

**Universität Leipzig  
Fakultät für Mathematik und Informatik  
Institut für Informatik**

**Modellierung phänotypischer Beschreibungen  
auf der Grundlage von Bio-Ontologien**

**Bachelorarbeit**

Leipzig, November 2011

vorgelegt von

Stumpf, Frank  
Studiengang B.Sc. Informatik

**Betreuender Hochschullehrer: Prof. Dr. Heinrich Herre  
Fakultät für Mathematik und Informatik  
Institut für Informatik  
Arbeitsgruppe Formale Konzepte**

**Betreuender Assistent: Dipl.-Inf. Frank Loebe  
Fakultät für Mathematik und Informatik  
Institut für Informatik  
Abteilung Intelligente Systeme**

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>3</b>
1.1 Hintergrund	3
1.2 Phenotype and Trait Ontology	4
<b>2. Repräsentation von Phänotypen</b>	<b>7</b>
2.1 Formate für Bio-Ontologien	7
2.1.1 OBO Flat File Format	7
2.1.2 OWL/RDF-Syntax und ihre zugrundeliegende Beschreibungslogik	9
2.2 EQ-Modell	14
2.3 Cross-Products	15
<b>3. Verwendung von Phänotypen in biologischen Datenbanken</b>	<b>17</b>
3.1 Annotation	17
3.2 Grundlegende Anfragen	18
<b>4. Verbesserung des herkömmlichen EQ-Modells</b>	<b>22</b>
4.1 Vertauschungsproblem	22
4.1.1 Problembeschreibung	22
4.1.2 Lösungsansatz	24
4.2 Kontextinsensitivität von <i>inheres_in</i> und <i>towards</i>	27
4.2.1 Problembeschreibung	27
4.2.2 Lösungsansätze	29
4.2.2.1 Relationale GFO-Rollen	32
4.2.2.2 Kontextsensitive Properties	36
4.2.2.3 Künstlicher Relator	39
4.2.2.4 Vergleich der Varianten	40
<b>5. Schlusswort</b>	<b>42</b>
<b>6. Literaturverzeichnis</b>	<b>44</b>
<b>7. Anlagenverzeichnis</b>	<b>50</b>
<b>8. Anlagen</b>	<b>51</b>
<b>9. Erklärung</b>	<b>70</b>

# 1. Einleitung

## 1.1 Hintergrund

Noch nie hatte eine Gesellschaft so viel Wissen zu verwalten, wie die heutige. Allein im indexierbaren Teil des World Wide Web sind 13,8 Milliarden Dokumente verfügbar (siehe [deK11]). Dies entspricht rund 10,24 Petabyte an Informationen (bei einer zugrunde gelegten durchschnittlichen Dokumentgröße von 816 KB (siehe [Sou11])). Zudem ist die Wachstumsrate des Internets mit ungefähr 1 Million neuer Dokumente am Tag (siehe [Hey08]) enorm.

Im Zeitalter computergestützter Wissensverarbeitung muss man sich die Frage stellen, wie all diese Informationen effektiv und effizient genutzt werden können. Schließlich nutzt einem die Fülle an Daten nichts, wenn man nicht die Information findet, die man sucht. Stellt man beispielsweise die Anfrage „In welchem Jahr ist G. W. Leibniz geboren?“ an eine Suchmaschine, so möchte man nicht alle Dokumente finden, die diese Information enthalten könnten, sondern möglichst die Antwort „1646“. Hierzu reichen jedoch vielfach die Methoden des Information Retrieval und verschiedene Ansätze des Text Mining nicht mehr aus. Vielmehr muss man die Semantik von Texten betrachten, die sich Menschen erschließt, Computern jedoch nicht so einfach.

Ein vielversprechender Ansatz, der zumindest den Grundstein für kommende Lösungsansätze bilden könnte, sind Ontologien. Bei einer Ontologie (siehe [Jan08] und [Rus04] als Einführung) handelt es sich um eine Art der Wissensrepräsentation, die „[...] defines a set of representational primitives with which to model a domain of knowledge or discourse.“ [Gru09, 2. Abschnitt]. Eine Domäne wird also durch Definition ihrer Elemente und deren Beziehungen untereinander beschrieben. Setzt man nun ein Dokument mit einer Ontologie gleicher Domäne in Verbindung, so verleiht dies dem Dokument eine gewisse Semantik.

Besonders nützlich erweist sich dies, aufgrund des hohen Informationsgrades, beim Umgang mit wissenschaftlichen Texten oder Datensätzen, beispielsweise aus der Biomedizin. Aufgrund dessen hat sich im Jahr 2001 die

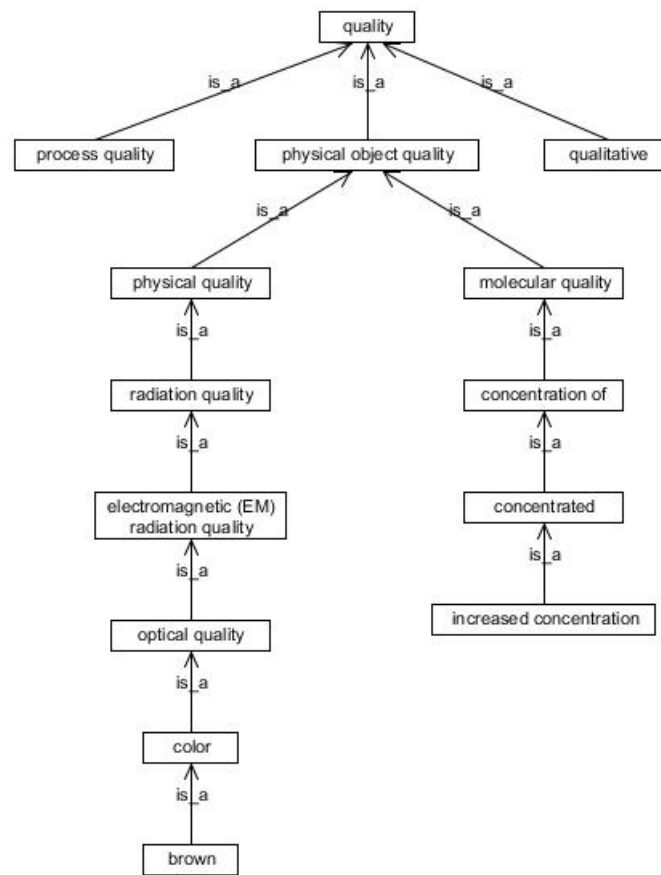
Open Biological and Biomedical Ontology Foundry (Abk.: OBO Foundry) gegründet. Sie sieht ihre Kernaufgabe (siehe [Mun10b]) in der geeigneten Partitionierung der biologischen und biomedizinischen Domäne und der Sicherstellung logischer Kohärenz zwischen Ontologien des Fachgebiets. Diese Maßnahmen sollen dazu dienen Ontologien, wie Chemical Entities of Biological Interest (Abk.: ChEBI), Mouse adult gross anatomy (Abk.: MA) und Phenotype and Trait Ontology (Abk.: PATO), zu entwickeln und zu verwalten.

Von besonderem Interesse bei dieser Arbeit ist die Ontologie PATO (für entsprechendes OBO-Dokument siehe [Gko11]). Es handelt sich dabei um eine Ontologie von Qualitäten. Dementsprechend sind dort Konzepte wie *concentration of* (PATO:0000033) oder *color* (PATO:0000014) definiert. Diese Qualitäten können in Verbindung mit Konzepten spezies-spezifischer Ontologien (z. B. ChEBI oder MA) zur Beschreibung von Phänotypen genutzt werden. Hierbei treten jedoch einige Probleme auf, die im weiteren Verlauf der Arbeit aufgezeigt und gelöst werden.

## 1.2 Phenotype and Trait Ontology

Die, im Jahre 2002 von Michael Ashburner vorgeschlagene (siehe [Gko04]), Phenotype and Trait Ontology ist die einzige Ontologie von Qualitäten zur Beschreibung von Phänotypen. Als Phänotyp ist „[...] *a collection of characteristics that arise through the expression of the genes of an organism, in an environment [...]*“ [Mun10b, S. 1] zu verstehen. Diese Charakteristika (auch: phänotypische Charakteristika oder Phäne) können morphologischer, physiologischer oder psychologischer Art sein (siehe [Wik11b]). So ist z. B. „braunes Auge“ eine morphologische Charakteristik, „erhöhte Eisenkonzentration in der Milz“ eine physiologische und „abnormale Futterpräferenz“ eine psychologische Charakteristik. Die Kollektion dieser phänotypischen Charakteristika „braunes Auge und erhöhte Eisenkonzentration in der Milz und abnormale Futterpräferenz“ bildet den Phänotyp eines Organismus. Im weiteren Verlauf der Arbeit werden, entsprechend der üblichen Literatur, einzelne Phäne bereits als Phänotypen bezeichnet.

Als Ontologie mit dem Zweck der Definition phänotypischer Charakteristika beschreibt PATO Qualitäten (auch: Eigenschaften) und deren Beziehungen untereinander. Eine Qualität (siehe [Mun06]) bezeichnet in PATO Attribute, wie *color* (PATO:0000014) oder *concentration of* (PATO:0000033) und deren korrespondierende Werte, wie *brown* (PATO:0000952) oder *increased concentration* (PATO:0001162). Diese Werte können nur qualifizierend sein, d. h. konkrete Angaben zur Quantität, wie z. B. „10 mg“ für das Attribut *concentration of* sind nicht möglich. Es kann zudem eine Unterscheidung in monadische und relationale Qualitäten angestellt werden (siehe [Gko07]), die so jedoch nur noch in abgeschwächter Form in PATO vorzufinden ist (siehe [Bal10]). Eine monadische Qualität inhäriert lediglich in einer Entität und benötigt dazu keine weitere. Ein Beispiel hierfür ist *brown*, das nur eine Entität (z. B. ein Auge) färbt. Relationale Qualitäten benötigen hingegen mindestens zwei Entitäten um ausgeprägt zu werden. So zeigt *increased concentration* eine erhöhte Konzentration einer Substanz in einem Stoffgemisch an und verbindet somit zwei Entitäten miteinander. Da Eigenschaften die einzigen Konzepte in der Ontologie sind, ergibt sich eine, durch Supersumption (auch: Generalisierung) gekennzeichnete, Hierarchie von Qualitäten. Die Wurzel des Supersumptionsbaumes bildet der Term (in etwa auch: Konzept, Klasse, Kategorie oder Universalie) *quality* (PATO:0000001). Je näher man den Blättern entgegenkommt, desto spezieller werden die Qualitäten. Ein Beispiel dessen lässt sich Abbildung 1 entnehmen. An dem dort dargestellten Baum lässt sich erkennen, dass jedes Konzept lediglich über eine *is\_a*-Verbindung (siehe [Smi05]) mit seinen Oberkonzepten verbunden ist. Bei PATO handelt es sich jedoch nicht nur um eine einfache Taxonomie, weshalb neben der *is\_a*-Relation auch andere Beziehungen existieren. Beispiele hierfür sind *has\_part*, *realized\_by* oder *has\_ratio\_quality*. Eine vollständige Auflistung der Relationen lässt sich [Gko11] entnehmen.



**Abbildung 1:** Ausschnitt aus dem Supersumptionsbaum von PATO

Der eben beschriebene Aufbau von PATO gilt erst seit 2006. Davor fand eine Aufspaltung in Attribut- und Werte-Hierarchie statt. Nähere Informationen hierzu lassen sich [Ope09], [Gko04], [Gko05], [Lew06] und [Mun06] entnehmen.

## 2. Repräsentation von Phänotypen

### 2.1 Formate für Bio-Ontologien

#### 2.1.1 OBO Flat File Format

Die OBO Foundry hat mit dem OBO Flat File Format ein proprietäres textuelles Format zur Repräsentation von Ontologien spezifiziert, dass aktuell in der Version 1.4 vorliegt. Dieses wird sowohl vom Ontologie-Editor OBO-Edit (siehe [OBOOJ]), als auch von allen, von der OBO Foundry verwalteten, Ontologien genutzt, sodass auch PATO darin vorliegt. Daher folgt eine kurze Einführung in die wesentlichsten Bestandteile des OBO Flat File Formats. Für eine ausführlichere Erklärung sei hier auf [Day06], [Mun10a] und [Hor07] verwiesen.

Jedes OBO-Dokument beginnt mit einem Header, der Meta-Daten, wie Format-Version oder Erstellungsdatum und Referenzen zu anderen Ontologien enthalten kann. Darauf folgt eine Sequenz von Absätzen (engl.: Stanzas). Diese Absätze bilden den Kern bei der Ontologie-Repräsentation, da in diesen die Konzepte, Instanzen und Relationen beschrieben werden. Das OBO Flat File Format unterscheidet zwischen drei verschiedenen Typen von Absätzen. Der erste Typ *[Term]* signalisiert, dass es sich beim korrelierenden Absatz um eine Konzeptbeschreibung handelt. Der zweite Typ *[Instance]* gibt an, dass der folgende Absatz eine Instanz eines Konzepts definiert und der dritte Typ *[Typedef]* zeigt eine Property-Beschreibung (Definition einer Relation) an. Der, für den weiteren Verlauf, wichtigste Typ dürfte *[Term]* sein, da in PATO und anderen Bio-Ontologien hauptsächlich Konzepte beschrieben und diese auch für spätere Cross-Products oder Annotationen verwendet werden. Ein Beispiel kann vorab Beschreibung 2 entnommen werden.

*[Term]*  
*id: PATO:0001162*  
*name: increased concentration*  
*def: „A concentration which is higher relative to the normal or average.“*  
*[PATO:C:GVG]*  
*subset: relational\_slim*  
*subset: value\_slim*  
*synonym: „high concentration“ EXACT []*  
*is\_a: PATO:0001159 ! implied link automatically realized ! concentrated*  
*intersection\_of: PATO:0001159 ! concentrated*  
*intersection\_of: increased\_in\_magnitude\_relative\_to PATO:0000461 ! normal*

**Beschreibung 2:** *[Term]*-Abschnitt für das Konzept *increased concentration* (PATO:0001162) in PATO

Für einen *[Term]*-Abschnitt ist lediglich das Tag *id* verpflichtend. Dieses soll dem Term (in etwa auch: Konzept, Klasse, Kategorie oder Universalie) eine Identifizierungsnummer zuweisen, die eine global eindeutige Bezugnahme ermöglicht. Diese Identifizierungsnummer besteht, gemäß [RutOJ], aus dem ID-Raum (z. B. PATO oder MP) und einer, in diesem Raum, eindeutigen Nummer, wodurch die Interoperabilität zwischen verschiedenen Ontologien der OBO Foundry unterstützt werden soll. Auf das *id*-Tag können weitere optionale Tags folgen, die für eine gültige Term-Beschreibung jedoch nicht notwendig sind (auch wenn diese den Term eigentlich erst beschreiben). Die wichtigsten dieser sind in Tabelle 3 aufgelistet und beschrieben.



Tag	Beschreibung
<i>name</i>	Mit dem <i>name</i> -Tag kann dem Term ein informeller Name zugewiesen werden, um die Lesbarkeit für den Menschen zu erhöhen.
<i>def</i>	Mit diesem Tag ist es möglich dem Term eine informelle Beschreibung zu geben. Sie ist nur für menschliche Leser bestimmt, da die maschinelle Verarbeitung hierfür eingeschränkt ist. Die Definition ist in Anführungsstriche zu setzen und zum Schluss folgt die Angabe der Quelle.
<i>subset</i>	Dieses Tag zeigt an, dass der Term zu einer, im Header definierten, Term-Teilmenge korrespondiert. Für PATO können dies <i>abnormal_slim</i> , <i>absent_slim</i> , <i>attribute_slim</i> , <i>cell_quality</i> , <i>disposition_slim</i> , <i>relational_slim</i> , <i>scalar_slim</i> und <i>value_slim</i> sein.
<i>synonym</i>	Das <i>synonym</i> -Tag ermöglicht es dem Term einen Namen zu geben, der synonym zum eigentlichen verwendet werden kann. Das Synonym wird in Anführungsstriche gesetzt. Darauf können eine Angabe der Synonym-Exaktheit, ein Typ und Quellen folgen.
<i>is_a</i>	Das <i>is_a</i> -Tag zeigt die Generalisierungsbeziehung zwischen dem, zur angegebenen Term-ID korrespondierenden, Konzept und dem Konzept, zu dem die Term-Beschreibung gehört, an.
<i>intersection_of</i>	Mit <i>intersection_of</i> lässt sich der Schnitt über angegebene Konzepte bilden. Je Tag darf nur eine Term-ID oder ein Paar aus Relation-ID und Term-ID angegeben werden. Es wird empfohlen, dass die Schnittbildung dem Genus-Differentia-Muster folgt, d. h. ein Tag gibt eine Term-ID an und die anderen geben Paare aus Relation-ID und Term-ID an.
<i>union_of</i>	Das <i>union_of</i> -Tag bildet die Vereinigung über alle angegebenen Konzepte und kann analog zu <i>intersection_of</i> genutzt werden.
<i>relationship</i>	Mit diesem Tag lässt sich die Beziehung zwischen dem Konzept, zu dem die Term-Beschreibung gehört und dem angegebenen Konzept darstellen. Die ID der Relation wird noch vor der ID des Terms angegeben. Diese Beziehung ist jedoch nur eine notwendige, jedoch keine hinreichende.

**Tabelle 3:** die wichtigsten optionalen Tags für *[Term]*-Abschnitte

### 2.1.2 OWL/RDF-Syntax und ihre zugrundeliegende Beschreibungslogik

Dieses Kapitel stellt die wichtigsten Konzepte der Web Ontology Language (Abk.: OWL) vor und liefert zugleich eine Übersetzung in das OBO Flat File Format und in die formale Syntax der Beschreibungslogik SROIQ. Da dieses Kapitel lediglich

einführenden Charakter besitzt, sei an dieser Stelle auf [Hit08], [Hit09], [Smi04], [Man09], [Pat04], [Hor03], [Hor07] und [Hor09] als OWL-Literatur und auf [Baa02], [Fra0J], [Kol0J] und [Sch0J] als Abhandlungen zur Beschreibungslogik verwiesen. Bei OWL handelt es sich um eine Beschreibungssprache zur Realisierung von Web-Ontologien. Sie ist domänenübergreifend die meistverwendete Sprache dieser Art und liegt aktuell in der Version 2.0 vor. Neben der hier beschriebenen OWL/RDF-Syntax, sind zwei weitere Formate gebräuchlich. Die funktionale Syntax ist sehr kompakt und wird in vielen Dokumenten des World Wide Web Consortiums (Abk.: W3C) verwendet und die Manchester-Syntax ähnelt stark der formalen Repräsentation in der Beschreibungslogik und findet daher oft in wissenschaftlichen Artikeln Anwendung. Das offizielle Austauschformat bildet jedoch die OWL/RDF-Syntax, weshalb diese auch im weiteren Verlauf des Kapitels verwendet wird. Es handelt sich dabei um eine Erweiterung des Resource Description Framework Schemas (Abk.: RDFS), welches selbst auf die Extensible Markup Language (Abk.: XML) aufsetzt. Ähnlich wie OBO-Dokumente beginnt auch ein OWL-Dokument mit einem Header, der an dieser Stelle aber vernachlässigt wird. Stattdessen soll ein genauerer Blick auf die beschriebenen Entitäten geworfen werden. Diese sind die gleichen, wie schon aus dem OBO Flat File Format bekannt. Folglich können Klassen (siehe Beschreibung 4), Instanzen und Properties definiert werden.

```

<owl:Class rdf:ID="PATO_0001652">
  <rdfs:subClassOf rdfs:resource="#PATO_0000140"/>
  <rdfs:subClassOf>
    <owl:Class>
      <owl:Restriction>
        <owl:onProperty rdfs:resource="#has_part"/>
        <owl:someValuesFrom
          rdfs:resource="#PATO_0000133"/>
        </owl:Restriction>
      </owl:Class>
    </rdfs:subClassOf>
    <rdfs:label> alignment </rdfs:label>
  </owl:Class>

```

**Beschreibung 4:** Teil-Beschreibung der Qualität *alignment* (PATO:0001652) in OWL/RDF-Syntax

Eine OWL-Klasse entspricht im OBO Flat File Format einem Term, also einer Menge von Individuen. Eingeleitet wird deren Beschreibung durch das `<owl:Class>`-Tag und abgeschlossen durch `</owl:Class>`. Das einleitende Tag besitzt ein Attribut `rdf:ID` mit dem eine Identifikationsnummer für die Klasse definiert wird. Zwischen den beiden `<owl:Class>`-Tags kann eine Sequenz aus `<rdfs:subClassOf>`- und `<owl:equivalentClass>`-Tags stehen, die implizit über *AND* miteinander verbunden sind. Durch Verwendung von `<owl:intersectionOf>` kann diese Verbindung explizit gemacht werden und die Realisierung einer *OR*-Verknüpfung ist durch `<owl:unionOf>` möglich. Das Tag `<rdfs:subClassOf>` zeigt eine Subsumptionsbeziehung (auch: Spezialisierung) und `<owl:equivalentClass>` eine Äquivalenzbeziehung an. Verbindungen zu anderen Klassen können in OWL mit Properties beschrieben werden, wozu das Tag `<owl:onProperty>` Anwendung findet. Die Quantifizierung der Property wird durch die Wahl eines der folgenden Tags geregelt.

<code>&lt;owl:someValuesFrom&gt;</code>	Existenzquantifizierung
<code>&lt;owl:allValuesFrom&gt;</code>	Allquantifizierung
<code>&lt;owl:hasValue&gt;</code>	alle Instanzen müssen den, im <code>&lt;hasValue&gt;</code> -Tag, spezifizierten Wert erfüllen
<code>&lt;owl:Cardinality&gt;</code>	Angabe einer Kardinalität

Die Übersetzung verschiedener Tags des OBO Flat File Formats in die OWL/RDF-Syntax kann der Tabelle 5 entnommen werden. Beschreibung 6 zeigt die zur Beschreibung 4 äquivalente Repräsentation im OBO-Format. Für genauere Angaben zum Mapping sei auf [Mun07], [Mun10a], [Hor07], [Tir11] und [Gol07] verwiesen.

OBO Flat File Format	OWL/RDF-Syntax
<i>id: ONT:01</i>	<code>&lt;owl:Class rdf:ID="ONT_01"/&gt;</code>
<i>name: A</i>	<code>&lt;rdfs:label&gt; A &lt;/rdfs:label&gt;</code>
<i>is_a: ONT:00</i>	<code>&lt;rdfs:subClassOf rdfs:resource="#ONT_00"/&gt;</code>
<i>intersection_of: ONT:02</i> <i>intersection_of: ONT:03</i>	<code>&lt;owl:intersectionOf rdf:parseType="Collection"&gt;   &lt;owl:Class rdf:about="#ONT_02"/&gt;   &lt;owl:Class rdf:about="#ONT_03"/&gt; &lt;/owl:intersectionOf&gt;</code>
<i>union_of: ONT:02</i> <i>union_of: ONT:03</i>	<code>&lt;owl:unionOf rdf:parseType="Collection"&gt;   &lt;owl:Class rdf:about="#ONT_02"/&gt;   &lt;owl:Class rdf:about="#ONT_03"/&gt; &lt;/owl:unionOf&gt;</code>
<i>relationship: r ONT:02</i>	<code>&lt;owl:Restriction&gt;   &lt;owl:onProperty rdfs:resource="#r"/&gt;   &lt;owl:someValuesFrom rdfs:resource="#ONT_02"/&gt; &lt;/owl:Restriction&gt;</code>

**Tabelle 5:** Übersetzung wichtiger Tags des OBO Flat File Formats in die OWL/RDF-Syntax

```
[Term]
id: PATO:0001652
name: alignment
is_a: PATO:0000140
relationship: has_part PATO:0000133
```

**Beschreibung 6:** Übersetzung von Beschreibung 4 in das OBO Flat File Format

Das DL-Fragment (auch: Beschreibungslogik-Fragment) von OWL 2.0 ist zur Beschreibungslogik SROIQ äquivalent, sodass Klassen, Instanzen und Properties auch formal in deren Syntax repräsentiert werden können. In Tabelle 7 ist ein Ausschnitt der, für diese Arbeit, wichtigsten Konstruktoren gegeben und Tabelle 8 enthält die Übersetzung wichtiger Bestandteile der OWL/RDF-Syntax in die Syntax von SROIQ. Die Beschreibung 9 enthält eine zur Beschreibung 4 äquivalente Repräsentation in der formalen SROIQ-Syntax.

Konstruktor	Syntax	Semantik
atomares Konzept	$A$	$A^I \subseteq \Delta^I$
atomare Rolle	$r$	$r^I \subseteq \Delta^I \times \Delta^I$
Äquivalenz	$C \equiv D$	$C^I = D^I$
Subsumption	$C \sqsubseteq D$	$C^I \subseteq D^I$
Konjunktion	$C \sqcap D$	$C^I \cap D^I$
Existenz-Restriktion	$\exists r.C$	$\{x \mid \exists y. \langle x, y \rangle \in r^I \wedge y \in C^I\}$

**Tabelle 7:** Semantik der, für diese Arbeit wichtigen, DL-Konstrukturen bei gegebener Interpretation  $I = (\Delta^I, \cdot^I)$

OWL/RDF-Syntax	Syntax von SROIQ
<code>&lt;owl:Class rdf:ID="A"/&gt;</code>	$A$
<code>&lt;owl:Class rdf:ID="C"&gt; &lt;owl:equivalentClass rdfs:resource="#D"/&gt; &lt;/owl:Class&gt;</code>	$C \equiv D$
<code>&lt;owl:Class rdf:ID="C"&gt; &lt;rdf:subClassOf rdfs:resource="#D"/&gt; &lt;/owl:Class&gt;</code>	$C \sqsubseteq D$
<code>&lt;owl:intersectionOf rdf:parseType="Collection"&gt; &lt;owl:Class rdf:about="#C"/&gt; &lt;owl:Class rdf:about="#D"/&gt; &lt;/owl:intersectionOf&gt;</code>	$C \sqcap D$
<code>&lt;owl:Restriction&gt; &lt;owl:onProperty rdf:resource="#r"/&gt; &lt;owl:someValuesFrom rdf:resource="#C"/&gt; &lt;/owl:Restriction&gt;</code>	$\exists r.C$

**Tabelle 8:** Übersetzung der, für diese Arbeit wichtigen, Bestandteile der OWL/RDF-Syntax in die formale Syntax der Beschreibungslogik SROIQ

$PATO\_0001652 \sqsubseteq PATO\_0000140 \sqcap \exists has\_part.PATO\_0000133$

**Beschreibung 9:** formale DL-Repräsentation von Beschreibung 4

## 2.2 EQ-Modell

Zur formalen Beschreibung einer phänotypischen Charakteristik (auch: Phän oder Phänotyp) wird häufig das Entity-Quality-Modell (Abk.: EQ-Modell; siehe [Mun10b] und [Was09]) herangezogen. Jede, dem Modell entsprechende, Beschreibung (auch: EQ-Beschreibung) besteht aus „[...] *Q*, the type of quality (characteristic) that the genotype affects; *E*, the type of entity that bears the quality; *E2*, an additional optional entity type, for relational qualities; *M*, a modifier.“ [Mun10b, S. 3]. Dieser Aufbau folgt dem Genus-Differentia-Muster, bei dem alle beschreibenden Bestandteile sich auf einen Genus beziehen. In diesem Fall bildet *Q* den Genus und *E*, *E2* und *M* sind die differenzierenden Bestandteile, die in Beziehung zu *Q* stehen. Die Qualität *Q* entspricht einem Konzept aus PATO und die Entitäten *E* und *E2* stammen aus spezies-spezifischen Ontologien, wie MA oder ChEBI. Der Modifier *M* ist, entgegen der Bezeichnung, eigentlich auch nur eine Qualität und stammt daher auch aus PATO.

Das EQ-Modell beschreibt nicht nur den Aufbau einer Phänotyp-Beschreibung, sondern spezifiziert auch eine Syntax, in der diese Beschreibungen dargestellt werden können. Diese ist sehr einfach, da auf die Bezeichner *Q*, *E*, *E2* und *M* lediglich ein Gleichheitszeichen und das passende Konzept aus einer entsprechenden Ontologie folgen. Ein Schema der Syntax lässt sich Schema 10 und Beispiele lassen sich Beschreibung 11 und 12 entnehmen.

$$\begin{aligned} \underline{Q} &= \langle Q \rangle \\ E &= \langle E \rangle \\ E2 &= \langle E2 \rangle \\ M &= \langle M \rangle \end{aligned}$$

**Schema 10:** Schema für Beschreibungen in der EQ-Syntax

$$\begin{aligned} \underline{Q} &= \textit{brown} \\ E &= \textit{eye} \end{aligned}$$

**Beschreibung 11:** Beschreibung des Phänotyps „braunes Auge“ in EQ-Syntax

*Q = concentration*  
*E = urine*  
*E2 = calcium*  
*M = abnormal*

**Beschreibung 12:** Beschreibung des Phänotyps „abnormale Calciumkonzentration im Urin“ in EQ-Syntax

Die EQ-Syntax wird jedoch nur in wissenschaftlichen Arbeiten verwendet und besitzt keine praktische Relevanz. Stattdessen werden EQ-Beschreibungen häufig im OBO Flat File Format oder in einer OWL-Syntax dargestellt. Stellvertretend soll hier die Übersetzung der EQ-Syntax in das OBO Flat File Format dargelegt werden.

Ein Phänotyp entspricht dem Schnitt (auch: Konjunktion) über die Qualität *Q* und den Qualitäten, die in *E* inhärieren, mit *E2* verbunden sind und durch *M* qualifiziert werden. Der Schnitt über die eben genannten Konzepte kann über das Tag *intersection\_of* realisiert werden. Die Inhärenz von Qualitäten in *E* kann durch die *inheres\_in*-Relation, die Beziehung zu *E2* mit *towards* und die Qualifizierungsbeziehung durch *qualifier* beschrieben werden. Somit ergibt sich ein Schema, wie in Schema 13 dargestellt.

*intersection\_of: <Q>*  
*intersection\_of: inheres\_in <E>*  
*intersection\_of: towards <E2>*  
*intersection\_of: qualifier <M>*

**Schema 13:** Schema zur Beschreibung von Phänotypen im OBO Flat File Format

### 2.3 Cross-Products

Mit dem eben beschriebenen EQ-Modell lassen sich Phänotypen durch Komposition verschiedener Konzepte beschreiben. Dieses Vorgehen wird post-koordinierte Beschreibung genannt, da der Phänotyp (genauer: seine Beschreibung) erst bei Bedarf entsteht. Neben dieser Variante, existiert jedoch auch die Möglichkeit den Phänotyp bereits vor Bedarf, also prä-koordiniert, zu beschreiben. Entsprechend sind diese Phänotypen in speziellen Ontologien, wie der Mammalian Phenotype Ontology (Abk.: MP; siehe [Smi11]) oder der Human Phenotype Ontology (Abk.: HP), definiert. Ein

Beispiel einer solchen Beschreibung lässt sich Beschreibung 14 entnehmen. Dort lässt sich auch erkennen, dass der Phänotyp lediglich informell beschrieben ist, was eine maschinelle Verarbeitung prä-koordinierter Phänotypen stark einschränkt. Daher schlug Mungall in [Mun10b] eine Sammlung von Mappings zwischen prä- und post-koordinierten Phänotypen vor, was er als Cross-Products (dt.: Kreuzprodukte) bezeichnet.

*id: MP:0008809*  
*name: increased spleen iron level*  
*def: „increase in the amount of iron present in the spleen tissue“*  
*is\_a: MP:0008739 ! abnormal spleen iron level*

**Beschreibung 14:** post-koordinierte Beschreibung des Phänotyps „erhöhte Eisenkonzentration in der Milz“ (Konzept in MP)

„The descriptions are drawn from the cross-product of two sets of classes: the set of PATO classes and the set of classes from other OBO ontologies. For example, MP-XP is a collection of mappings between individual MP classes and their corresponding EQ descriptions.“ [Mun10b, S. 5]. Formal entspricht ein Cross-Product also der Äquivalentsetzung eines prä-koordinierten Phänotyps und seiner EQ-Beschreibung (siehe Beschreibung 15). Dies wiederum vereint die Vorteile beider Ansätze. Die Tatsache, dass es sich weiterhin um prä-koordinierte Phänotypen handelt, beschränkt den Gebrauch auf die, in ihrer Anzahl limitierten, Konzepte einer Ontologie (z. B. MP), sodass die Erzeugung sinnloser oder inkonsistenter Beschreibungen vermieden wird. Andererseits sind diese Phänotypen nun auch formal beschrieben, weshalb sie eine gewisse Semantik besitzen und diese dazu verwendet werden kann, Anfragen an beispielsweise biologische Datenbanken besser lösen zu können.

*id: MP:0008809 ! increased spleen iron level*  
*intersection\_of: PATO:0001162 ! increased concentration*  
*intersection\_of: towards CHEBI:18248 ! iron*  
*intersection\_of: inheres\_in MA:0000141 ! spleen*

**Beschreibung 15:** Cross-Product (Konzept in MP-XP) aus Beschreibung 8 und deren formale Definition



### 3. Verwendung von Phänotypen in biologischen Datenbanken

#### 3.1 Annotation

Unter Annotation versteht man im Allgemeinen den „[...] *act or process of furnishing critical commentary or explanatory notes.*“ [Hou09]. Diese Anmerkungen haben bezüglich einer Informationsquelle lediglich ergänzenden Charakter und sollen daher den strukturellen Aufbau dieser nicht stören (siehe [Wik11a]). Beispiele hierfür sind Angaben über Inhaltsstoffe, die separiert eine Speisekarte ergänzen sollen oder aber auch spezielle Sprachelemente einer Programmiersprache, um Metadaten zu beschreiben (z. B. `@Override` in Java).

Ein weiteres Beispiel ist die Annotation diverser Eigenschaften an elementare Bausteine, wie Allele oder Proteine, in biologischen Datenbanken. Neben den genetischen Informationen, den Genotypen, sind vor allem phänotypische Charakteristika interessant, die an den Genotyp annotiert werden (siehe Abbildung 16). Dabei ist ein Vermerk eines Phänotyps so zu verstehen, dass dieser bei Existenz des entsprechenden elementaren biologischen Bausteins ebenfalls auftritt. So hat beispielsweise die Existenz des Allels `Add2tm1Lp` (MGI:2149065; siehe [The11]) u. a. den Phänotyp „microcytosis“ zur Folge (siehe Abbildung 16). Auf der Grundlage solcher Phänotyp-Annotationen können über Datenbankabfragen defekte biologische Bausteine detektiert werden und so Krankheiten oder sonstige Defekte automatisiert diagnostiziert werden. Des Weiteren treibt die phänotypische Annotation die Dokumentation von Krankheitsbildern (ein Teil des Phänotyps eines Organismus) voran, da Wissenschaftler hierdurch einen einfacheren Zugang zum aktuellen Erkenntnisstand erhalten.

Nomenclature	Symbol: <b>Add2<sup>tm1Llp</sup></b>	Red blood cell and histological spleen comparison of Add2 <sup>tm1Llp</sup> /Add2 <sup>tm1Llp</sup> and +/- wildtype mouse. Show the 1 image(s) involving this allele.
	Name: adducin 2 (beta); targeted mutation 1, Luanne L Peters MGI ID: MGI:2149065 Synonyms: band 3 <sup>+</sup> , beta-adducin <sup>+</sup> Gene: Add2 Location: Chr6:86028078-86069549 bp, + strand Genetic Position: Chr6, 37.55 cM	
Mutation origin	Germline Transmission: Earliest citation of germline transmission: J:71029 Parent Cell Line: J1 (ES Cell) Strain of Origin: 129S4/SvJae	<b>Name des Allels</b>
Mutation description	Allele Type: Targeted (knock-out) Mutations: Insertion, Intragenic deletion A neomycin resistance cassette replaced a genomic fragment containing exons 9 to 13. Western blot analysis did not detect protein in RBC ghosts from homozygous mutant mice. Northern blot analysis using a probe for exons 9 to 12 did not detect transcript in homozygous mutant brain, spleen, or kidney. Northern blots using a neoR gene probe indicated that in homozygous mutant mice, transcripts are produced from antisense neoR sequence, and in spleen membranes the 4kb chimeric transcript is translated into a 55kDa protein that is detectable by Western blot. RT-PCR analysis detected a transcript that is a direct splice from exon 8 to exon 14, and which in homozygous mutant brain and spleen does not produce a polypeptide that is detectable by Western blot analysis. (J:71029)	
Find Mice (IMSR)	Mouse strains and cell lines available from the International Mouse Strain Resource (IMSR) Carrying this Mutation: Mouse Strains: 1 strain available Cell Lines: 0 lines available Carrying any Add2 Mutation: 3 strains or lines available	
Phenotype summary	Phenotype Summary by Mammalian Phenotype terms Key: hm homozygous ht heterozygous (show or hide all annotated terms) Genotypes are listed in the next section. Affected Systems Genotypes: hm1 hm2 cx3 behavior/neurological abnormal contextual conditioning behavior abnormal cued conditioning behavior abnormal spatial learning hematopoietic system abnormal erythrocyte morphology decreased hematocrit increased mean corpuscular hemoglobin concentration decreased mean corpuscular volume anemia anisopoikilocytosis microcytosis spherocytosis reticulocytosis	<b>bedingte Phänotypen</b>

Abbildung 16: Ausschnitt aus der Definition des Allels Add2<sup>tm1Llp</sup> (MGI:2149065) mit seinen phänotypischen Annotationen

### 3.2 Grundlegende Anfragen

An eine biologische Datenbank sind u. a. die nachfolgenden vier Anfragen bezüglich Phänotypen interessant, weshalb alle später vorgestellten Lösungsansätze diese unterstützen müssen.

A<sub>1</sub>: Generiere eine Ergebnismenge, die alle biologischen Bausteine enthält, die u. a. das übergebene Phän auslösen.

A<sub>2</sub>: Generiere eine Ergebnismenge, die alle biologischen Bausteine enthält, die Phäne auslösen, welche in irgendeiner Art über eine explizite Subsumptionsbeziehung

zum übergebenen Phän stehen.

A<sub>3</sub>: Generiere eine Ergebnismenge, die alle biologischen Bausteine enthält, die generellere Phäne, als den übergebenen auslösen.

A<sub>4</sub>: Generiere eine Ergebnismenge, die alle biologischen Bausteine enthält, die mindestens die, in der Anfrage, konjunctierten Phäne enthalten.

Es dürfte ersichtlich sein, dass A<sub>1</sub> die restriktivste Anfrage ist und A<sub>3</sub> die generellste, sodass die Ergebnismengen der ersten drei Anfragen in dem, in Formel 17, dargestellten Zusammenhang stehen. Anfrage A<sub>4</sub> nimmt eine Sonderrolle ein, da ihr die Konjunktion von Phänen übergeben werden kann und ein direkter Vergleich mit den anderen Anfragen somit nicht möglich ist.

$$\text{Ergebnismenge}(A_1) \subseteq \text{Ergebnismenge}(A_2) \subseteq \text{Ergebnismenge}(A_3)$$

**Formel 17:** Zusammenhang zwischen den ersten drei Anfragen an biologische Datenbanken

Die erste Anfrage A<sub>1</sub> bekommt ein Phän übergeben und sucht in der Datenbank nach biologischen Grundbausteinen (z. B. Allele oder Proteine), die mindestens dieses Phän bedingen. Damit ist sie zugleich die logisch einfachste, da hierbei nicht inferiert oder sonstige Schlüsse gezogen werden müssen. Das Datenbanksystem, in dem die biologischen Bausteine mit ihren Annotationen organisiert sind, muss lediglich ermitteln in welchen Zeilen der Datenbank der übergebene String enthalten ist und die entsprechenden Bausteine in die Ergebnismenge schreiben. Sucht ein Nutzer beispielsweise nach biologischen Grundbausteinen, die das Phän *increased spleen iron level* (MP:0008809) auslösen, so liefert das Datenbanksystem genau die Bausteine zurück, die dieses Phän bedingen.

Die Anfrage A<sub>2</sub> sucht, im Gegensatz zu A<sub>1</sub>, nicht nur nach dem übergebenen Phän, sondern auch nach dessen Spezialisierungen, die in entsprechenden Ontologien definiert sind. Sie erfordert somit das Verfolgen von Pfaden in einer Generalisierungshierarchie. Das bedeutet, dass durch Inferenz zunächst die Unterkonzepte eines übergebenen Konzepts (in diesem Fall Phänotyp) ermittelt werden müssen und anschließend nach all diesen in der Datenbank gesucht wird. Während für prä-koordinierte Phänotypen, wie solche aus MP, diese Hierarchie bereits durch die Ontologie gegeben ist, muss für

post-koordinierte Phäne zunächst temporär eine solche Generalisierungshierarchie erzeugt werden. Diese Anfrage liefert eine Obermenge der Ergebnisse von  $A_1$ , da sie in die Anfragen  $A_{1,P}$ ,  $A_{1,S_1}$ ,  $A_{1,S_2}$ , ...,  $A_{1,S_n}$  aufgeteilt werden kann, wobei  $A_{1,P}$  (entspricht dem eigentlichen  $A_1$ ) der Suche nach biologischen Grundbausteinen, die genau das übergebene Phän  $P$  bedingen und  $A_{1,S_i}$  der Suche nach allen biologischen Grundbausteinen, die genau die Spezialisierung  $S_i$  von  $P$  bedingen, entsprechen. Die Ergebnismenge von  $A_2$  entspricht nun genau der Vereinigung der Ergebnisse von  $A_{1,P}$ ,  $A_{1,S_1}$ ,  $A_{1,S_2}$ , ...,  $A_{1,S_n}$ . So liefert eine Anfrage nach *abnormal spleen iron level* (MP:0008739; entspricht  $P$ ) alle Bausteine, die dieses Phän oder deren Spezialisierungen, wie *increased spleen iron level* (entspricht  $S_1$ ) oder *decreased spleen iron level* (MP:0008808; entspricht  $S_2$ ), bedingen.

Als Erweiterung der zweiten Anfrage werden in  $A_3$  nicht nur nach biologischen Grundbausteinen gesucht, die Phänotypen bedingen, welche explizite Spezialisierungen des übergebenen Phänotyps sind, sondern auch implizite. Hierzu müssen u. a. die Definitionen von Phänotypen ausgewertet werden, um eine eventuelle Subsumptionsbeziehung zum gesuchten Konzept prüfen zu können. Neben den impliziten Unterkonzepten, werden auch explizite in die Ergebnismenge geschrieben, weshalb die Ergebnismenge von  $A_3$  eine Obermenge der Ergebnismenge von  $A_2$  ist. Die Begründung ist ähnlich, wie die obige für  $A_1 \subseteq A_2$ , d. h. auch hier kann  $A_3$  in speziellere Anfragen aufgeteilt werden, sodass die Vereinigung aller Ergebnismengen dem Ergebnis von  $A_3$  entspricht. Ein Beispiel hierfür wäre die Suche nach *increased concentration and inheres\_in some spleen*. Diese Anfrage würde alle biologischen Bausteine zurückliefern, die genau dieses Phän (entspricht Anfrage  $A_1$ ) und deren explizite Spezialisierungen (entspricht Anfrage  $A_2$ ) bedingen (sofern solche existieren). Daneben sucht diese Anfrage aber auch nach Phänen, deren Beschreibung spezieller als *increased concentration and inheres\_in some spleen* ist und ermittelt demnach alle erhöhten Konzentration in der Milz, wie *increased spleen iron level*, woraufhin die bedingenden biologischen Bausteine zurückgeliefert werden.

Der Anfrage  $A_4$  wird die Konjunktion von Phänen übergeben. Sie initiiert die Suche nach Annotationseinträgen, die eines der konjungierten Phäne, deren Spezialisierung oder die Spezialisierung der Konjunktion enthält und ermittelt somit alle biologischen

Grundbausteine, die den Phänotyp oder eine Spezialisierung bedingen. Z. B. könnte eine Anfrage *increased spleen iron level and microcytosis* lauten. In der Datenbank würde daraufhin nach biologischen Bausteinen gesucht werden, die *increased spleen iron level* und *microcytosis* (MP:0002813), deren Spezialisierungen oder Spezialisierungen von *increased spleen iron level and microcytosis* bedingen.

## 4. Verbesserung des herkömmlichen EQ-Modells

So wie das EQ-Modell momentan definiert ist, birgt es einige Schwächen. So sind beispielsweise phänotypische Beschreibungen nach diesem Schema mehrdeutig und zudem noch ontologisch inadäquat. Des weiteren gibt es bei der notwendigen Klassifizierung der beteiligten Entitäten einigen Interpretationsspielraum, der durch das Modell nicht eingeschränkt wird. Die nachfolgenden zwei Kapitel werden diese Probleme genauer aufzeigen und Lösungen hierfür anbieten.

### 4.1 Vertauschungsproblem

#### 4.1.1 Problembeschreibung

Das Vertauschungsproblem (siehe [Loe11]) ist eine Art Ambiguität, die bei der Aggregation von post-koordinierten Phänen in Phänotypen (gemeint sind deren Beschreibungen) entsteht. Eine Phänotyp-Beschreibung entsteht durch den Schnitt (auch: Konjunktion) mehrerer phänotypischer Charakteristika. Beschreibung 18 definiert einen solchen Phänotyp, der durch das Allel *Add2<sup>tm1Llp</sup>* (MGI:2149065) bedingt wird. Es dürfte ersichtlich sein, dass dessen Bestandteile *IncreasedSpleenIronLevel* (MP:0008809) und *Microcytosis* (MP:0002813) prä-koordiniert sind. Des weiteren ist erkennbar, dass hier keine Mehrdeutigkeit vorliegt. Das Individuum mit dem Phänotyp *PartialPhenotypeOfAdd2tm1Llp* leidet sowohl an einer erhöhten Eisen-Konzentration in der Milz als auch an einer verringerten Größe der Erythrozyten. Werden jetzt die Konzepte *IncreasedSpleenIronLevel* und *Microcytosis* durch ihre formalen post-koordinierten Beschreibungen ersetzt, so entsteht die Beschreibung 19. Auf den ersten Blick sagt diese dasselbe aus wie Beschreibung 18. Tatsächlich sind beide Beschreibungen jedoch nicht äquivalent zueinander. Der Grund hierfür ist die Schnittbildung sowohl über die Phäne zur Bildung des Phänotyps als auch über die Qualitäten (gemeint sind die Qualität  $Q$ , die Qualitäten, die in  $E$  inhärieren oder mit  $E2$

verbunden sind und der Modifier  $M$ ) zur Bildung von Phänen. Der  $\sqcap$ -Operator ist kommutativ und assoziativ, wodurch  $(A \sqcap B) \sqcap C \equiv (A \sqcap C) \sqcap B$  gilt. Dementsprechend können Konzepte, die Ziel derselben Rolle sind, innerhalb einer Phänotyp-Beschreibung beinahe beliebig vertauscht werden, ohne dass diese invalide wird. Beschreibung 20 zeigt neben der intendierten eine weitere mögliche Bedeutung des Phänotyps *PartialPhenotypeOfAdd2tm1Llp*. Aus einer erhöhten Eisen-Konzentration in der Milz und einer verringerten Größe der Erythrozyten können durch äquivalente Umformungen eine erhöhte Eisen-Konzentration in den Erythrozyten und eine verringerte Größe der Milz entstehen.

$$\textit{PartialPhenotypeOfAdd2tm1Llp} \equiv \textit{IncreasedSpleenIronLevel} \sqcap \textit{Microcytosis}$$

**Beschreibung 18:** partielle Beschreibung des vom Allel  $\textit{Add2}^{\textit{tm1Llp}}$  bedingten Phänotyps auf Grundlage prä-koordinierter Phäne

$$\begin{aligned} \textit{PartialPhenotypeOfAdd2tm1Llp} \\ \equiv (\textit{IncreasedConcentration} \sqcap \exists \textit{inheres\_in.Spleen} \sqcap \exists \textit{towards.Iron}) \sqcap \\ (\textit{DecreasedSize} \sqcap \exists \textit{inheres\_in.Erythrocyte}) \end{aligned}$$

**Beschreibung 19:** partielle Beschreibung des vom Allel  $\textit{Add2}^{\textit{tm1Llp}}$  bedingten Phänotyps auf Grundlage post-koordinierter Phäne

$$\begin{aligned} \textit{PartialPhenotypeOfAdd2tm1Llp} \\ \equiv (\textit{IncreasedConcentration} \sqcap \exists \textit{inheres\_in.Spleen} \sqcap \exists \textit{towards.Iron}) \sqcap \\ (\textit{DecreasedSize} \sqcap \exists \textit{inheres\_in.Erythrocyte}) \\ \equiv (\textit{IncreasedConcentration} \sqcap \exists \textit{inheres\_in.Erythrocyte} \sqcap \exists \textit{towards.Iron}) \sqcap \\ (\textit{DecreasedSize} \sqcap \exists \textit{inheres\_in.Spleen}) \end{aligned}$$

**Beschreibung 20:** Äquivalenz zwischen zwei eigentlich völlig verschiedenen Phänotypen

Auch wenn an dieser Stelle der Eindruck entsteht, dass dieses Problem künstlich konstruiert sei, hat es starke praktische Relevanz. Ein logisch denkender Mensch wird erkennen, dass eine Vertauschung den Sinn des Phänotyps zerstören würde, nicht so ein logischer Beweiser. Dieser sieht nur die Syntax und kennt nicht die Intention des Autors, die dahinter steckt. Wann immer es eine äquivalente Umformung für nötig hält, wird das Programm sie vornehmen. Die Folge sind falsche Schlüsse und fehlerhafte

Ergebnismengen. Dies ist natürlich ein fundamentaler Fehler in der Modellierung solcher Phänotypen. Biologische Datenbanken, wie Mouse Genome Informatics (Abk.: MGI) oder Zebrafisch Model Organism Database (Abk.: ZFIN), wurden mit dem Ziel eindeutiger Repräsentation von Wissen entworfen. Lässt man jedoch einen Spielraum bei der Interpretation von Phänotypen, so nutzt der gesamte Aufwand, der in die Entwicklung dieser Datenbanken investiert wurde, nichts. Daher wird im nächsten Abschnitt ein Lösungsansatz für das Problem dargestellt.

#### 4.1.2 Lösungsansatz

Wie bereits beschrieben, resultiert das Vertauschungsproblem aus der Kommutativität und Assoziativität des Schnitts (auch: Konjunktion) und der Aggregation von Phänen in Phänotypen. Um nun eine Vertauschung der verschiedenen Konzepte innerhalb einer Phänotyp-Beschreibung und die damit verbundene Ambiguität zu verhindern, müssen die Qualitäten in der Beschreibung des Phäns gebunden werden zu dem sie ursprünglich gehören. Dies kann mit einer Rolle *has\_phene* (siehe Beschreibung 21; es handelt sich dabei nicht um die gleiche Rolle wie in [Hoe11]) realisiert werden, deren Definitionsbereich durch die Menge aller Lebewesen und deren Wertebereich durch das Konzept *quality* (PATO:0000001) gegeben ist. Die Beschreibung eines Phäns kann so in der Rolle gebunden werden, wodurch ein neues anonymes Konzept entsteht. Dieses wird im weiteren Verlauf als atomarer Phänotyp bezeichnet und enthält (auf Instanzebene) genau die Lebewesen, in denen dieser atomare Phänotyp inhäriert. Durch die Verbindung von mindestens zwei atomaren Phänotypen über den Schnitt entsteht, entgegen der biologischen Definition (ein Phänotyp entsteht durch Vereinigung mehrerer Phäne), ein komplexer Phänotyp, der wiederum die Lebewesen enthält (auf Instanzebene) in denen er inhäriert.

Dadurch, dass die Beschreibungen der Phäne jeweils in einer *has\_phene*-Rolle gebunden sind, ist lediglich die Vertauschung ganzer atomarer Phänotypen untereinander oder die Vertauschung von Qualitäten innerhalb der Beschreibung eines Phäns möglich (siehe Beschreibung 22). Die problematische Vertauschung von Konzepten zwischen Phänen ist hingegen nicht mehr möglich, da zwar (1) das



Herausziehen der, in einer Rolle gebundenen, Konjunktion zweier Konzepte legitim ist, (2) deren erneute Hereinnahme jedoch nicht (siehe Formel 23). Aussage (1) kann bewiesen werden, indem man zeigt, dass  $(\exists r.(A \sqcap B))^I \subseteq (\exists r.A \sqcap \exists r.B)^I$  gilt. Hierzu wird ein  $x \in (A \sqcap B)^I$  fixiert, für das  $(y,x) \in r^I$  mit  $y \in (\exists r.(A \sqcap B))^I$  gilt. Da  $x$  im Schnitt von  $A^I$  und  $B^I$  enthalten ist, muss  $x \in A^I$  und  $x \in B^I$  gelten, sodass  $y \in (\exists r.A)^I$  und  $y \in (\exists r.B)^I$  gilt. Daraus folgt, dass  $y$  im Schnitt  $(\exists r.A \sqcap \exists r.B)^I$  enthalten sein muss und Aussage (1) damit bewiesen ist. Für Aussage (2) muss gezeigt werden, dass  $(\exists r.A \sqcap \exists r.B)^I \not\subseteq (\exists r.(A \sqcap B))^I$  gilt. Dies kann durch ein Gegenbeispiel erfolgen. Betrachtet man die Interpretation  $I$  mit  $A^I = \{a,b\}$ ,  $B^I = \{b,c\}$  und  $r^I = \{(d,b), (e,a), (e,c)\}$ , so wird ersichtlich, dass  $(\exists r.A \sqcap \exists r.B)^I = \{d,e\}$ , jedoch  $(\exists r.(A \sqcap B))^I = \{d\}$  gilt, sodass  $(\exists r.A \sqcap \exists r.B)^I \not\subseteq (\exists r.(A \sqcap B))^I$  gelten muss und damit auch Aussage (2) bewiesen ist.

Neben der Disambiguierung phänotypischer Beschreibungen macht die Einführung der *has\_phene*-Relation die Beschreibung zugleich ontologisch adäquater. Im herkömmlichen Beschreibungsmodell (EQ-Modell) wurde der Schnitt über disjunkte Phäne gebildet, der natürlich leer ist. So kann beispielsweise keine Instanz von *IncreasedConcentration* (PATO:0001162) gleichzeitig Instanz von *DecreasedSize* (PATO:0000587) sein, weshalb der Phänotyp *PartialPhenotypeOfAdd2tm1Llp* aus Beschreibung 18 selbst keine Instanzen besitzt. Adäquat wäre hier die Schnittbildung über atomare Phänotypen, also den Mengen von Lebewesen, in denen *IncreasedConcentration* und *DecreasedSize* inhärieren. Genau das bewirkt die *has\_phene*-Rolle.

$$\begin{aligned}
 & \textit{PartialPhenotypeOfAdd2tm1Llp} \\
 \equiv & \exists \textit{has\_phene}. ( \textit{IncreasedConcentration} \sqcap \\
 & \quad \exists \textit{inheres\_in.Spleen} \sqcap \\
 & \quad \exists \textit{towards.Iron} ) \sqcap \\
 & \exists \textit{has\_phene}. ( \textit{DecreasedSize} \sqcap \\
 & \quad \exists \textit{inheres\_in.Erythrocyte} )
 \end{aligned}$$

**Beschreibung 21:** jedes Phän wird über die Rolle *has\_phene* mit dem Organismus verbunden, in dem das Phän inhäriert

$$\begin{aligned}
& \text{PartialPhenotypeOfAdd2tmLlp} \\
& \equiv \exists \text{has\_phene.} ( \text{IncreasedConcentration} \sqcap \\
& \quad \exists \text{inheres\_in.Spleen} \sqcap \\
& \quad \exists \text{towards.Iron} ) \sqcap \\
& \quad \exists \text{has\_phene.} ( \text{DecreasedSize} \sqcap \\
& \quad \quad \exists \text{inheres\_in.Erythrocyte} ) \\
& \equiv \exists \text{has\_phene.} ( \text{DecreasedSize} \sqcap \\
& \quad \quad \exists \text{inheres\_in.Erythrocyte} ) \sqcap \\
& \quad \exists \text{has\_phene.} ( \text{IncreasedConcentration} \sqcap \\
& \quad \quad \exists \text{inheres\_in.Spleen} \sqcap \\
& \quad \quad \exists \text{towards.Iron} ) \\
& \equiv \exists \text{has\_phene.} ( \exists \text{towards.Iron} \sqcap \\
& \quad \quad \text{IncreasedConcentration} \sqcap \\
& \quad \quad \exists \text{inheres\_in.Spleen} ) \sqcap \\
& \quad \exists \text{has\_phene.} ( \exists \text{inheres\_in.Erythrocyte} \sqcap \\
& \quad \quad \text{DecreasedSize} ) \\
& \neq \exists \text{has\_phene.} ( \text{IncreasedConcentration} \sqcap \\
& \quad \quad \exists \text{inheres\_in.Erythrocyte} \sqcap \\
& \quad \quad \exists \text{towards.Iron} ) \sqcap \\
& \quad \exists \text{has\_phene.} ( \text{DecreasedSize} \sqcap \\
& \quad \quad \exists \text{inheres\_in.Spleen} )
\end{aligned}$$

**Beschreibung 22:** durch die Einführung der Rolle *has\_phene* ist die Vertauschung atomarer Phänotypen untereinander und die Vertauschung von Qualitäten in der EQ-Beschreibung möglich, die Vertauschung von Konzepten zwischen Phänen jedoch nicht

- (1)  $\exists r.(A \sqcap B)$  impliziert  $\exists r.A \sqcap \exists r.B$
- (2)  $\exists r.A \sqcap \exists r.B$  impliziert nicht  $\exists r.(A \sqcap B)$

**Formel 23:** das Herausziehen der, in einer Rolle *r* gebunden, Konjunktion zweier Konzepte *A* und *B* ist möglich, deren Hereinnahme jedoch nicht

## 4.2 Kontextinsensitivität von *inheres\_in* und *towards*

### 4.2.1 Problembeschreibung

Phänotypen werden im EQ-Modell durch Verbindung der Entität *E* und der Qualität *Q*, die in *E* inhäriert, via *inheres\_in*-Relation gebildet. Handelt es sich bei *Q* um eine relationale Qualität, so wird zusätzlich die Entität *E2* über die *towards*-Relation mit *Q* verbunden (zusätzlich kann ein Modifier *M* über *qualifier* mit *Q* verbunden werden, was für das nachfolgend beschriebene Problem jedoch unkritisch ist). Die Verwendung von *inheres\_in* und *towards* ist dabei kontextinsensitiv (siehe [Loe11]), d. h. jede Qualität wird immer über dieselben Relationen mit ihren korrelierenden Entitäten verbunden. Dies bringt einige Vorteile mit sich. So sind beispielsweise die Beschreibungen sehr kompakt und sind dank ihrer statischen Struktur (*Q* ist immer über *inheres\_in* und *towards* mit *E* und *E2* verbunden) vollkommen kompatibel zum relationalen Datenbankparadigma, d. h. die Bestandteile einer phänotypischen Beschreibung (also Qualität, Entitäten und Modifier) lassen sich in Spalten einer Datenbanktabelle übertragen.

Die Kontextinsensitivität stellt allerdings auch ein Problem dar. Der Autor einer EQ-Beschreibung hat lediglich die Möglichkeit eine Entität als inhäriert oder als sonstig mit der Qualität verbunden zu klassifizieren. In Beschreibung 24 wurde z. B. *IncreasedConcentration* als in *Spleen* (MA:0000141) inhärierend und zu *Iron* (CHEBI:18248) als sonstig verbunden angenommen. Genauso gut könnte es aber auch andersherum sein. Natürlich ist *Iron* in *Spleen* konzentriert, was für *Spleen* als inhärierte Entität sprechen würde. Andererseits ist es ja *Iron*, das sich in *Spleen* einbringt, sich also dort konzentriert, was für *Iron* als inhäriert spricht. Die Frage nach dem Träger der Eigenschaft *IncreasedConcentration* sollte eher von Philosophen beantwortet werden, statt von Biologen und Biomedizinern, den Autoren solcher Beschreibungen. Beschreibung 25 zeigt einen noch merkwürdigeren Fall. Sie definiert einen erhöhten Abstand des linken zum rechten Auge. Obwohl beide Entitäten *LeftEye* (MA:0002949) und *RightEye* (MA:0002950) die gleiche Rolle im Kontext *IncreasedDistance* (PATO:0000374) ausfüllen und somit auch über die gleiche Relation mit der Qualität

verbunden sein müssten, wird *LeftEye* als die inhärierte und *RightEye* als die sonstig in Verbindung stehende Entität interpretiert.

*IncreasedSpleenIronLevel*

$\equiv \exists hasPhene. (IncreasedConcentration \sqcap \exists inheres\_in.Spleen \sqcap \exists towards.Iron)$

**Beschreibung 24:** *Spleen* wurde als inhärierte Entität klassifiziert, obwohl genauso gut *Iron* der Träger der Eigenschaft sein könnte

*OcularHypertelorism*

$\equiv \exists hasPhene. (IncreasedDistance \sqcap \exists inheres\_in.LeftEye \sqcap \exists towards.RightEye)$

**Beschreibung 25:** trotz, dass *LeftEye* und *RightEye* gleichberechtigt sind, wurde *LeftEye* als inhäriert ausgewählt

Insgesamt lässt sich also beobachten, dass die Klassifizierung der Entitäten innerhalb einer EQ-Beschreibung nicht ganz so einfach ist. Häufig sind beide Varianten vertretbar. Dies kann jedoch zu Inkonsistenzen zwischen phänotypischen Beschreibungen führen. Annotiert beispielsweise ein Autor den Phänotyp *IncreasedSpleenIronLevel*, so wie in Beschreibung 24 gezeigt, an die Allele *X* und *Y* und ein anderer Autor den selben Phänotyp, jedoch mit *Iron* als inhärierte und *Spleen* als sonstig verbundene Entität, an Allel *Z*, so ist die Konsistenz des Datenbestandes nicht mehr vollständig gewahrt. Dies hat zur Folge, dass eine Query wie Anfrage 26 an eine Bio-Datenbank lediglich die Ergebnismenge  $\{X, Y\}$  zurückliefern würde, obwohl das Allel *Z* den gleichen Phänotyp wie *X* und *Y* auslöst. Der Grund hierfür liegt darin, dass der zweite Autor die Entitäten *Spleen* und *Iron* im Phänotyp *IncreasedSpleenIronLevel* anders als der erste Autor klassifizierte und die Query im Stil der ersten Phänotyp-Variante formuliert war. Des Weiteren lässt sich erkennen, dass selbst bei Zugrundeliegen einer einheitlich repräsentierten Datenbank (wenn *IncreasedSpleenIronLevel* für *X*, *Y* und *Z* gleich formuliert wäre) Queries falsche Ergebnismengen generieren können. Dies ist allgemein der Fall, wenn sich die Entitäten-Klassifizierung in Query und Datenbank unterscheidet. Anfrage 26 würde daher eine leere Ergebnismenge zurückliefern, wenn die, von *X* und *Y* bedingten, Phänotypen genauso wie der von Allel *Z* ausgelöste, definiert wären.

```

SELECT *
FROM AllelesWithPhenotypes
WHERE quality="IncreasedConcentration" AND
      inheres_in="Spleen" AND
      towards="Iron"

```

**Anfrage 26:** mögliche SQL-Formulierung zur Ermittlung aller Allele, die den Phänotyp *IncreasedSpleenIronLevel* bedingen

Ein weiteres Problem, das aus der Kontextinsensitivität logisch folgt ist, dass es für einen Leser einer EQ-Beschreibung nicht sofort ersichtlich ist, welche Rolle eine Entität innerhalb des Kontextes des Phänotyps spielt. So kann er dies lediglich aus seinem Weltwissen folgern, also dem Wissen über die Vereinbarkeit einer Entität mit einer bestimmten Rolle. Dieses Vorgehen führt natürlich in fast allen Fällen zu einer eindeutigen Zuordnung der Funktion zu einer Entität. Dennoch wäre es wünschenswert die Rolle einer Entität ohne Recherche möglicher Beziehungen gleich ablesen zu können.

#### 4.2.2 Lösungsansätze

Um die Klassifizierungsproblematik der Entitäten zu umgehen, wäre es zunächst denkbar, statt der Relation *towards* nochmals *inheres\_in* zur Verbindung der Qualität *Q* mit der Entität *E2* zu verwenden. Dieses Vorgehen wäre jedoch nicht zielführend, da zum einen die Verbindung zwischen der Qualität und den beiden Entitäten kontextinsensitiv bleiben würde und zum anderen vorgegeben wird, dass beide Entitäten im Kontext der Qualität die gleiche Funktion ausfüllen. Der zweite Punkt würde dazu führen, dass z. B. Anfragen, wie die nach allen erhöhten Konzentrationen der Erythrozyten (*increased concentration and inheres\_in some erythrocyte*) nicht beantwortet werden könnten, da weder in der Query noch in den Annotationseinträgen der Datenbank feststellbar wäre, ob die Erythrozyten irgendwo erhöht konzentriert wären oder etwas in den selbigen erhöht konzentriert wäre.

Um nun das Problem der kontextinsensitiven Anbindung von Entitäten an die Qualität zu lösen, muss also das EQ-Modell verändert werden. Hierzu werden in den nachfolgenden Abschnitten die folgenden Varianten vorgestellt.

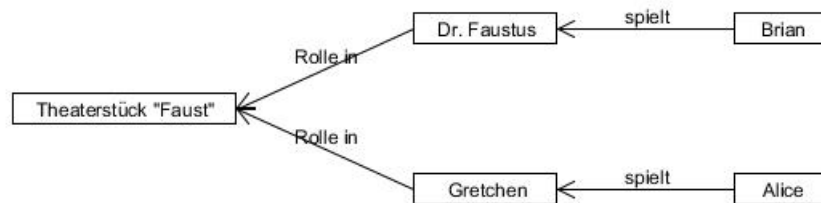
- Verwendung von relationalen GFO-Rollen
- Nutzung kontextsensitiver Properties
- Verwendung eines künstlichen Relators

Diese drei Varianten werden auf einige Konzepte der General Formal Ontology (Abk.: GFO; siehe [Her07] und [Her10]) aufbauen, weshalb zuvor eine kurze Einführung in die Begriffe Fakt, Relation, Relator und relationale Rolle notwendig ist. Bei GFO handelt es sich um eine, von der Research Group Ontologies in Medicine (Abk.: Onto-Med) entwickelte, Top-Level-Ontologie.

Relationen sind „[...] *entities that bind things of the real world together whereas facts are constituted by several related entities considered together with their relation.*“ [Her07, S. 42]. Die Bindung der Entitäten (zu verstehen als Argumente) im Kontext der Relation schafft Fakten. Diese Fakten hängen demnach existentiell von den Argumenten (auch: Relata) und der Relation, die diese Fakten bindet, ab, d. h. der Austausch einer dieser Entitäten würde die Eliminierung dieses und die Existenz eines anderen Fakts bedingen. GFO geht davon aus, dass Relationen, aufgrund der existentiellen Abhängigkeit ihrer konstituierten Fakten von ihren Argumenten, kategorischen Charakter besitzen. Demnach müssen auch Instanzen dieser Kategorie existieren, die in GFO-Terminologie Relatoren (auch: Relationsindividuen) genannt werden. Diese Relatoren bieten eine Art interne Struktur an, die erlaubt festzustellen, welche Rolle (siehe [Loe07] und [Ste00]) ein Argument der Relation im Fakt spielt. Es existieren verschiedene Arten solcher Rollen, von denen die relationale Rolle in diesem Kontext als die am interessantesten erscheint. Es handelt sich dabei um Verbindungsstücke zwischen dem Relator und genau einem Argument. Mit anderen Worten spielt ein Argument eine Rolle innerhalb eines definierten Kontextes (in diesem Fall kann der Relator als Kontext interpretiert werden).

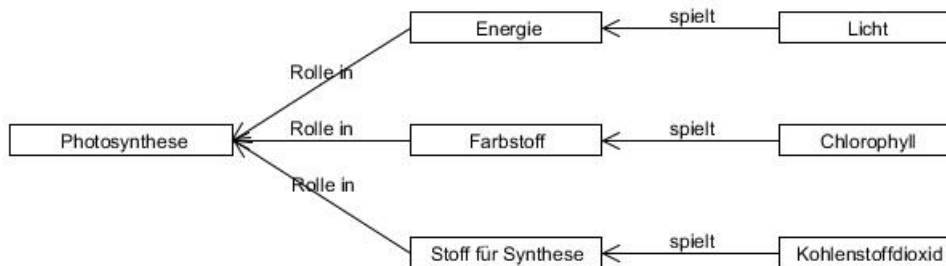
Als Beispiel (siehe Abbildung 27) soll an dieser Stelle das Theaterwesen dienen, aus dem der Begriff „Rolle“ entlehnt wurde. Der Kontext ist durch ein bestimmtes Theaterstück, in diesem Fall Goethe's „Faust“, gegeben. Dieses Stück bietet im Wesentlichen die Rollen „Dr. Faustus“ und „Gretchen“. Nun müssen Darsteller gefunden werden, die diese Rolle ausfüllen können. Am Theater sind die drei Darsteller

Bob, Brian und Alice beschäftigt. Sie treten somit als Kandidaten für Spieler auf. Der Regisseur entscheidet sich für Brian als die Besetzung der Rolle des „Dr. Faustus“ und für Alice als die Besetzung von „Gretchen“. Somit wird der Fakt geschaffen, dass Brian und Alice im Kontext des Theaterstücks „Faust“, über die Rollen „Dr. Faustus“ und „Gretchen“ auftreten.



**Abbildung 27:** Darstellung des Faktens, dass Brian und Alice über die Rollen Dr. Faustus und Gretchen im Theaterstück „Faust“ auftreten

Ein weiteres Beispiel ist in Abbildung 28 gezeigt. Dort ist die Modellierung der allgemeinen Photosynthese zu sehen. Im Kontext dieses Prozesses tritt Licht als Energiequelle, Chlorophyll als Farbstoff und Kohlenstoffdioxid als Stoff für die Synthese auf.



**Abbildung 28:** Darstellung des Faktens, dass Licht, Chlorophyll und Kohlenstoffdioxid über die Rollen Energie, Farbstoff und Stoff für Synthese in der allgemeinen Photosynthese auftreten

#### 4.2.2.1 Relationale GFO-Rollen

In diesem Kapitel wird zur semantischen Untersetzung der Relation zwischen Entität und Qualität das relationale Rollenkonzept der GFO (Roles-as-Classes) genutzt. Allgemein bedeutet dies, dass im Kontext eines Relators Argumente eine Rolle ausfüllen, d. h. relationale Rollen dienen als Verbindungsstück zwischen einem Argument (in diesem Sinn auch Spieler genannt) und dem Relator (in diesem Sinn auch Kontext genannt) in dem es auftritt. Eine Rolle selbst kann nur von einem Spieler eingenommen werden, aber in beliebig vielen Kontexten auftreten. Die Relation zwischen einem Spieler und einer relationalen Rolle ist durch *plays* und zwischen Rolle und Kontext durch *role\_of* gegeben. Die Verwendung dieser beiden Relationen führt zu einer spieler-orientierten Sichtweise, d. h. jede so gebildete Beschreibung definiert ein Konzept, dessen Instanzen Spieler sind (siehe Formel 29).

$$\textit{PlayersThatPlayARoleOfAContext} \equiv \textit{Player} \sqcap \exists \textit{plays}.(\textit{Role} \sqcap \exists \textit{role\_of}.\textit{Context})$$

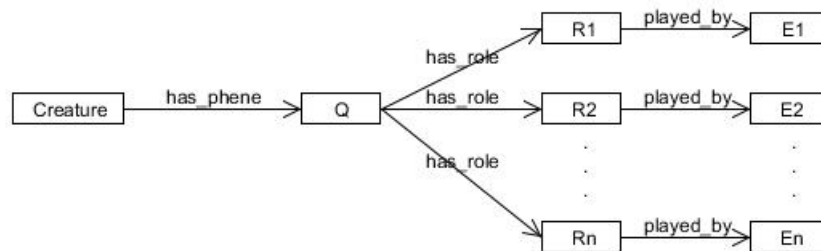
**Formel 29:** eine Beschreibung in der relationale Rollen mit den Relationen *plays* und *role\_of* genutzt werden, definiert ein Konzept, dessen Instanzen Spieler sind

Wird das Rollenkonzept auf phänotypische Beschreibungen adaptiert, so bilden Konzepte aus spezies-spezifischen Ontologien die Spieler und relationale Qualitäten aus PATO die Kontexte. Wie oben beschrieben, ist das allgemeine Rollenkonzept von GFO spieler-orientiert. Im Gegensatz dazu sind phänotypische Beschreibungen jedoch kontext-orientiert, sodass für die Adaption statt der Relationen *plays* und *role\_of* deren Inverse *played\_by* und *has\_role* verwendet werden müssen. Das Schema einer phänotypischen Beschreibung im EQ-Modell lässt sich Abbildung 30 entnehmen. Wie schon angedeutet, lässt sich dieses Schema nur auf relationale Qualitäten anwenden. Dies liegt daran, dass Rollen Bestandteile von Relatoren, den Instanzen von Relationen, sind. Das charakterisierende Merkmal von Relationen ist die Bezugnahme auf mindestens zwei Argumente. Genau das ist bei monadischen Qualitäten nicht so. Diese inhärieren lediglich in einer Entität, ohne das sie eine zweite dazu bräuchten. Nun gibt es zwei Möglichkeiten monadische Qualitäten dennoch einzubeziehen. Die erste liegt in



der Nutzung der Inhärenz als Charakterisierung der Beziehung zwischen Qualität und Entität und die zweite in der Nutzung kontextsensitiver Properties. Bei kontextsensitiven Properties handelt es sich um ein Derivat des eigentlichen relationalen GFO-Rollenkonzepts, welches auf OWL-Properties aufbaut. Es bietet demnach ähnliche Vorteile wie GFO-Rollen und ist der Inhärenz-Charakterisierung vorzuziehen. Weitere Informationen können dem nächsten Kapitel 4.2.2.2 entnommen werden.

Beschreibung 31 zeigt ein, dem Schema entsprechendes, Beispiel. Die relationale Qualität *IncreasedConcentration* (PATO:0001162) stellt den Kontext innerhalb des Phäns *IncreasedSpleenIronLevel* (MP:0008809) dar. Dieser Kontext sieht zwei Rollen vor, die mit *ConcentratingRole* und *ConcentratedRole* bezeichnet sind. Folglich wird *ConcentratingRole* durch ein Stoffgemisch oder Organ, in diesem Fall *Spleen* (MA:0000141) und *ConcentratedRole* durch eine Substanz, hier *Iron* (CHEBI:18248), ausgefüllt.



**Abbildung 30:** Schema phänotypischer Beschreibungen, die auf relationalen GFO-Rollen basieren

$$\begin{aligned}
 & \textit{IncreasedSpleenIronLevel} \\
 \equiv & \exists \textit{has\_phene}. (\textit{IncreasedConcentration} \sqcap \\
 & \quad \exists \textit{has\_role}. (\textit{ConcentratingRole} \sqcap \exists \textit{played\_by}. \textit{Spleen}) \sqcap \\
 & \quad \exists \textit{has\_role}. (\textit{ConcentratedRole} \sqcap \exists \textit{played\_by}. \textit{Iron}))
 \end{aligned}$$

**Beschreibung 31:** Beschreibung des Phänotyps „erhöhte Eisenkonzentration in der Milz“ basierend auf dem Rollenkonzept der GFO

Alle Rollen sind so in einer Hierarchie angeordnet, dass sie das Oberkonzept *BiologicalRole* direkt oder indirekt spezialisieren. *BiologicalRole* ist wiederum ein

Unterkonzept vom Top-Level-Konzept *Relational Role*. Wie in Beschreibung 32 zu sehen ist, werden die Rollen einer Qualität in deren Definition vermerkt, sodass Autoren phänotypischer Beschreibungen alles Wesentliche direkt dort ablesen können. Nicht für jede relationale Qualität sind eigene Rollen zu definieren. Oftmals können sie diese von ihren Oberkonzepten erben. Die Granularität in der Rollen definiert werden, liegt im Ermessen des Modellierers. Sie sollte jedoch stets zum besseren Verständnis zur Verwendung der Qualitäten im Rahmen entsprechender Beschreibungen beitragen. In einigen Testversuchen hat sich jedoch oftmals die Beschreibung des speziellsten Attributs als geeigneter Ort für die Rollen-Definition erwiesen.

```

<owl:Class rdf:about="#PATO_000033">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    concentration of
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0002182"/>
  <rdfs:subClassOf>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <owl:Restriction>
          <owl:onProperty rdf:resource=
            "file:/C:/test/Roles.owl#has_role"/>
          <owl:someValuesFrom rdf:resource=
            "file:/C:/test/Roles.owl#Role_03"/>
        </owl:Restriction>
        <owl:Restriction>
          <owl:onProperty rdf:resource=
            "file:/C:/test/Roles.owl#has_role"/>
          <owl:someValuesFrom rdf:resource=
            "file:/C:/test/Roles.owl#Role_04"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </rdfs:subClassOf>
</owl:Class>

```

**Vermerk der Rollen**

**Beschreibung 32:** Beschreibung der Qualität *concentration of* mit ihren Rollen *Role\_03* (*ConcentratingRole*) und *Role\_04* (*ConcentratedRole*) in OWL/RDF-Syntax

Es lässt sich leicht erkennen, dass die Verwendung von Rollen innerhalb phänotypischer Beschreibungen die, durch die Kontextinsensitivität des herkömmlichen EQ-Modells

bedingten, Probleme korrigiert. So muss sich der Autor einer EQ-Beschreibung nicht mehr der Frage stellen, welche Entität die inhärierte und welche die sonstig verbundene ist. An der Benennung der Rollen kann er eindeutig die Funktion der Entitäten im Kontext einer Qualität erkennen und so Inkonsistenzen auf logischer Ebene verhindern. Auch den Lesern solcher Beschreibungen hilft dies den Sinn eines Phäns besser und schneller zu verstehen. Das Rollenkonzept bietet noch einen weiteren Vorteil. Es ermöglicht die Definition von Qualitäten mit einer höheren Arität als zwei. Mit dem bisherigen EQ-Modell war dies nicht möglich, da Qualitäten nur über *inheres\_in* und *towards* mit Entitäten verbunden werden konnten und eine doppelte Nutzung einer der Relationen zu Ambiguitäten geführt hätte. Die Zahl der Rollen innerhalb eines Kontextes ist hingegen nicht beschränkt, sodass theoretisch unbegrenzt viele Entitäten mit einer Qualität in Beziehung stehen können. Diese Variabilität kann dazu genutzt werden, um beispielsweise Bedingungen oder andere Informationen, wie Messmethoden, in Phäne zu integrieren (siehe Beschreibung 33).

*IncreasedSpleenIronLevel*

$$\equiv \exists has\_phene.(IncreasedConcentration \sqcap$$

$$\quad \exists has\_role.(ConcentratingRole \sqcap \exists played\_by.Spleen) \sqcap$$

$$\quad \exists has\_role.(ConcentratedRole \sqcap \exists played\_by.Iron) \sqcap$$

$$\quad \exists has\_role.(MeasurementMethod \sqcap \exists played\_by.MethodA))$$

**Beschreibung 33:** Beispiel für die Integration der Information über eine Messmethode in eine phänotypische Beschreibung

Die Nutzung des GFO-Rollenkonzepts birgt jedoch im Wesentlichen auch drei Nachteile. Zum einen sind die neuen Beschreibungen, aufgrund des höheren Verschachtelungsgrads, nicht mehr so kompakt. Dies stellt jedoch kein großes Problem dar, da EQ-Beschreibungen meistens mit Hilfe von speziellen Werkzeugen erstellt werden. Zum anderen sind post-koordinierte Phäne nach diesem Modell, im Gegensatz zu Phänen nach dem herkömmlichen EQ-Modell, nicht mit dem relationalen Datenbankparadigma kompatibel, d. h. ihre Bestandteile lassen sich nicht in Spalten einer Tabelle übertragen. Im alten Modell war dies unproblematisch. Die Qualität ließ sich in eine Spalte *quality*, die inhärierte Entität in *inheres\_in* und die sonstig in

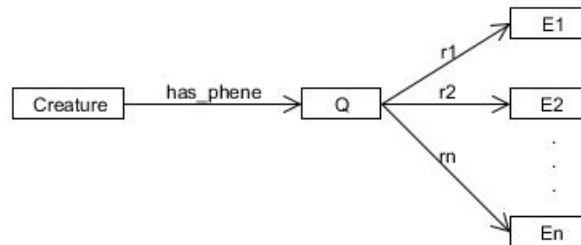
Verbindung stehende in eine Spalte *towards* eintragen. Denkbar wäre eine Adaptierung dieses Vorgehens für das neue Modell. So könnte die Qualität auch in eine Spalte *quality* und die Entität *i* in eine Spalte *entity\_i* eingetragen werden. Dies würde aber dazu führen, dass auf Datenbankebene die Qualität mit den Entitäten kontextinsensitiv verbunden wäre, was wieder zu den oben beschriebenen Problemen führt. Demnach wäre die zu bevorzugende Art post-koordinierte Phäne abzuspeichern, sie als String in eine Spalte einer Tabelle einzutragen und für Anfragen einen Parser darauf anzuwenden. Alternativ können auch prä-koordinierte Phäne annotiert werden, da deren Bedeutung in Cross-Products ermittelt werden kann. Als dritter Nachteil stellt sich die inkompatible Behandlung monadischer und relationaler Qualitäten heraus. Während das GFO-Rollenkonzept auf relationale Qualitäten anwendbar ist, muss für monadische Qualitäten auf kontextsensitive Properties zurückgegriffen werden. Dies könnte bei Autoren phänotypischer Beschreibungen zu Irritationen führen.

#### 4.2.2.2 Kontextsensitive Properties

Neben der Umsetzung des Rollenkonzepts durch relationale GFO-Rollen ist auch die Verwendung kontextsensitiver Properties (auch: Roles-as-Properties) möglich. Dabei handelt es sich lediglich um eine andere syntaktische Ausprägung der GFO-Rollen, d. h. auch hier können Spieler in einem definierten Kontext bestimmte Rollen ausfüllen. Der Unterschied zur ersten Variante besteht in der Realisierung der Rollen als OWL-Properties, statt als Klassen. Des weiteren ist diese Variante sowohl auf relationale als auch auf monadische Qualitäten anwendbar. Ein Schema für EQ-Beschreibungen nach diesem Modell lässt sich Abbildung 34 entnehmen. Dort lässt sich erkennen, dass eine kontextsensitive Property, ähnlich wie *inherits\_in* und *towards*, eine Qualität mit einer, in diesem Kontext, auftretenden Entität verbindet, diesen dabei jedoch berücksichtigt.

Beschreibung 35 zeigt ein, dem Schema entsprechendes, Beispiel für eine phänotypische Beschreibung. Dargestellt ist die Definition des atomaren Phänotyps *IncreasedSpleenIronLevel* (MP:0008809), in den sich *Iron* (CHEBI:18248) als eine konzentrierende Substanz und *Spleen* (MA:0000141) sich als ein konzentrierendes

Organ einbringen. Ausgedrückt wird dies über die OWL-Properties *concentrates* und *is\_concentrated\_in*.



**Abbildung 34:** Schema phänotypischer Beschreibungen, die auf kontextsensitiven Properties basieren

$$\begin{aligned}
 & \textit{IncreasedSpleenIronLevel} \\
 \equiv & \exists \textit{has\_phene}. ( \textit{IncreasedConcentration} \sqcap \\
 & \quad \exists \textit{concentrates.Iron} \sqcap \\
 & \quad \exists \textit{is\_concentrated\_in.Spleen} )
 \end{aligned}$$

**Beschreibung 35:** Beschreibung des Phänotyps „erhöhte Eisenkonzentration in der Milz“ basierend auf kontextsensitiven Properties

Alle kontextsensitiven Properties sind in einer Hierarchie organisiert, deren Wurzel durch *relates\_to* gegeben ist. Jede andere Property subsumiert diese direkt oder indirekt. Die Granularität der Property-Definitionen liegt im Ermessen des Modellierers, d. h. die Entscheidung für welche Ebenen der Qualitäten-Hierarchie neue Properties definiert werden, ist pauschal nicht zu treffen. Hierfür ist der Detailgrad der Qualitäten auf einer Ebene zu verschieden. In einigen Testversuchen hat sich jedoch oft das spezielleste Attribut (z. B. *color* (PATO:0000014)) in der Hierarchie als geeigneter Property-Träger erwiesen. Dennoch existieren viele Ausnahmen, in denen sich ein oder mehrere generellere Attribute (z. B. *concentration of* (PATO:0000033) statt *concentrated* (PATO:0001159)) besser eignen. Der Verweis auf Properties findet sich direkt in der Definition der Qualitäten (siehe Beschreibung 36).

```

<owl:Class rdf:about="#PATO_0000033">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    concentration of
  </rdfs:label>
  <rdfs:subClassOf>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <owl:Restriction>
          <owl:onProperty rdf:resource=
            "#concentrates"/>
          <owl:someValuesFrom rdf:resource=
            "file:/C:/test/Entities.owl#Entity_01"/>
        </owl:Restriction>
        <owl:Restriction>
          <owl:onProperty rdf:resource=
            "#is_concentrated_in"/>
          <owl:someValuesFrom rdf:resource=
            "file:/C:/test/Entities.owl#Entity_01"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#PATO_0002182"/>
</owl:Class>

```

**Vermerk der  
kontextsensitiven  
Properties**

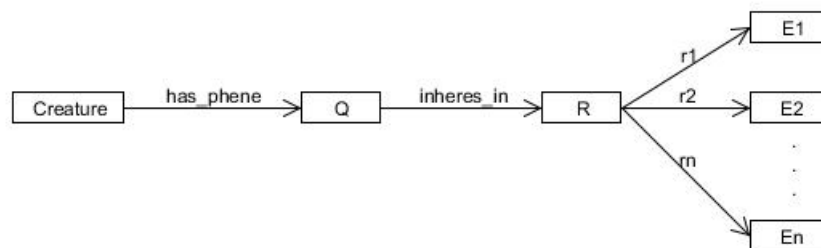
**Beschreibung 36:** Beschreibung der Qualität *concentration of* mit ihren Properties *concentrates* und *is\_concentrated\_in* in OWL/RDF-Syntax

Im Wesentlichen besitzt das Konzept der kontextsensitiven Properties die gleichen Vor- und Nachteile wie die GFO-Rollen-Variante. So löst es das Problem der kontextinsensitiven Charakterisierung von Entitäten im phänotypischen Kontext und gestattet gleichzeitig die Bezugnahme auf mehr als zwei Entitäten innerhalb eines Phäns. Hinzu kommt noch, dass Beschreibungen, die kontextsensitive Properties nutzen, im Gegensatz zu denen der GFO-Rollen-Variante, sehr kompakt sind und sich schematisch für monadische und relationale Qualitäten gleichen. Die begrenzte Kompatibilität zum relationalen Datenbankparadigma bleibt jedoch erhalten.

#### 4.2.2.3 Künstliche Relatoren

In den beiden vorangegangenen Varianten wurden relationale Qualitäten als Relationen aufgefasst, die Entitäten in einem Phän banden. Diese Variante weicht von einem solchen Schema ab und ersetzt die relationale Qualität durch einen künstlichen Relator, der die, aus spezies-spezifischen Ontologien stammenden, Entitäten verbindet. Die Rollen, welche die Entitäten ausfüllen, können z. B. durch kontextsensitive Properties realisiert sein. Die eigentliche relationale Qualität wird nur noch als monadisch betrachtet und inhäriert im künstlichen Relator. Ein Schema für entsprechende Beschreibungen kann Abbildung 37 entnommen werden. Wie schon angedeutet, ist dieses Schema nur auf relationale Qualitäten anwendbar. Monadische Qualitäten sind, wie schon in der ersten Variante, über kontextsensitive Properties mit ihrer Entität verbunden.

Beschreibung 38 zeigt eine, dem Schema entsprechende, Beispiel-Beschreibung des Phäns *IncreasedSpleenIronLevel* (MP:0008809). Der Relator *ConcentrationRelator* verbindet das, sich lösende, *Iron* (CHEBI:18248) mit dem konzentrierenden Organ *Spleen* (MA:0000141) und inhäriert *IncreasedConcentration* (PATO:0001162).



**Abbildung 37:** Schema phänotypischer Beschreibungen, die auf künstlichen Relatoren basieren

$$\begin{aligned}
& \textit{IncreasedSpleenIronLevel} \\
& \equiv \exists \textit{has\_phene}. ( \textit{IncreasedConcentration} \sqcap \\
& \quad \exists \textit{inheres\_in}. ( \textit{ConcentrationRelator} \sqcap \\
& \quad \quad \exists \textit{dissolve.Iron} \sqcap \\
& \quad \quad \exists \textit{solute\_in.Spleen} ) )
\end{aligned}$$

**Beschreibung 38:** Beschreibung des Phänotyps „erhöhte Eisenkonzentration in der Milz“, basierend auf dem künstlichen Relator *ConcentrationRelator*

Jeder künstlich eingeführte Relator erbt direkt oder indirekt vom GFO-Konzept *Relator*. Die Granularität, in der Relatoren definiert werden, liegt auch hier im Ermessen des Modellierers. Als geeignet erscheint eine Definition jedoch an der Stelle, wo in der ersten Variante Rollen zur korrelierenden Qualität (im Fall von *ConcentratingRole* ist die korrelierende Qualität *ConcentrationOf* (PATO:0000033)) eingeführt wurden.

Auch diese Variante löst das Problem der Kontextinsensitivität und ermöglicht zugleich mehr als zwei Entitäten in Beziehung zu setzen. Zudem können Relationen umfassender charakterisiert werden und im Gegensatz zur GFO-Rollen-Variante sind Beschreibungen nach diesem Modell relativ kompakt. Als negativ ist hier zu erachten, dass es ebenfalls keine Unterstützung des relationalen Datenbankparadigmas bietet und monadische und ursprünglich relationale Qualitäten nicht analog behandelt. Des Weiteren muss die Ontologie PATO nicht nur erweitert werden, wie in den vorangegangenen Modellen, sondern vollständig überarbeitet werden. Wie am Anfang des Kapitels erläutert, werden relationale Qualitäten fortan als monadisch betrachtet, wodurch die Verweise zur Term-Teilmenge *relational\_slim* entfernt werden müssen.

#### 4.2.2.4 Vergleich der Varianten

In den vorangegangenen Kapiteln wurden drei Varianten zur Lösung des beschriebenen Problems der kontextinsensitiven Bindung von Entitäten an eine Qualität vorgestellt. Jede dieser Varianten basierte in irgendeiner Weise auf dem Rollenmodell und besaß unterschiedliche Vor- und Nachteile. In der nachfolgenden Tabelle 39 werden diese nochmals kurz dargestellt.



	<b>herkömmliches EQ-Modell</b>	<b>relationale GFO-Rollen</b>	<b>kontextsensitive Properties</b>	<b>künstlicher Relator</b>
<b>Berücksichtigung des Kontextes</b>	nein	ja	ja	ja
<b>maximale Anzahl verbundener Entitäten</b>	2	unbegrenzt	unbegrenzt	unbegrenzt
<b>Kompaktheit der Beschreibung</b>	kompakt	lang	kompakt	relativ kompakt
<b>Möglichkeit zusätzliche Aussagen über die Relation zu treffen</b>	nein	nein	nein	ja
<b>analoge Behandlung monadischer und relationaler Qualitäten</b>	ja	nein	ja	nein
<b>Unterstützung des relationalen Datenbankparadigmas</b>	ja	nein	nein	nein
<b>erfüllt Anforderungen, um alle Anfragen aus Kapitel 3.2 lösen zu können</b>	ja	ja	ja	ja
<b>erforderliche Beschreibungslogik</b>	SH	SHIQ (wegen Nutzung von GFO)	SHIQ (wegen Nutzung von GFO)	SHIQ (wegen Nutzung von GFO)

**Tabelle 39:** Vergleich des herkömmlichen EQ-Modells mit den drei vorgestellten Varianten

## 5. Schlusswort

Die vorangegangenen vier Kapitel haben ein Modell für phänotypische Beschreibungen und deren Verwendung in Cross-Products und Annotationen vorgestellt. Des weiteren wurden zwei signifikante Probleme des Modells aufgezeigt und zugleich gelöst. Das Vertauschungsproblem war für verschiedene Mehrdeutigkeiten innerhalb eines Phänotyps verantwortlich. Es ermöglichte den Austausch von Konzepten über die Grenzen der Phäne hinweg. Mit der Verbindung zwischen Lebewesen und ihren inhärierenden Phänotypen über eine *has\_phene*-Relation und der damit verbundenen Bindung von Konzepten innerhalb einer Phän-Beschreibung konnte dieses Problem jedoch gelöst werden. Für das zweite Problem der kontextinsensitiven Verbindung zwischen Entitäten und einer Qualität wurden drei mögliche Ansätze vorgestellt. Im Prinzip basierten alle auf dem Rollenkonzept der GFO, das vorsieht Entitäten über Rollen mit ihrem Relator zu verbinden.

Obwohl mit der Lösung der zwei Probleme aus Kapitel 4 auch zwei wichtige Fragen beantwortet werden konnten, sind noch längst nicht alle Fragen geklärt. So stellt sich beispielsweise die Frage wie mit Entitäten verfahren werden soll, die laut einer phänotypischen Beschreibung gar nicht existieren sollten (siehe [Hoe07]). In Beschreibung 40 wird das Phän *AbsentSpleen* (MP:0000690) definiert. Die Property *lacks* verbindet darin eine Instanz der Qualität *LacksAllPartsOfType* (PATO:0002000) mit einer Instanz von *Spleen* (MA:0000141). Wenn jedoch eine nicht existierende *Spleen* beschrieben wird, kann somit auch keine Instanz von *Spleen* existieren, die über *lacks* mit *LacksAllPartsOfType* verbunden ist.

$$\begin{aligned} & \textit{AbsentSpleen} \\ \equiv & \exists \textit{has\_phene}. ( \textit{LacksAllPartsOfType} \sqcap \\ & \quad \exists \textit{lacks.Spleen} \sqcap \\ & \quad \exists \textit{jointed\_in.ImmuneSystem} ) \end{aligned}$$

**Beschreibung 40:** obwohl keine Instanz von *Spleen* existiert, wird sie in der Beschreibung des Phäns *AbsentSpleen* über *lacks* mit *LacksAllPartsOfType* verbunden

Ein weiteres Problem kristallisiert sich aus dem Gebrauch von Modifiern heraus. Es erscheint relativ merkwürdig, dass *abnormal* (PATO:0000460) eine Qualität ist, die häufig dazu verwendet wird eine präsente Qualität als von der Norm abweichend zu beschreiben, jedoch beispielsweise keine Qualität *increased* existiert, die *abnormal* spezialisieren würde. Stattdessen findet sich „increased“ in vielen Qualitäten, wie *increased concentration* (PATO:0001162), wieder, die selbst als, im Vergleich zur Norm, erhöht definiert sind ( $\exists$ *increased\_in\_magnitude\_relative\_to.Normal*). Eine Auslagerung von *increased* als eigenständige Qualität würde viel konsistenter wirken. Die Liste noch nicht gelöster Probleme im Zusammenhang phänotypischer Beschreibungen ließe sich noch weiter fortführen. Die Beantwortung all dieser Fragen wird noch viel wissenschaftliche Arbeit auf diesem Gebiet erfordern.

## 6. Literaturverzeichnis

- [Baa02] Baader, Franz; Nutt, Werner: Basic Description Logics. In: The Description Logic Handbook. Cambridge University Press, Cambridge 2002: S. 47 - 100. Verfügbar unter: <http://www.inf.unibz.it/~franconi/dl/course/dlhb/dlhb-02.pdf> (zuletzt abgerufen am 16.06.2011)
- [Bal10] Balhoff, James P.; Dahdul, Waslia M.; Kothari, Cartik R.; Lapp, Hilmar; Lundberg, John G.; Mabee, Paula; Midford, Peter E.; Westerfield, Monte; Vision, Todd J.: Phenex: Ontological Annotation of Phenotypic Diversity. In: PloS ONE. PloS, San Francisco 2010: v.5(5). Verfügbar unter: <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0010500> (zuletzt abgerufen am 16.06.2011)
- [Day06] Day-Richter, John: The OBO Flat File Format Specification, version 1.2. 2006. Verfügbar unter: [http://www.geneontology.org/GO.format.obo-1\\_2.shtml](http://www.geneontology.org/GO.format.obo-1_2.shtml) (zuletzt abgerufen am 15.06.2011)
- [deK11] de Kunder, Maurice: The size of the World Wide Web. 2011. Verfügbar unter: <http://www.worldwidewebsite.com/> (zuletzt abgerufen am 16.06.2011)
- [FraOJ] Franconi, Enrico: Description Logics - Designing Knowledge Bases. o. J.. Verfügbar unter: <http://www.inf.unibz.it/~franconi/dl/course/slides/kbs/kbs-modelling.pdf> (zuletzt abgerufen am 16.06.2011)
- [Gko04] Gkoutos, Georgios V.; Green, Eain C. J.; Mallon, Ann-Marie; Hancock, John M.; Davidson, Duncan: BUILDING MOUSE PHENOTYPE ONTOLOGIES. In: Pacific Symposium on Biocomputing, PubMed, Bethesda 2004: v.9(S. 178 - 189). Verfügbar unter: <http://helix-web.stanford.edu/psb04/gkoutos.pdf> (zuletzt abgerufen am 04.09.2011)
- [Gko05] Gkoutos, Georgios V.; Green, Eain C. J.; Mallon, Ann-Marie; Hancock, John M.; Davidson, Duncan: Using ontologies to describe mouse phenotypes. In: Genome Biology. PubMed Central, Bethesda 2005: v.6(1). Verfügbar unter: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC549069/> (zuletzt abgerufen am 15.06.2011)

- [Gko07] Gkoutos, Georgios V.: Report on status of mouse phenotype ontologies. 2007. Verfügbar unter: [http://www.casimir.org.uk/storyfiles/42.0.wp4\\_gkoutos.pdf](http://www.casimir.org.uk/storyfiles/42.0.wp4_gkoutos.pdf) (zuletzt abgerufen am 15.06.2011)
- [Gko11] Gkoutos, Geogios V.: quality.obo. 2011. Verfügbar unter: <http://pato.googlecode.com/svn/trunk/quality.obo> (zuletzt abgerufen am 19.06.2011)
- [Gol07] Golbreich, Christine; Horrocks, Ian: The OBO to OWL mapping, GO to OWL 1.1!. In: OWLED 2007, Innsbruck 2007. Verfügbar unter: <http://www.comlab.ox.ac.uk/people/ian.horrocks/Publications/download/2007/GoHo07a.pdf> (zuletzt abgerufen am 16.06.2011)
- [Gru09] Gruber, Tom: Ontology. In: Encyclopedia of Database Systems. Springer-Verlag, o. O. 2009. Verfügbar unter: <http://tomgruber.org/writing/ontology-definition-2007.htm> (zuletzt abgerufen am 16.06.2011)
- [Her07] Herre, Heinrich; Heller, Barbara; Burek, Patryk; Hoehndorf, Robert; Loebe, Frank; Michalek, Hannes: General Formal Ontology (GFO) - A Foundational Ontology Integrating Objects and Processes. 2007. Verfügbar unter: <http://www.onto-med.de/Archiv/ontomed2002/en/theories/gfo/part1-drafts/gfo-part1-v1-0-1.pdf> (zuletzt abgefufen am 15.06.2011)
- [Her10] Herre, Heinrich: General Formal Ontology (GFO) - A Foundational Ontology for Conceptual Modelling. 2010. Verfügbar unter: <http://www.onto-med.de/publications/2010/gfo-basic-principles.pdf> (zuletzt abgerufen am 15.06.2011)
- [Hey08] Heyer, Gerhard; Quasthoff, Uwe; Wittig, Thomas: Text Mining: Wissensrohstoff Text - Konzepte, Algorithmen, Ergebnisse. W3L-Verlag, Herdecke 2008: S. 3.
- [Hit08] Hitzler, Pascal; Krötzsch, Markus; Rudolph, Sebastian; Sure, York: Semantic Web. Springer-Verlag, Berlin 2008: S. 33 - 45, S. 123 - 172.
- [Hit09] Hitzler, Pascal; Krötzsch, Markus; Rudolph, Sebastian: OWL 2 Rules (Part 1). 2009. Verfügbar unter: <http://www.semantic-web-book.org/w/images/d/dc/OWL2-Rules-Part-1-ESWC09.pdf> (zuletzt abgerufen am 16.06.2011)
- [Hoe07] Hoehndorf, Robert; Loebe, Frank; Kelso, Janet; Herre, Heinrich: Representing default knowledge in biomedical ontologies: application to the integration of anatomy and phenotype ontologies. In: BMC Bioinformatics. BioMed Central, London 2007: v.8. Verfügbar unter: <http://www.biomedcentral.com/1471-2105/8/377> (zuletzt abgerufen am 15.06.2011)

- [Hoe11] Hoehndorf, Robert; Schofield, Paul N.; Gkoutos, Georgios V.:  
PhenomeNET: a whole-phenome approach to disease gene discovery. In:  
Nucleic Acids Research. Oxford University Press, Oxford 2011: v.39(18).  
Verfügbar unter:  
<http://nar.oxfordjournals.org/content/early/2011/07/06/nar.gkr538.full>  
(zuletzt abgerufen am 13.11.2011)
- [Hor03] Horrocks, Ian; Patel-Schneider, Peter F.; van Harmelen, Frank:  
From SHIQ and RDF to OWL:  
The Making of a Web Ontology Language. In: Journal of Web Semantics.  
Elsevier, Shannon 2003: v.1(1). Verfügbar unter:  
<http://www.websemanticsjournal.org/index.php/ps/article/view/24/22>  
(zuletzt abgerufen am 16.06.2011)
- [Hor07] Horrocks, Ian: OBO Flat File Format Syntax and Semantics and Mapping  
to OWL Web Ontology Language. 2007. Verfügbar unter:  
<http://www.cs.man.ac.uk/~horrocks/obo> (zuletzt abgerufen am  
15.06.2011)
- [Hor09] Horridge, Matthew; Patel-Schneider, Peter F.:  
OWL 2 Web Ontology Language - Manchester Syntax. 2009.  
Verfügbar unter:  
<http://www.w3.org/TR/2009/NOTE-owl2-manchester-syntax-20091027/>  
(zuletzt abgerufen am 16.06.2011)
- [Hou09] Houghton Mifflin Company: annotation. 2009. Verfügbar unter:  
<http://www.thefreedictionary.com/annotation> (zuletzt abgerufen am  
05.09.2011)
- [Jan08] Jansen, Ludger; Smith, Barry: Biomedizinische Ontologie -  
Wissen strukturieren für den Informatik-Einsatz. vdf Hochschulverlag,  
Zürich 2008: S. 15 - 45, S. 155 - 172.
- [KolOJ] Kolb: Beschreibungslogik. o. J.. Verfügbar unter:  
[http://www.ling.uni-potsdam.de/~kolb/textual-entailment/  
Beschreibungslogik.pdf](http://www.ling.uni-potsdam.de/~kolb/textual-entailment/Beschreibungslogik.pdf) (zuletzt abgerufen am 16.06.2011)
- [Lew06] Lewis, Suzanna E.:  
PaTO: Towards a description of organismal Phenotypes. 2006.  
Verfügbar unter: [http://archive.niees.ac.uk/talks/egenomics\\_2006/  
eGenomics2006\\_lewis.ppt](http://archive.niees.ac.uk/talks/egenomics_2006/eGenomics2006_lewis.ppt) (zuletzt abgerufen am 15.06.2011)
- [Loe07] Loebe, Frank: Abstract vs. social roles -  
Towards a general theoretical account of roles. In: Applied Ontology.  
IOS Press, Amsterdam 2007: v.2. Verfügbar unter:  
[http://www.informatik.uni-leipzig.de/~loebe/papers/  
loebe-f-2007-127-a.pdf](http://www.informatik.uni-leipzig.de/~loebe/papers/loebe-f-2007-127-a.pdf) (zuletzt abgerufen am 15.06.2011)

- [Loe11] Loebe, Frank; Stumpf, Frank; Hoehndorf, Robert; Herre, Heinrich: Towards Improving Phenotype Representation in OWL. In: OBML 2011 Workshop Proceedings, IMISE-Reports. Universität Leipzig, Leipzig 2011: v.1(H).
- [Man09] Mandl, Michael: Web Ontology Language: OWL. 2009. Verfügbar unter: [http://www.cis.uni-muenchen.de/~leiss/description-logic-09/mandl\\_OWL.pdf](http://www.cis.uni-muenchen.de/~leiss/description-logic-09/mandl_OWL.pdf) (zuletzt abgerufen am 16.06.2011)
- [Mun06] Mungall, Christopher J.: Some thoughts on PATO. 2006. Verfügbar unter: <http://ontology.buffalo.edu/smith/ppt/FuGo2006/PATO2.ppt> (zuletzt abgerufen am 16.06.2011)
- [Mun07] Mungall, Christopher J.; Gkoutos, Georgios V.; Washington, Nicole; Lewis, Suzanna E.: Representing Phenotypes in OWL. In: OWLED 2007, Innsbruck 2007. Verfügbar unter: [http://www.webont.org/owled/2007/PapersPDF/paper\\_40.pdf](http://www.webont.org/owled/2007/PapersPDF/paper_40.pdf) (zuletzt abgerufen am 15.06.2011)
- [Mun10a] Mungall, Christopher J.; Ireland, Amelia: The OBO Flat File Format Guide, version 1.4. 2010. Verfügbar unter: [http://www.geneontology.org/GO.format.obo-1\\_4.shtml](http://www.geneontology.org/GO.format.obo-1_4.shtml) (zuletzt abgerufen am 15.06.2011)
- [Mun10b] Mungall, Christopher J.; Gkoutos, Georgios V.; Smith, Cynthia L.; Haendel, Melissa A.; Lewis, Suzanna E.; Ashburner, Michael: Integrating phenotype ontologies across multiple species. In: Genome Biology. PubMed Central, Bethesda 2010: v.11(1). Verfügbar unter: <http://genomebiology.com/2010/11/1/R2> (zuletzt abgerufen am 15.06.2011)
- [OBOOJ] OBO-Edit Working Group: OBO-Edit. The OBO Ontology Editor. o. J.. Verfügbar unter: <http://oboedit.org/> (zuletzt abgerufen am 04.09.2011)
- [Ope09] Open Biological and Biomedical Ontologies Foundry: PATO:About. 2009. Verfügbar unter: <http://obofoundry.org/wiki/index.php/PATO:About> (zuletzt abgerufen am 15.06.2011)
- [Pat04] Patel-Schneider, Peter F.; Horrocks, Ian: OWL Web Ontology Language - Semantics and Abstract Syntax Section 2. Abstract Syntax. 2004. Verfügbar unter: <http://www.w3.org/TR/2004/REC-owl-semantics-20040210/syntax.html> (zuletzt abgerufen am 16.06.2011)
- [Rus04] Russell, Stuart; Norvig, Peter: Künstliche Intelligenz - Ein moderner Ansatz, 2. Auflage. Pearson Studium, München 2004: S. 397 - 463
- [RutOJ] Ruttenberg, Alan; Courtot, Melanie; Mungall, Christopher J.: OBO Foundry Identifier Policy. o. J.. Verfügbar unter: <http://www.obofoundry.org/id-policy.shtml> (zuletzt abgerufen am 19.06.2011)

- [SchOJ] Schmitt, P. H.: Beschreibungslogiken. o. J.. Verfügbar unter: <http://i12www.ira.uka.de/~pschmitt/NKL/NKLFolien.pdf/35.pdf> (zuletzt abgerufen am 16.06.2011)
- [Smi04] Smith, Michael K.; Welty, Chris; McGuinness, Deborah L.; Köstlbacher, Anton (Übers.); Diehn, Max (Übers.); Stephan, Andreas (Übers.): OWL Web Ontology Language - Guide. 2004. Verfügbar unter: <http://www.semaweb.org/dokumente/w3/TR/2004/REC-owl-guide-20040210-DE.html> (zuletzt abgerufen am 16.06.2011)
- [Smi05] Smith, Barry; Ceusters, Werner; Klagges, Bert; Köhler, Jacob; Kumar, Anand; Lomax, Jane; Mungall, Christopher J.; Neuhaus, Fabian; Rector, Alan L.; Rosse, Cornelius: Relations in biomedical ontologies. In: Genome Biology. PubMed Central, Bethesda 2005: v.6(5). Verfügbar unter: <http://genomebiology.com/2005/6/5/R46> (zuletzt abgerufen am 15.06.2011)
- [Smi11] Smith, Cynthia L.; Bello, Susan M.; Goldsmith, Carroll W.; Eppig, Janan T.: mammalian\_phenotype.obo. 2011. Verfügbar unter: [http://obo.cvs.sourceforge.net/viewvc/obo/obo/ontology/phenotype/mammalian\\_phenotype.obo](http://obo.cvs.sourceforge.net/viewvc/obo/obo/ontology/phenotype/mammalian_phenotype.obo) (zuletzt abgerufen am 04.09.2011)
- [Ste00] Steinmann, Friedrich: Formale Modellierung mit Rollen. 2000. Verfügbar unter: <http://www.fernuni-hagen.de/ps/pubs/HabilitationsschriftSteinmann.pdf> (zuletzt abgerufen am 16.06.2011)
- [Sou11] Souders, Steve: Interesting stats. 2011. Verfügbar unter: <http://www.httparchive.org/interesting.php> (zuletzt abgerufen am 04.09.2011)
- [The11] The Jackson Laboratory: Add2tm1Llp. 2011. Verfügbar unter: <http://www.informatics.jax.org/javawi2/servlet/WIFetch?page=alleleDetail&key=3542> (zuletzt abgerufen am 04.09.2011)
- [Tir11] Tirmizi, Syed Hamid; Aitken, Stuart; Moreira, Dilvan A.; Mungall, Christopher J.; Sequeda, Juan; Shah, Nigam H.; Miranker, Daniel P.: Mapping between the OBO and OWL ontology languages. In: JOURNAL OF BIOMEDICAL SEMANTICS. BioMed Central, London 2011: v.2(1). Verfügbar unter: <http://www.jbiomedsem.com/content/2/S1/S3> (zuletzt abgerufen am 16.06.2011)
- [Was09] Washington, Nicole L.; Haendel, Melissa A.; Mungall, Christopher J.; Ashburner Michael; Westerfield, Monte, Lewis, Suzanna E.: Linking Human Diseases to Animal Models Using Ontology-Based Phenotype Annotation. In: PloS Biology. PloS, San Francisco 2009: v.7(11). Verfügbar unter: <http://www.neuro.uoregon.edu/westerf/pdfs/Washington.pdf> (zuletzt abgerufen am 16.06.2011)



- [Wik11a] Wikimedia Foundation Inc.: Annotation. 2011. Verfügbar unter:  
<http://de.wikipedia.org/wiki/Annotation> (zuletzt abgerufen am  
13.11.2011)
- [Wik11b] Wikimedia Foundation Inc.: Phänotyp. 2011. Verfügbar unter:  
<http://de.wikipedia.org/wiki/Phänotyp> (zuletzt abgerufen am 13.11.2011)

## **7. Anlagenverzeichnis**

<b>8.1 Beispiel-Implementierung</b>	<b>51</b>
8.1.1 Beispiel-Implementierung für relationale GFO-Rollen	51
8.1.2 Beispiel-Implementierung für kontextsensitive Properties	56
8.1.3 Beispiel-Implementierung für künstlichen Relator	61
<b>8.2 Glossar</b>	<b>66</b>

## 8. Anlagen

### 8.1 Beispiel-Implementierung

Um die, in den letzten Kapiteln entwickelten, Modelle zu veranschaulichen, folgen nun Ausschnitte aus Beispiel-Implementierungen für die Qualitäten-Ontologie PATO und der Cross-Product-Ontologie MP-XP. In diesen Implementierungen wurden Lösungsansätze für das Vertauschungsproblem und die kontextinsensitive Verbindung zwischen Qualität und Entitäten berücksichtigt. Die vollständigen Beispiel-Implementierungen können der beiliegenden CD-ROM entnommen werden.

#### 8.1.1 Beispiel-Implementierung für relationale GFO-Rollen

```
<owl:Class rdf:about="#PATO_0000117">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    size
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0000051"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#sizes"/>
      <owl:someValuesFrom rdf:resource=
        "file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:about="#PATO_0000587">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    decreased size
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0000117"/>
</owl:Class>
```

```

<owl:Class rdf:about="#PATO_0000033">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    concentration of
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0002182"/>
  <rdfs:subClassOf rdf:resource=
    "http://www.onto-med.de/ontologies/gfo.owl#Relator"/>
  <rdfs:subClassOf>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <owl:Restriction>
          <owl:onProperty rdf:resource="#has_role"/>
          <owl:someValuesFrom rdf:resource="#Role_02"/>
        </owl:Restriction>
        <owl:Restriction>
          <owl:onProperty rdf:resource="#has_role"/>
          <owl:someValuesFrom rdf:resource="#Role_03"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:about="#PATO_0001159">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    concentrated
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0000033"/>
</owl:Class>

<owl:Class rdf:about="#PATO_0001162">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    increased concentration
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0001159"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#qualifier"/>
      <owl:someValuesFrom rdf:resource="#PATO_0000460"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

```

```

<owl:Class rdf:about="#Role_01">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    BiologicalRole
  </rdfs:label>
  <rdfs:subClassOf rdf:resource=
    "http://www.onto-med.de/ontologies/gfo.owl#Relational_role"/>
</owl:Class>

<owl:Class rdf:about="#Role_02">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    ConcentratingRole
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#Role_01"/>
</owl:Class>

<owl:Class rdf:about="#Role_03">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    ConcentratedRole
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#Role_01"/>
</owl:Class>

<owl:ObjectProperty rdf:about="#has_phene">
  <rdfs:domain rdf:resource="#Creature_01"/>
  <rdfs:range rdf:resource="#PATO_0000001"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#sizes">
  <rdfs:domain rdf:resource="#PATO_0000117"/>
  <rdfs:range rdf:resource="file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
  <owl:subPropertyOf rdf:resource="#relates_to"/>
</owl:ObjectProperty>

```

**Beschreibung 41:** Ausschnitt aus der, um relationale Rollen erweiterten, Phenotype and Trait Ontology (PATO\_RC)

```

<owl:Class rdf:about="#MP_0002813">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    microcytosis
  </rdfs:label>
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource=
        "file:/C:/Beispiel/PATO_RC.owl#has_phene"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:intersectionOf rdf:parseType="Collection">
            <rdf:Description rdf:about=
              "file:/C:/Beispiel/PATO_RC.owl#PATO_0000587"/>
            <owl:Restriction>
              <owl:onProperty rdf:resource=
                "file:/C:/Beispiel/PATO_RC.owl#sizes"/>
              <owl:someValuesFrom rdf:resource=
                "file:/C:/Beispiel/Entities.owl#CL_0000232"/>
            </owl:Restriction>
          </owl:intersectionOf>
        </owl:Class>
      </owl:someValuesFrom>
    </owl:Restriction>
  </owl:equivalentClass>
</owl:Class>

```

```

<owl:Class rdf:about="#MP_0008809">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    increased spleen iron level
  </rdfs:label>
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource=
        "file:/C:/Beispiel/PATO_RC.owl#has_phene"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:intersectionOf rdf:parseType="Collection">
            <rdf:Description rdf:about=
              "file:/C:/Beispiel/PATO_RC.owl#PATO_0001162"/>
            <owl:Restriction>
              <owl:onProperty rdf:resource=
                "file:/C:/Beispiel/PATO_RC.owl#has_role"/>
              <owl:someValuesFrom>
                <owl:Class>
                  <owl:intersectionOf rdf:parseType="Collection">
                    <rdf:Description rdf:about=

```

```

        "file:/C:/Beispiel/PATO_RC.owl#Role_02"/>
    <owl:Restriction>
        <owl:onProperty rdf:resource=
            "file:/C:/Beispiel/PATO_RC.owl#played_by"/>
        <owl:someValuesFrom rdf:resource=
            "file:/C:/Beispiel/Entities.owl#MA_0000141"/>
    </owl:Restriction>
</owl:intersectionOf>
</owl:Class>
</owl:someValuesFrom>
</owl:Restriction>
<owl:Restriction>
    <owl:onProperty rdf:resource=
        "file:/C:/Beispiel/PATO_RC.owl#has_role"/>
    <owl:someValuesFrom>
        <owl:Class>
            <owl:intersectionOf rdf:parseType="Collection">
                <rdf:Description rdf:about=
                    "file:/C:/Beispiel/PATO_RC.owl#Role_03"/>
                <owl:Restriction>
                    <owl:onProperty rdf:resource=
                        "file:/C:/Beispiel/PATO_RC.owl#played_by"/>
                    <owl:someValuesFrom rdf:resource=
                        "file:/C:/Beispiel/Entities.owl#CHEBI_18248"/>
                </owl:Restriction>
            </owl:intersectionOf>
        </owl:Class>
    </owl:someValuesFrom>
</owl:Restriction>
</owl:intersectionOf>
</owl:Class>
</owl:someValuesFrom>
</owl:Restriction>
</owl:equivalentClass>
</owl:Class>

```

**Beschreibung 42:** Ausschnitt aus der, um relationale Rollen erweiterten, MP\_XP (MP\_XP\_RC)

### 8.1.2 Beispiel-Implementierung für kontextsensitive Properties

```
<owl:Class rdf:about="#PATO_0000117">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    size
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0000051"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#sizes"/>
      <owl:someValuesFrom rdf:resource=
        "file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:about="#PATO_0000587">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    decreased size
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0000117"/>
</owl:Class>
```



```

<owl:Class rdf:about="#PATO_000033">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    concentration of
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0002182"/>
  <rdfs:subClassOf rdf:resource=
    "http://www.onto-med.de/ontologies/gfo.owl#Relator"/>
  <rdfs:subClassOf>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <owl:Restriction>
          <owl:onProperty rdf:resource="#concentrates"/>
          <owl:someValuesFrom rdf:resource=
            "file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
        </owl:Restriction>
        <owl:Restriction>
          <owl:onProperty rdf:resource="#is_concentrated_in"/>
          <owl:someValuesFrom rdf:resource=
            "file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:about="#PATO_0001159">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    concentrated
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_000033"/>
</owl:Class>

<owl:Class rdf:about="#PATO_0001162">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    increased concentration
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0001159"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#qualifier"/>
      <owl:someValuesFrom rdf:resource="#PATO_0000460"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

```

```
<owl:ObjectProperty rdf:about="#has_phene">
  <rdfs:domain rdf:resource="#Creature_01"/>
  <rdfs:range rdf:resource="#PATO_0000001"/>
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:about="#relates_to">
  <rdfs:domain rdf:resource="#PATO_0000001"/>
  <rdfs:range rdf:resource="file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:about="#sizes">
  <rdfs:domain rdf:resource="#PATO_0000117"/>
  <rdfs:range rdf:resource="file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
  <owl:subPropertyOf rdf:resource="#relates_to"/>
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:about="#concentrates">
  <rdfs:domain rdf:resource="#PATO_0000033"/>
  <rdfs:range rdf:resource="file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
  <owl:subPropertyOf rdf:resource="#relates_to"/>
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:about="#is_concentrated_in">
  <rdfs:domain rdf:resource="#PATO_0000033"/>
  <rdfs:range rdf:resource="file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
  <owl:subPropertyOf rdf:resource="#relates_to"/>
</owl:ObjectProperty>
```

**Beschreibung 43:** Ausschnitt aus der, um kontextsensitive Properties erweiterten, Phenotype and Trait Ontology (PATO\_CSP)

```

<owl:Class rdf:about="#MP_0002813">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    microcytosis
  </rdfs:label>
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource=
        "file:/C:/Beispiel/PATO_CSP.owl#has_phene"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:intersectionOf rdf:parseType="Collection">
            <rdf:Description rdf:about=
              "file:/C:/Beispiel/PATO_CSP.owl#PATO_0000587"/>
            <owl:Restriction>
              <owl:onProperty rdf:resource=
                "file:/C:/Beispiel/PATO_CSP.owl#sizes"/>
              <owl:someValuesFrom rdf:resource=
                "file:/C:/Beispiel/Entities.owl#CL_0000232"/>
            </owl:Restriction>
          </owl:intersectionOf>
        </owl:Class>
      </owl:someValuesFrom>
    </owl:Restriction>
  </owl:equivalentClass>
</owl:Class>

```

```

<owl:Class rdf:about="#MP_0008809">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    increased spleen iron level
  </rdfs:label>
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource=
        "file:/C:/Beispiel/PATO_CSP.owl#has_phene"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:intersectionOf rdf:parseType="Collection">
            <rdf:Description rdf:about=
              "file:/C:/Beispiel/PATO_CSP.owl#PATO_0001162"/>
            <owl:Restriction>
              <owl:onProperty rdf:resource=
                "file:/C:/Beispiel/PATO_CSP.owl#concentrates"/>
              <owl:someValuesFrom rdf:resource=
                "file:/C:/Beispiel/Entities.owl#CHEBI_18248"/>
            </owl:Restriction>
            <owl:Restriction>
              <owl:onProperty rdf:resource=
                "file:/C:/Beispiel/PATO_CSP.owl#is_concentrated_in"/>
              <owl:someValuesFrom rdf:resource=
                "file:/C:/Beispiel/Entities.owl#MA_0000141"/>
            </owl:Restriction>
          </owl:intersectionOf>
        </owl:Class>
      </owl:someValuesFrom>
    </owl:Restriction>
  </owl:equivalentClass>
</owl:Class>

```

**Beschreibung 44:** Ausschnitt aus der, um kontextsensitive Properties erweiterten, MP-XP (MP\_XP\_CSP)

### 8.1.3 Beispiel-Implementierung für künstliche Relatoren

```
<owl:Class rdf:about="#PATO_0000117">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    size
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0000051"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#sizes"/>
      <owl:someValuesFrom rdf:resource=
        "file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

```
<owl:Class rdf:about="#PATO_0000587">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    decreased size
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0000117"/>
</owl:Class>
```

```
<owl:Class rdf:about="#PATO_0000033">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    concentration of
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0002182"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#inheres_in"/>
      <owl:someValuesFrom rdf:resource="#Relator_01"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

```
<owl:Class rdf:about="#PATO_0001159">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    concentrated
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0000033"/>
</owl:Class>
```

```

<owl:Class rdf:about="#PATO_0001162">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    increased concentration
  </rdfs:label>
  <rdfs:subClassOf rdf:resource="#PATO_0001159"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#qualifier"/>
      <owl:someValuesFrom rdf:resource="#PATO_0000460"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

```

```

<owl:Class rdf:about="#Relator_01">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    ConcentrationRelator
  </rdfs:label>
  <rdfs:subClassOf rdf:resource=
    "http://www.onto-med.de/ontologies/gfo.owl#Relator"/>
  <owl:equivalentClass>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <owl:Restriction>
          <owl:onProperty rdf:resource="#concentrates"/>
          <owl:someValuesFrom rdf:resource=
            "file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
        </owl:Restriction>
        <owl:Restriction>
          <owl:onProperty rdf:resource="#is_concentrated_in"/>
          <owl:someValuesFrom rdf:resource=
            "file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </owl:equivalentClass>
</owl:Class>

```

```

<owl:ObjectProperty rdf:about="#has_phene">
  <rdfs:domain rdf:resource="#Creature_01"/>
  <rdfs:range rdf:resource="#PATO_0000001"/>
</owl:ObjectProperty>

```

```
<owl:ObjectProperty rdf:about="#inheres_in">
  <rdfs:domain rdf:resource="#PATO_0000001"/>
  <rdfs:range rdf:resource="http://www.onto-med.de/ontologies/gfo.owl#Relator"/>
  <owl:subPropertyOf rdf:resource="#relates_to"/>
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:about="#sizes">
  <rdfs:domain rdf:resource="#PATO_0000117"/>
  <rdfs:range rdf:resource="file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
  <owl:subPropertyOf rdf:resource="#relates_to"/>
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:about="#concentrates">
  <rdfs:domain rdf:resource="#PATO_0000033"/>
  <rdfs:range rdf:resource="file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
  <owl:subPropertyOf rdf:resource="#relates_to"/>
</owl:ObjectProperty>
```

```
<owl:ObjectProperty rdf:about="#is_concentrated_in">
  <rdfs:domain rdf:resource="#PATO_0000033"/>
  <rdfs:range rdf:resource="file:/C:/Beispiel/Entities.owl#BiologicalEntity_01"/>
  <owl:subPropertyOf rdf:resource="#relates_to"/>
</owl:ObjectProperty>
```

**Beschreibung 45:** Ausschnitt aus der, um künstliche Relatoren erweiterten, Phenotype and Trait Ontology (PATO\_AR)

```

<owl:Class rdf:about="#MP_0002813">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    microcytosis
  </rdfs:label>
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource=
        "file:/C:/Beispiel/PATO_AR.owl#has_phene"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:intersectionOf rdf:parseType="Collection">
            <rdf:Description rdf:about=
              "file:/C:/Beispiel/PATO_AR.owl#PATO_0000587"/>
            <owl:Restriction>
              <owl:onProperty rdf:resource=
                "file:/C:/Beispiel/PATO_AR.owl#sizes"/>
              <owl:someValuesFrom rdf:resource=
                "file:/C:/Beispiel/Entities.owl#CL_0000232"/>
            </owl:Restriction>
          </owl:intersectionOf>
        </owl:Class>
      </owl:someValuesFrom>
    </owl:Restriction>
  </owl:equivalentClass>
</owl:Class>

```



```

<owl:Class rdf:about="#MP_0008809">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    increased spleen iron level
  </rdfs:label>
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource=
        "file:/C:/Beispiel/PATO_AR.owl#has_phene"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:intersectionOf rdf:parseType="Collection">
            <rdf:Description rdf:about=
              "file:/C:/Beispiel/PATO_AR.owl#PATO_0001162"/>
            <owl:Restriction>
              <owl:onProperty rdf:resource=
                "file:/C:/Beispiel/PATO_AR.owl#inheres_in"/>
              <owl:someValuesFrom>
                <owl:Class>
                  <owl:intersectionOf rdf:parseType="Collection">
                    <rdf:Description rdf:about=
                      "file:/C:/Beispiel/PATO_AR.owl#Relator_01"/>
                    <owl:Restriction>
                      <owl:onProperty rdf:resource=
                        "file:/C:/Beispiel/PATO_AR.owl#concentrates"/>
                      <owl:someValuesFrom rdf:resource=
                        "file:/C:/Beispiel/Entities.owl#CHEBI_18248"/>
                    </owl:Restriction>
                    <owl:Restriction>
                      <owl:onProperty rdf:resource=
                        "file:/C:/Beispiel/PATO_AR.owl#is_concentrated_in"/>
                      <owl:someValuesFrom rdf:resource=
                        "file:/C:/Beispiel/Entities.owl#MA_0000141"/>
                    </owl:Restriction>
                  </owl:intersectionOf>
                </owl:Class>
              </owl:someValuesFrom>
            </owl:Restriction>
          </owl:intersectionOf>
        </owl:Class>
      </owl:someValuesFrom>
    </owl:Restriction>
  </owl:equivalentClass>
</owl:Class>

```

**Beschreibung 46:** Ausschnitt aus der, um künstliche Relatoren erweiterten, MP-XP (MP\_XP\_AR)

## 8.2 Glossar

Annotation	Unter Annotation versteht man die Bereitstellung von Anmerkungen, die eine Informationsquelle ergänzen sollen.
Beschreibung	siehe → Beschreibung, phänotypische
Beschreibung, phänotypische	1. Bedeutung: Unter einer phänotypischen Beschreibung ist die formale Repräsentation eines → Phänotyps zu verstehen. 2. Bedeutung: Unter einer phänotypischen Beschreibung kann auch die formale Repräsentation eines → Phäns verstanden werden. Diese Repräsentation basiert in dieser Arbeit einheitlich auf dem → EQ-Modell.
Charakteristik, phänotypische	Merkmale eines → Lebewesens werden als phänotypische Charakteristika bezeichnet. Ein → Phänotyp ist eine Menge aus phänotypischen Charakteristika. Phänotypische Charakteristika werden häufig einfach als Phänotyp bezeichnet.
Cross-Product	Cross-Products sind spezielle Ontologien, in denen Konzepte einer Phänotyp-Ontologie formal definiert werden. Ein Beispiel eines solchen Cross-Products ist MP-XP.
Entität	1. Bedeutung: Entität umfasst allgemein im Kontext der Beschreibungslogik Konzepte und deren Instanzen. 2. Bedeutung: Im Phänotyp-Kontext ist „Entität“ als ein Organismus, ein Teil eines Organismus, eine chemische Substanz oder ähnlichem zu verstehen. Entitäten stammen aus → spezies-spezifischen Ontologien und werden zur Beschreibung → phänotypischer Charakteristika genutzt.
Entity-Quality-Modell	Das Entity-Quality-Modell ist ein Modell zur Beschreibung phänotypischer Charakteristika, dass auf EQ-Beschreibungen basiert.
EQ-Modell	siehe → Entity-Quality-Modell
EQ-Beschreibung	EQ-Beschreibungen dienen zur formalen Repräsentation einer → phänotypischen Charakteristik. Sie bestehen aus einer → Qualität $Q$ , einer → Entität $E$ , die Träger von $Q$ ist, einer zusätzlichen Entität $E2$ für → relationale Qualitäten und einem optionalen → Modifier $M$ .

General Formal Ontology	Die General Formal Ontology ist eine, von der Onto-Med entwickelte, Top-Level-Ontologie.
GFO	siehe → General Formal Ontology
Lebewesen	Ein Lebewesen ist ein Organismus, in dem ein Phänotyp inhäriert.
Mammalian Phenotype Ontology	Die Mammalian Phenotype Ontology ist eine → Ontologie → prä-koordinierter Phänotypen.
Modifier	Bei einem Modifier handelt es sich um eine spezielle → Qualität aus → PATO, die dazu verwendet wird andere Qualitäten zu charakterisieren. Ein Beispiel für einen Modifier ist <i>abnormal</i> .
MP	siehe → Mammalian Phenotype Ontology
OBO Flat File Format	Das OBO Flat File Format ist ein proprietäres Format der → OBO Foundry zur Ontologie-Repräsentation.
OBO Foundry	siehe → Open Biological and Biomedical Ontology Foundry
Ontologie	Bei einer Ontologie handelt es sich um eine Art der Wissensrepräsentation zur Beschreibung von Konzepten, Instanzen und ihren verbindenden Relationen innerhalb einer bestimmten Domäne.
Ontologie, spezies-spezifische	Die Konzepte einer spezies-spezifischen Ontologie sind Entitäten (2. Bedeutung). Ein Beispiel einer solchen Ontologie ist MA.
Open Biological and Biomedical Ontology Foundry	Die Open Biological and Biomedical Ontology Foundry ist eine Organisation, die sich um die Partitionierung der biologischen und biomedizinischen Domäne und der Sicherstellung logischer Kohärenz von → Ontologien des Fachgebietes kümmert.
OWL	siehe → Web Ontology Language
OWL-Property	Eine OWL-Property ist eine Relation zwischen zwei Entitäten (1. Bedeutung). In OWL wird sie durch das <owl:onProperty>-Tag gekennzeichnet.
PATO	siehe → Phenotype and Trait Ontology
Phän	siehe → Charakteristik, phänotypische
Phänotyp	Ein Phänotyp ist eine Menge aus phänotypischen Charakteristika, die durch den Genotyp oder der Interaktion des Phän-Trägers (Lebewesen) mit seiner Umwelt bestimmt wird. Phänotypen bezeichnen häufig auch Phäne.

Phänotyp, atomarer	Ein atomarer Phänotyp ist das Konzept $\exists has\_phene.Phene$ , wobei <i>Phene</i> ein Phän bezeichnet.
Phänotyp, komplexer	Ein komplexer Phänotyp bezeichnet die Konjunktion atomarer Phänotypen.
Phänotyp, prä-koordinierter	Bei einem prä-koordinierten Phänotyp handelt es sich um eine $\rightarrow$ phänotypische Beschreibung, die bereits vor Bedarf erstellt wurde. Prä-koordinierte Phänotypen sind meist nur informell beschrieben, weshalb sie schlecht für die maschinelle Verarbeitung geeignet sind. Definiert sind sie z. B. in $\rightarrow$ MP.
Phänotyp, post-koordinierter	Bei einem post-koordinierten Phänotyp handelt es sich um eine $\rightarrow$ phänotypische Beschreibung, die erst bei Bedarf erstellt wird. Meist werden sie im $\rightarrow$ EQ-Modell beschrieben.
Phenotype and Trait Ontology	Die Phenotype and Trait Ontology ist eine $\rightarrow$ Ontologie aus $\rightarrow$ Qualitäten, die zur Beschreibung $\rightarrow$ post-koordinierter Phänotypen verwendet werden können.
Property Relation	siehe $\rightarrow$ OWL-Property
	1. Bedeutung: Allgemein handelt es sich bei einer Relation um eine Beziehung zwischen zwei $\rightarrow$ Entitäten (1. Bedeutung).
	2. Bedeutung: Relation wird in dieser Arbeit im Kontext formaler Repräsentationen in Beschreibungslogik teilweise synonym zu „Rolle“ verwendet.
	3. Bedeutung: Im Kontext der $\rightarrow$ GFO handelt es sich bei einer Relation um eine Entität kategorialen Charakters, die analog zur 1. Bedeutung die Beziehung zwischen zwei $\rightarrow$ Entitäten beschreibt.
Relator	Ein Relator ist die Instanz einer GFO-Relation.
Rolle	1. Bedeutung: Im Kontext der Beschreibungslogik bezeichnet eine Rolle die Relation zwischen zwei $\rightarrow$ Entitäten (1. Bedeutung).
	2. Bedeutung: Im Kontext von $\rightarrow$ GFO handelt es sich bei einer relationalen Rolle um eine $\rightarrow$ Entität (1. Bedeutung) zur Verbindung eines $\rightarrow$ Relators mit einem seiner Argumente.
Qualität	Qualitäten sind Konzepte aus $\rightarrow$ PATO und umfassen sowohl Attribute als auch deren korrespondierenden qualifizierenden Werte.

Qualität, monadische	Bei einer monadischen Qualität handelt es sich um eine → Qualität, die lediglich mit einer → Entität (2. Bedeutung) verbunden ist (z. B. durch Inhärenz).
Qualität, relationale	Eine relationale Qualität ist immer mit mindestens zwei → Entitäten (2. Bedeutung) verbunden.
Web Ontology Language	Die Web Ontology Language ist die meist verwendete Beschreibungssprache zur Realisierung von Web-Ontologien.

## **9. Erklärung**

Ich versichere, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe, insbesondere sind wörtliche oder sinngemäße Zitate als solche gekennzeichnet. Mir ist bekannt, dass Zuwiderhandlung auch nachträglich zur Aberkennung des Abschlusses führen kann.

Leipzig, 15.11.2011