham, S. 178–190, 2015, URL: https://link.springer.com/10.1007/978-3-319-19027-3_15.

[Ni22] Niemand, S.: Vom Rechtstext zum regelkonformen Geschäftsprozess: Konstruktion und Anwendung von Prozessmodellen, die auf gesetzlichen und vertraglichen Regelungen basieren, Dissertation, Kiel: Christian-Albrechts-Universität zu Kiel, 25. März 2022, URL: https://nbn-resolving.org/urn:nbn:de:gbv:8:3-2022-00134-4.

[No20] Nolte, F. R.: Text to Process Model: Automating Process Model Creation from Text, Münster: Westfälische Wilhelms-Universität, Dez. 2020.

[Ob13] Object Management Group: Business Process Model and Notation (BPMN), Version 2.0, 9. Dez. 2013, URL: https://www.omg.org/spec/BPMN/2.0.2/PDF.

[OL21] OLG Dresden: Urteil 4 U 1158/21, 30. Nov. 2021, URL: https://openjur.de/u/2381765.html, Stand: 27. 01. 2023.

[Pa86] Paivio, A.: Mental Representations : A Dual Coding Approach. Oxford Univ. Press [u.a.], Oxford [u.a.], 1986.

[Pe07] Peffers, K.; Tuunanen, T.; Rothenberger, M.; Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24/, S. 45–77, 1. Jan. 2007.

[Pe20] Peñaloza, R.: Introduction to Probabilistic Ontologies. In (Manna, M.; Pieris, A., Hrsg.): Reasoning Web. Declarative Artificial Intelligence. Bd. 12258, Springer International Publishing, Cham, S. 1–35, 2020, URL: http://link.springer.com/10.1007/978-3-030-60067-9_1.

[Pe22] Peci, F.: Datenschutzklassifizierung von Geschäftsprozessmodellen mit Hilfe von NLP, Bachelorarbeit, 31. März 2022.

[PH20] Paaß, G.; Hecker, D.: Künstliche Intelligenz: Was steckt hinter der Technologie der Zukunft? Springer Fachmedien, Wiesbaden, 2020.

[PMB17] Pullonen, P.; Matulevičius, R.; Bogdanov, D.: PE-BPMN: Privacy-Enhanced Business Process Model and Notation. In: Lecture Notes in Computer Science. Business Process Management 2017. Bd. 10445, 10. Aug. 2017.

[PR15] Pohl, K.; Rupp, C.: Basiswissen Requirements Engineering: Aus- und Weiterbildung nach IREB-Standard zum Certified Professional for Requirements Engineering Foundation Level. dpunkt.verlag GmbH, Heidelberg, 2015.

[PS16] Pereira, J. L.; Silva, D.: Business Process Modeling Languages: A Comparative Framework. In (Rocha, Á.; Correia, A. M.; Adeli, H.; Reis, L. P.; Mendonça Teixeira, M., Hrsg.): New Advances in Information Systems and Technologies. Bd. 444, Springer International Publishing, Cham, S. 619–628, 2016, URL: http://link.springer.com/10.1007/978-3-319-31232-3_58.

[Pu19a] Pullonen, P.; Tom, J.; Matulevičius, R.; Toots, A.: Privacy-Enhanced BPMN: Enabling Data Privacy Analysis in Business Processes Models. Software and Systems Modeling/18, 2019, URL: https://link.springer.com/article/10.1007/s10270-019-00718-z.

[Pu19b] Puttrus, I.: Technische und organisatorische Maßnahmen nach DSGVO - Masterarbeit, Apr. 2019.

[Ra94] Raskin, J.: Viewpoint: Intuitive Equals Familiar. Communications of the ACM 37/9, S. 17–18, 1. Sep. 1994, URL: https://dl.acm.org/doi/10.1145/182987.584629.

[Re20] Reinhardt, K.: Einführung: Warum Digitalisierung wichtig ist. In (Reinhardt, K., Hrsg.): Digitale Transformation der Organisation: Grundlagen, Praktiken und Praxisbeispiele der digitalen Unternehmensentwicklung. Springer Fachmedien, Wiesbaden, S. 1–9, 2020.

[RE22] Rat der EU; Europäischer Rat: Eine digitale Zukunft für Europa, 15. Dez. 2022, URL: https://www.consilium.europa.eu/de/policies/a-digital-future-for-europe/, Stand: 10. 04. 2023.

[RFP07] Rodriguez, A.; Fernández-Medina, E.; Piattini, M.: A BPMN Extension for the Modeling of Security Requirements in Business Processes. IEICE Transactions on Information and Systems E90D/, 1. März 2007.

[Ri15] Riguzzi, F.; Bellodi, E.; Lamma, E.; Zese, R.: Reasoning with Probabilistic Ontologies. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence. IJCAI. S. 4316, 2015.

[Ri20] Risius, P.: Digitaisierung der Ausbildung. Neue Kompetenzen für eine Arbeitswelt im Wandel. Netzwerk Q 4.0, 2020.

[Ro18] Roßnagel, A.: Umsetzung der Unionsregelungen zum Datenschutz: Erste Erfahrungen mit der Datenschutz-Grundverordnung aus rechtswissenschaftlicher Sicht. Datenschutz und Datensicherheit - DuD 42/12, S. 741–745, Dez. 2018, URL: http://link.springer.com/10.1007/s11623-018-1037-7.

[Ro19] Robles, R. R.: Development of a BPMN Model Validation and Automatic Correction Tool for SAP Solution Manager - Bachelorarbeit, 4. Sep. 2019.

[Rü21] Rühl, G.: Europa Sollte in Der Digitalisierung Verstärkt Zusammenarbeiten. In (Flick, C. M., Hrsg.): Neue Konstellationen Der Gegenwart: Annäherungen, Institutionen Und Legitimität. Wallstein Verlag, 2021.

[Sc15] Schach, A.: Corporate Language/Manual. In (Schach, A., Hrsg.): Advertorial, Blogbeitrag, Content-Strategie & Co.: Neue Texte der Unternehmenskommunikation. Springer Fachmedien, Wiesbaden, S. 83–93, 2015.

[Sc18] Schmedt, M.: Hinweise und Empfehlungen zur ärztlichen Schweigepflicht, Datenschutz und Datenverarbeitung in der Arztpraxis, Deutsches Ärzteblatt Online, 9. März 2018, URL: https://www.aerzteblatt.de/down.asp?id=20474, Stand: 27. 03. 2023.

[Sc20] Schäffer, U.: Digitalisierung ist kein Selbstzweck, sie muss sich rechnen. Controlling & Management Review 64/1, S. 16–21, 1. Jan. 2020.

[SCV11] Stroppi, L. J. R.; Chiotti, O.; Villarreal, P. D.: Extending BPMN 2.0: Method and Tool Support. In (Dijkman, R.; Hofstetter, J.; Koehler, J., Hrsg.): Business Process Model and Notation. Lecture Notes in Business Information Processing, Springer, Berlin, Heidelberg, S. 59–73, 2011.

[SDG17] Salnitri, M.; Dalpiaz, F.; Giorgini, P.: Designing Secure Business Processes with SecBPMN. Software & Systems Modeling 16/3, S. 737–757, 1. Juli 2017, URL: https://doi.org/10.1007/s10270-015-0499-4.

[Se21] Seifert, C.: Was die Digitalisierung dem Fußball bringt – und warum die analoge Kommunikation trotzdem wichtig bleibt. In (Hildebrandt, A.; Landhäußer, W., Hrsg.): CSR und Digitalisierung: Der digitale Wandel als Chance und Herausforderung für Wirtschaft und Gesellschaft. Management-Reihe Corporate Social Responsibility, Springer, Berlin, Heidelberg, S. 851–853, 2021.

[SGN07] Sadiq, S.; Governatori, G.; Namiri, K.: Modeling Control Objectives for Business Process Compliance. In (Alonso, G.; Dadam, P.; Rosemann, M., Hrsg.): Business Process Management. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, S. 149–164, 2007.

[SMB21] Seckler, C.; Mauer, R.; vom Brocke, J.: Design Science in Entrepreneurship: Conceptual Foundations and Guiding Principles. Journal of Business Venturing Design 1/1, S. 100004, 1. Juli 2021, URL: https://www.sciencedirect.com/science/article/pii/S2667277422000019.

[Sn15] Snoeck, M.; Moreno-Montes de Oca, I.; Haegemans, T.; Scheldeman, B.; Hoste, T.: Testing a Selection of BPMN Tools for Their Support of Modelling Guidelines. In (Ralyté, J.; España, S.; Pastor, Ó., Hrsg.): The Practice of Enterprise Modeling. Lecture Notes in Business Information Processing, Springer International Publishing, Cham, S. 111–125, 2015.

[So16] Sommerville, I.: Software Engineering. Pearson, Boston, 2016.

[SRZ22] Schönnenbeck, C.; Runschke, F.; Zang, D.: Camunda 8 vs. Camunda 7: Wo liegen die Unterschiede?, viadee, 31.10.22, URL: https://blog.viadee.de/camunda-8-release, Stand: 12. 03. 2023.

[SS21] Schenk, B.; Schneider, C.: Innovative Prozessmodellierung. In (Schenk, B.; Schneider, C., Hrsg.): Innovative Services und Prozesse für Kommunen: Wie mit innovativer Prozessmodellierung die öffentliche Verwaltung bürgernäher und digitaler werden kann. essentials, Springer Fachmedien, Wiesbaden, S. 7–28, 2021.

[St12] Stuht, T.; Speck, A.; Feja, S.; Witt, S.; Pulvermüller, E.: Rule Determination and Process Verification Using Business Capabilities. In (Sandkuhl, K.; Seigerroth, U.; Stirna, J., Hrsg.): The Practice of Enterprise Modeling. Bd. 134, Springer Berlin Heidelberg, Berlin, Heidelberg, S. 46–60, 2012, URL: http://link.springer.com/10.1007/978-3-642-34549-4_4.

[ST21] Schmidtner, M.; Timinger, H.: HyValue – Ein adaptives Referenzmodell für den hybriden Produktentstehungsprozess in der Automobilindustrie. In (Lehmann, L.; Engelhardt, D.; Wilke, W., Hrsg.): Kompetenzen für die digitale Transformation 2020. Springer, Berlin, Heidelberg, S. 37–48, 2021.

[SW20] Streim, A.; Weiß, R.: Jedes 2. Unternehmen verzichtet aus Datenschutzgründen auf Innovationen | Presseinformation | Bitkom e.V. 29. Sep. 2020, URL: https://www.bitkom.org/Presse/Presseinformation/Jedes-2-Unternehmen-verzichtet-aus-Datenschutzgruenden-auf-Innovationen, Stand: 15. 04. 2023.

[SZ15] Sang, K. S.; Zhou, B.: BPMN Security Extensions for Healthcare Process. In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. S. 2340–2345, Okt. 2015.

[Tr22] Tremp, H.: Agile objektorientierte Anforderungsanalyse: Planen – Ermitteln – Analysieren – Modellieren – Dokumentieren – Prüfen. Springer Fachmedien Wiesbaden, Wiesbaden, 2022.

[Uh21] Uhlemann, N.: Contributing to the Camunda Modeler and Getting Started with BPMN.io, Camunda, 31. März 2021, URL: https://camunda.com/blog/2021/03/contributing-to-the-camunda-modeler-and-getting-started-with-bpmn-io/, Stand: 11. 03. 2023.

[Ve21] Veigel, P.: Die Einwilligung nach der Datenschutz-Grundverordnung. Orientierungshilfe, hrsg. von Der Bayerische Landesbeauftragte für den Datenschutz, 1. Sep. 2021.

[Ve23] Velovic, S.: Entwicklung eines Plugins zur Visualisierung von datenschutzkritischen Aktivitäten und Datenobjekten in BPMN Modellen - Bachelorarbeit, 31. März 2023.

[We20] Weber, F.: Künstliche Intelligenz. In (Weber, F., Hrsg.): Künstliche Intelligenz für Business Analytics: Algorithmen, Plattformen und Anwendungsszenarien. Springer Fachmedien, Wiesbaden, S. 37–72, 2020.

[Wi22] Wikipedia: Camunda BPM, Wikipedia, 9. Nov. 2022, URL: https://de.wikipedia.org/w/index.php?title=Camunda_BPM, Stand: 11. 03. 2023.

[WSG21] Windrich, M.; Speck, A.; Gruschka, N.: Representing Data Protection Aspects in Process Models by Coloring. In (Gruschka, N.; Antunes, L. F. C.; Rannenberg, K.; Drogkaris, P., Hrsg.): Privacy Technologies and Policy. Bd. 12703, Springer International Publishing, Cham, S. 143–155, 2021, URL: https://link.springer.com/10.1007/978-3-030-76663-4_8.

[Za19] Zarour, K.; Benmerzoug, D.; Guermouche, N.; Drira, K.: A Systematic Literature Review on BPMN Extensions. Business Process Management Journal 26/6, S. 1473–1503, 18. Nov. 2019, URL: https://www.emerald.com/insight/content/doi/10.1108/BPMJ-01-2019-0040/full/html.

[ZAA20] Zareen, S.; Akram, A.; Ahmad Khan, S.: Security Requirements Engineering Framework with BPMN 2.0.2 Extension Model for Development of Information Systems. Applied Sciences 10/14, S. 4981, 20. Juli 2020, URL: https://www.mdpi.com/2076-3417/10/14/4981.

[ZL21] Ziehmann, J.; Lantow, B.: Agilität im Geschäftsprozessmanagement. In. INFORMATIK 2020. Lecture Notes in Informatics, Gesellschaft für Informatik, Bonn, 2021, URL: http://dl.gi.de/handle/20.500.12116/34771.

Abbildungsverzeichnis

Tabellenverzeichnis

Quellcodeverzeichnis

Abkürzungsverzeichnis

AWS	Amazon Web Services
BBO	BPMN 2.0 Based Ontology for Business Process Representation
BDSG-neu	Neues Bundesdatenschutzgesetz
BPM	Business Process Management
BPMN	Business Process Model and Notation
BPMN MIWG	BPMN Model Interchange Working Group
CAU	Christian-Albrechts-Universität zu Kiel
DS	Datenschutz
DSGVO	Datenschutz-Grundverordnung
DSK	Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder
eEPK	Erweiterte Ereignisgesteuerte Prozessketten
EStG	Einkommensteuergesetz
EU	Europäische Union
GKE	Google Kubernetes Engine
GPM	Geschäftsprozessmodell
KI	Künstliche Intelligenz
KK	Krankenkasse
LDSG	Landesdatenschutzgesetz
MFA	Medizinischer Fachangestellter
ML	Machine Learning
NLP	Natural Language Processing
OMG	Object Management Group
SaaS	Software as a Service
SDM	Standard-Datenschutzmodell
SGB IV	Sozialgesetzbuch Viertes Buch
SV	Sozialversicherung
TF-IDF	Term Frequency - Inverse Document Frequency
TLP	Traffic Light Protocol
TOM	Technische und organisatorische Maßnahmen
UML	Unified Modeling Language
XML	Extensible Markup Language