Towards Improving Phenotype Representation in OWL

Frank Loebe [1]*, Frank Stumpf [1], Robert Hoehndorf [2] and Heinrich Herre [3]

[1]Department of Computer Science, University of Leipzig, Germany
[2]Department of Genetics, University of Cambridge, UK
[3]Institute of Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Germany

ABSTRACT

Phenotype ontologies are used in species-specific databases for the annotation of mutagenesis experiments and to characterize human diseases. The Entity-Quality (EQ) formalism is a means to describe complex phenotypes based on one or more affected entities and a quality. EQ-based definitions have been developed for many phenotype ontologies, including the Human and Mammalian Phenotype ontologies. We analyze the OWL-based formalizations of complex phenotype descriptions based on the EQ model, identify several representational challenges and analyze potential solutions to address these challenges. In particular, we suggest a novel, role-based approach to represent *relational qualities* such as *Concentration of calcium in blood*, discuss its ontological foundation in the General Formal Ontology (GFO) and evaluate its representation in OWL and the benefits it can bring to the representation of phenotype annotations. Our analysis of OWL-based representation of phenotypes can contribute to improving consistency and expressiveness of formal phenotype descriptions.

1 INTRODUCTION

In recent years, molecular biology has made significant progress in understanding the mechanisms underlying human disease. Several studies investigate disease mechanisms in animals that serve as models for humans [30]. In particular, the targeted modification of the genetic markup of these organisms provides a powerful means to investigate the molecular mechanisms associated with heritable diseases in humans [8]. Large-scale mutagenesis projects are now underway with the aim to characterize the outcomes of null-mutations for every gene in an organism. The observable characteristics of these modified organisms (their phenotypes) are represented in model organism databases and can be utilized to suggest candidate genes for diseases for which no molecular origin is currently known [20].

To standardize the terminology used in describing phenotypes, multiple species-specific phenotype ontologies were developed. For example, the Mammalian Phenotype Ontology (MP) [33, 7] is used to characterize phenotypes in mice and other mammals, and the Worm Phenotype Ontology (WPO) [31] is used to characterize *C. elegans* phenotypes. The Human Phenotype Ontology (HPO) [29] describes phenotypes in humans and is applied for describing human diseases and individual patients.

To translate phenotypes across species and enable their comparison with human phenotypes and disease, a syntax for phenotype decompositions has been developed [5, 37, 26]. In this syntax, phenotypes are represented by a combination of a quality and one or more entities. The entities represent the entities that are affected by a phenotype and are either physiological processes and functions (from the Gene Ontology [2]) or anatomical structures as represented by species-specific anatomy ontologies. The Phenotypic Attribute and Trait Ontology (PATO) [9] is an ontology of qualities which is used to describe *how* an entity is affected within a phenotype. Entity-Quality (EQ) based specifications of phenotypes have been developed for several species-specific phenotype ontologies [26], including the HPO [29], MP [33, 6, 7], WPO [31], and others, thereby intergrating pre- and postcoordinated biomedical ontologies [32, 26].

Recently, mechanisms became available to enable the automated translation of phenotypes across different species [26, 20]. In these methods, ontologies are integrated through species-independent ontologies, and automated reasoning over the integrated ontologies enables the automated comparison of species-specific phenotype information across multiple species. This approach crucially relies on the formalization of phenotype information in ontologies and model organism databases. With the increasing application of ontologies for data analysis, improving the representation of phenotype ontologies has the potential to directly affect and advance scientific analyses and discoveries.

The EQ model is an important and widely used means for formalizing phenotype information in ontologies [4]. In greater detail, its main idea is to combine an 'entity class' (the E in EQ) from an anatomy or process ontology with a 'quality class' (the Q) from PATO. For example, the class *eye* (MA:000261 in the Mouse adult gross anatomy ontology (MA) [14]) as the E and the color *red* (PATO:0000322) for Q can be combined to form the class *Red eye*. The typical formal interpretation of EQ statements is that the combination refers to a specialization of the quality class Q such that it inheres in instances of the entity class E [26, p. 3],[25]. In the example, this yields the class *red that inheres in an eye* (cf. Fig. 1).

Relational qualities involve at least one additional entity besides E. In the semantics of EQ, a second entity can be attached to a quality via the relation `towards` [26, p. 3–5]. An example of this kind is the *concentration of iron in the spleen*, which can be formalized as a quality *concentration of* (PATO:0000033) inhering in *spleen* (MA:0000141) and connected via `towards` to *iron* (CHEBI:18248

Figure 1. EQ model. (Gray indicates the optional part for relational qualities.)

*to whom correspondence should be addressed

in the ontology of Chemical Entities of Biological Interest [21]), in order to define *abnormal spleen iron level* (MP:0008739).[1]

The term 'relational quality' as nowadays found in the bio-ontology community is typically used without further analysis, e.g., in [26] and, through [25], can be traced back to [27] where it seems to be meant synonymously with the more widely used term 'relation'. Notably, in the context of formal ontology, by 'relational qualities' sometimes constituents of particular relation instances are referred to (in contrast to the overall relation instances themselves), termed 'relational roles' in sect. 3.3.

While EQ descriptions characterize a phenotype, a related question pertains to the formalization of the *annotation* of organisms, genotypes and genes with EQ-based phenotype descriptions. In model organism databases such as the MGI database [6], genotypes like Add2^{tmLlp} (MGI:2149065) are annotated with a class like *abnormal spleen iron level* (MP:0008739). The intended meaning of this annotation is that organisms of a particular mouse strain that exhibit the described genotype (a targeted mutation of the *Add2* gene) within a specific environment will develop the *abnormal spleen iron level* phenotype. This complex relation can be simplified to improve performance of specific information retrieval tasks into a view in which the genotype is equivalent to the intersection of phenotypes and individual mice instances of their phenotypic annotations.

Only few efforts formally explore the compositional nature of phenotypes, i.e., how atomic phenotypes can be combined into more complex phenotypes such as in disease descriptions or in genotypes annotated with multiple phenotypes. In particular, the naive combination of phenotypes such as *red eye* with *short tail* is based on class intersections, and these lead to contradictory class definitions due to the disjointness of *color* (the super-class of *red*) and *size* (the super-class of *short*) [19]. More challenging are combinations of qualities which are hidden in the taxonomy of biomedical ontologies. For example, asserting that *red eye* is a sub-class of an *abnormal eye morphology* will imply that *red eye* is both a subclass of *morphology* and *color*. This will lead to another contradictory class definition due to the disjointness of *color* and *morphology* [18].

2 REPRESENTING PHENOTYPES IN OWL

2.1 Basic Problems

We see three *basic problems* that need to be addressed regarding the representation of phenotypes and the interpretation of EQ descriptions in terms of the Web Ontology Language (OWL) [36], in order to utilize automated and semantically correct reasoning to its full extent.

I. ontological foundation of complex phenotypes
II. representation of phenotypes in formal languages
III. ontological foundation of phenotype annotations

The first problem concerns the ontological foundation of complex phenotypes. To address this problem, we attempt to gain a clear understanding of the ontological nature of complex phenotypes and rely on an ontological framework for the explanation and foundation of complex phenotypes which does not depend on the expressive power of OWL. Once we obtained an understanding of the ontological nature of complex phenotypes, we investigate how to represent them in OWL, as a case of the second problem. The next step then is to apply this theory to existing descriptions of complex phenotype, such as those found in phenotypic annotation of diseases and genotypes in model organism annotations.

2.2 Issues of Formal Representation

The first basic problem requires further attention, but is widely discussed in biomedicine and formal ontology, e.g. see [24, 19, 35]. In the present paper, our focus is on the second problem and its application to formalizing phenotype annotations. In this regard we identify five interrelated *particular issues* that affect our analyses.

1. ontological adequacy / coherence of ontological interpretation
2. invalid permutations / ambiguities
3. relational expressiveness
4. consistency of domain modeling
5. formal reflection of annotations

Referring to *ontological adequacy*, we intend to find OWL representations that are close to the ontological understanding of phenotypes as *qualities*, similar to established ontological theories of phenotypes [25, 26].

While several approaches allow for representations of individual EQ statements in OWL, combining multiple EQ statements by means of their intersections may create incorrect [19, sect. 4.2, p. 3117] and sometimes contradictory statements [18]. For instance, consider the following OWL concept:

```
(red that inheresIn some eye) and
(short that inheresIn some tail)
```
(1)

Concept (1) is necessarily empty, because no instance of `red` is equally an instance of `short`. Furthermore, this formalization faces the problem of *permutations* (issue two), arising from the commutativity and associativity of intersections in OWL. In particular, the parentheses in example (1) are merely auxiliary for reading. The concept is formally equivalent to `(red that inheresIn some tail) and (short that inheresIn some eye)`. As a consequence, queries will deliver incorrect results if this mode of combining EQ statements is used.

The next two issues concern primarily phenotypes based on relational properties, like the *iron concentration in the spleen*. *Relational expressiveness* is used for referring to limitations of the arity of relations that can be specified with an EQ description. The current model does not allow for relational qualities of an arity greater than two. This may lead to undesirable consequences, since several applications of biomedical knowledge representation require relations of higher arity [34, 10]. This issue has been identified as a particularly important challenge for representing EQ-based phenotypes [25]. Closely connected to the number of arguments is the question of *inter-modeler consistency / harmonization*, cf. also [10]. This fourth issue refers to the question how to link (a class representing) a relation to (classes of) its arguments such that it is as unambigious as possible which argument connects to the relation in which way. In the current EQ model confusion can arise,

[1] Despite continued use of this example, we will not go into detailed ontological analyses of the relationship between iron and spleen, e.g., as particulars. In particular, iron as an amount of matter/quantity would deserve special treatment, cf. e.g. [12].

e.g., on whether *calcium concentration of blood* should be formalized as `concentration that inheresIn some blood and towards some calcium` or instead as `concentration that inheresIn some calcium and towards some blood`. The different positions may correlate with the community/background of modelers, e.g. whether a biologist or a chemist makes the assertion. Corresponding decisions are not only relevant for formalization, but likewise influence querying. For the particular case of concentrations, [13] proposes inherence in those entities that are concentrated in another in the context of an ontological analysis, i.e., inherence in calcium in the example. We comment on this in sect. 4, with hindsight regarding our analysis.

The fifth and final issue is the orientation and clarification of how *annotations* are interpreted, for any account of phenotype representations. This immediately links back to the ontological reading of phenotype representations and the third basic problem above.

3 ANALYSIS OF ALTERNATIVE APPROACHES

3.1 Spectrum of Solutions

In general, different approaches may be pursued in order to tackle the issues presented for the second basic problem. Like in [25], quality models that are fairly distinct from the EQ model may be (re-)considered. Another general change would be to concentrate on entities, i.e., primarily on the parts of an organism occurring in EQ descriptions, and to construct phenotype descriptions centering on them. E.g., the scheme E `that hasQuality some` Q follows this line of thought.[2]

We, however, focus first on solutions that limit the number of changes to the established interpretation of EQ descriptions. The latter are meanwhile widely in use, cf. e.g. [4], as are phenotype ontologies with their basic presupposition of providing (sub)concepts of *quality*. Therefore, the migration to new proposals should be facilitated by an approach with less changes compared to more radical revisions.

3.2 EQ Interpretations with regard to Annotations

What appears unavoidable is a more complex provision for annotations, at least if complex phenotypes formalized in OWL/description logic (DL) [3] shall be composable in terms of the usual intersection. Implicitly, this has already been observed in [19], to some extent also in connection with the EQ formalism. The following adheres to the understanding of annotations as outlined in sect. 1 and is inspired by the notion of *phenes* in [19]. Nevertheless, the subsequent variant differs in order to minimize changes to PATO and phenotype ontologies.

In order to solve especially the permutation problem of combined EQ descriptions, formally it suffices to have an "encapsulating" relation available. For instance, while (1) suffers from unwanted permutations, this is avoided in (2), where the encapsulating relation is termed `hasPheno`.

```
hasPheno some (red that inheresIn some eye) and
hasPheno some (short that inheresIn some tail)        (2)
```

Naturally, the question arises which ontological reading applies to `hasPheno`. We interpret (2) as a concept for classifying organisms (by two phenotype descriptions). The `hasPheno` relation belongs to an interpretive view/pattern that overlays common interconnections of entities, centering on the organism. In terms of the example, one may consider an organism O that has an eye E as its part, while there is a red R that inheres in E. Thus O is indirectly related with R in terms of common relations like inherence and part-of. In the phenotype view, this allows us to view O, as *phenotype bearer*, to exhibit R as a *pheno* of O. The latter connection is reflected by the `hasPheno` link between O and R. We require that each `hasPheno` link is "justified" by a chain of basic relations like *inheres-in*, *part-of*, *has-function*, *participates-in*, etc., that connects the entity in the pheno role with the one in the phenotype bearer role (PB in Fig. 2–4 below).

This approach leaves existing ontologies intact, resolves the first two particular issues identified, and accounts for the fifth, as well.

3.3 Enhancements for Relational Qualities

3.3.1 Purely Formal Extension On the remaining issues of relational expressiveness and consistency of domain modeling, we first observe that the current relational EQ model forms a special case of reifying (only binary) relations with *fixed* auxiliary relations, cf. the structural part of [1]. The main uncommon feature is the naming of those auxiliary relations as `inheresIn` and `towards`,[3] rather than using names counting arguments like $\texttt{argument}_1$ and $\texttt{argument}_2$. With the latter, an extension to n-ary relations is straightforward, which would solve the expressiveness issue. However, with fixed auxiliary relations there is no support for consistent domain modeling because the assignment of "values" to arguments is arbitrary. This may be the reason why all published variants of this pattern that we are aware of eventually suggest the *variable*, relation-specific naming of auxiliary relations [34, sect. 5.1],[28, 1].

Therefore, we do not see that changing the interpretation of relational EQ statements could be sidestepped, if inter-modeler consistent domain modeling is to be supported any further. Striving at the same time for ontological adequacy somewhat systematically, we adopt the model of relations and (relational)[4] roles from the General Formal Ontology (GFO) [15, 16], cf. also [23, 22].

3.3.2 Ontological Alternatives Using Relations In brief, relations in GFO are considered as categories of relators. *Relators* are

[2] Notably, this scheme is seen as equally eligible as phenotype description as the basic EQ scheme Q `that inheresIn some` E in [25, sect. 2.3, p. 5]. Giving preference to the basic EQ scheme appears to have been an arbitrary choice. In terms of their relationship to annotated entities the two schemes differ evidently. Nevertheless, the entity-focused scheme shares analogous problems to those expounded for the basic EQ scheme, in particular the permutation problem.

[3] Admittedly, `inheresIn` is meant to link to the ontological notion of inherence, whereas `towards` is introduced for rather technical reasons in [25] (circumventing an inherence relation of higher arity). It remains to be explored in greater detail whether `towards` can be adequately reinterpreted in terms of the notion of *external dependence*, see [11, esp. sect. 6.2.7].

[4] There are more types of roles in GFO, but for brevity we use roles and relational roles as synonyms herein. Note further that from here on 'role' is reserved for the ontological interpretation, whereas the meaning as set of pairs / as binary relation in the context of description logics and OWL is referred to as 'OWL property' or 'DL role'.

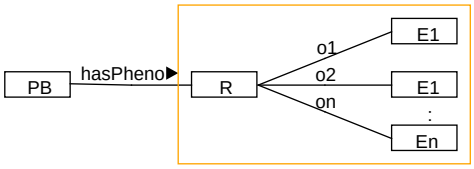

Figure 2. Roles-as-properties: Ontological roles encoded as OWL properties.

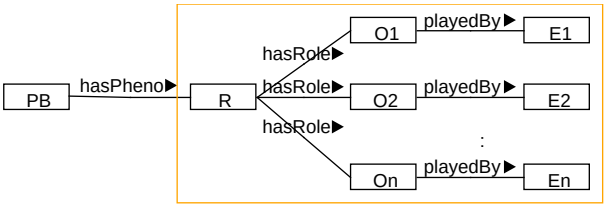

Figure 3. Roles-as-classes: Ontological roles modeled as classes in OWL.

ontological individuals akin to qualities, but with the power to mediate / connect entities. A relator consists of *role* individuals (via `hasRole` / `roleOf`) and each role individual, besides depending on the relator, depends on a *player* (via `playedBy` / `plays`). The term 'player' is relative to this approach; in general, arbitrary entities can play a role within a relation. At the categorial / class level, each relation R is associated with a set of role categories that forms the *role base* for this relation. Basically, that means for each relator of type R that its roles must instantiate one of the role categories in that set, cf. [22, sect. 3.3.3].

The GFO model of relations and roles can be encoded into an OWL representation in two obvious ways, termed *roles-as-properties* (Fig. 2) and *roles-as-classes* (Fig. 3). Common to both cases is to represent phenotype descriptions involving a relation R and (kinds of) entities $E_1, \ldots, E_n$ as argument restrictions. Either, corresponding to Fig. 2, roles are left implicit in the OWL properties $o_1, \ldots, o_n$, or, regarding Fig. 3, role categories are explicated as OWL classes $O_1, \ldots O_n$ (in between R and the E_i). Consider the example of *iron concentration in the spleen*, with the relation *concentration* and assuming that its two role categories are labeled *concentrated* (played by those entities concentrated in another) and *concentrator* (played by those entities within which another entity is concentrated). Then roles-as-properties yields in OWL

```
hasPheno some ( concentration and
                (concentrated some iron) and
                (concentrator some spleen) ),
```

whereas roles-as-classes leads to

```
hasPheno some ( concentration and
 (hasRole some (concentrated that playedBy some iron)) and
 (hasRole some (concentrator that playedBy some spleen)) ).
```

The first of these cases equals the above approach of using variable, relation-specific names for the auxiliary relations [34, sect. 5.1], [28, 1]. The second uses only two OWL properties `hasRole` and `playedBy` (and their inverses, possibly), but here this is unproblematic because the roles of the reified relation explicitly account for what is missing with fixed auxiliary relations without roles. Of course, both of these proposals will require a syntactic extension of the EQ model in order to capture the corresponding roles within EQ statements. Moreover, the roles-as-properties way may be simpler to reinterpret in other top-level ontological theories, because the roles presupposed by GFO are less explicit compared to roles-as-classes.

3.3.3 Ontological Alternative Using Relations and Qualities
The previous subsection suggests two ontologically inspired ways of understanding relational qualities like *concentration of* (PATO:0000033, hereafter CO) in EQ statements that cure the immediate deficiencies previously described. Both are based on a purely relational reading of CO (and relational qualities, in general), i.e., CO is merely considered as a noun form of the phrase *is concentrated in* (CI). For example, '(a particular amount of) *iron* I is concentrated in a (particular) *spleen* S' is a "relational proposition", stating that I is concentrated in S. This proposition can be true of false, depending on whether the relation CI applies to I and S or not, but there is nothing to be measured (neither quantitatively nor qualitatively).[5] In noun form, yet somewhat artificially, one may equivalently refer to 'there is concentration of I in S' (note that I and S are particulars).

However, we hold that CO comes in a second flavor, which is more amenable to specialization with notions like *increased concentration of* or to expressing specific values, e.g., $0.5g/l$. In phrases like 'the concentration of X in Y is $0.5g/l$, it appears more adequate to us to view CO as a proper quality which can be numerically quantified. Of course, immediately the question arises what that quality inheres in, which must be something that "includes" X and Y, not only one of the two. Here, computing the value of CO is instructive, which is based on values of qualities inhering solely in either X or Y, say, the weight of X and the volume of Y. The relationship between X and Y (of type CI, say) is characterized by the value within the CO phrase (in the second reading). Therefore, our current attempt of capturing relational qualities according to this analysis is to view them as inhering in particular relators, say a CI relator between X and Y. Admittedly, this is a deliberate, but no imperative choice among the possibilities within GFO. Other candidates for bearers of these qualities would be the overall relational fact, or one might consider the mereological sum of X and Y, in analogy to the inherence of relators in [11, sect. 6.2.7]. Regarding implementation in OWL, though, note that neither facts nor mereological sums are readily available on the basis of relators/relations and their arguments.

Eventually we arrive at a third approach, depicted in Fig. 4, where the relation is characterized by a quality. In the example, that means that CI is distinguished from CO, the latter being understood as a quality that inheres in CI relators / instances. Accordingly, we refer to this approach as *relator-based-quality*. Note that the intuitive term 'relational quality' experiences a formal-ontological reinterpretation from relations in the previous cases roles-as-properties and roles-as-classes to qualities proper (which are not relations) in the relator-based-qualities approach. Looking again at *iron concentration in the spleen*, assuming the roles-as-properties approach for modeling a relation `isConcentratedIn` (with roles like above) and a relational quality `concentration` yields in OWL

[5] Pursuing this line of thought further in the example, one may wonder what remains as the actual difference between CI and relations like 'is contained in' and 'is part of'.

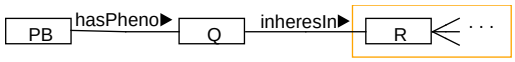

Figure 4. Relator-based-qualities: Relators characterized by qualities.

```
hasPheno some ( concentration that
        (inheresIn some (isConcentratedIn and
                         (concentrated some iron) and
                         (concentrator some spleen))) ).
```

This approach appears ontologically plausible to us currently, following the explanations above. Moreover, from the point of view of representation, it exhibits the beneficial property that CO is a "unary quality" like *color*, in the sense that it inheres in a single entity (a CI relator, which in turn accounts for the relational character of the quality). Any general account of qualities and quality values should thus be applicable to CO as it is to qualities like *color*. Furthermore, linking qualities to relators does not prescribe an overly specific relation model, but allows for adopting either of the approaches roles-as-properties and roles-as-classes in formalizing relations and roles, or even other theories (for which the quality bearer may require re-inspection).

4 DISCUSSION

Due to spatial limitations we focus the subsequent discussion mainly on aspects of the enhancements for relational qualities, where Table 1 compactly summarizes the approaches herein. Only minimal coverage of the introduction of the `hasPheno` relation can be given here. The latter is inspired by, but deviates from the notion of *phenes* and the *hasPhene* relation in [19]. Phenes may be understood as quality-like entities that reflect / abstract complex aspects that an organism is involved in. Accordingly, one immediate difference is that phenes are additional entities regarding those reflected aspects, whereas `hasPheno` bridges directly to one of the entities within those aspects. In any case, further comparison of the strengths and weaknesses of both views is a future task.

In connection with the general annotation-oriented interpretation, all three approaches for an improved account of relational qualities are designed to satisfy the issues identified in sect. 2.2, possibly varying in their degree of ontological adequacy. Concerning major disadvantages, clearly, all cases lead to significantly greater complexity of the representation through a considerable extension of vocabulary elements (see Table 1 for details). Concerning the "style" of reification embodied in roles-as-properties and roles-as-classes, there are also further unintended *technical issues*, surveyed in [10, sect. 2.2] (only with respect to roles-as-properties). At least in terms of reasoning, more precisely consistency checking and verifying entailments, those technical issues present no negative effects. Ibidem a number of potential *modeling shortcomings* are presented, in brief: (1) impeded manageability of the ontology, (2) purely technical nature of the additional vocabulary elements or at least an unclear ontological status, and (3) modeling diversity due to arbitrary splittings of reified relations, e.g. of reifying a 6-ary relation in terms of two ternary ones.

We disagree with all of these, yet to different degrees. Concerning (1), we agree that more vocabulary is involved which requires additional attention in *ontology maintenance*. But this can be countered by the mutual disjointness of relation, role, and non-relational classes and the use of distinct subsumption hierarchies / graphs for each category, within which relations, roles, and other classes can be organized manageably. Extra effort that remains is to determine role names for each relation when introducing the latter, which is a source of inter-modeler differences.[6] The *use of the ontology* may be less affected, if there are effective intermediate representations and user interfaces, cf. [25, p. 1]. (2) is wrong in the light of the GFO approach to relations and roles, where these are ontological entities and thus not of purely technical nature.[7] Criticism (3) appears not applicable in our case, because the reification directly uses roles instead of arbitrary k-ary "parts" of an n-ary relation (where $k < n$).

Moreover, we see significant advantages in modeling and expressiveness that arise from the use of roles. For instance, relations are not only unconstrained in the number of arguments per relation, but one may even use anadic relations (i.e., with a variable number of arguments) and such with optional arguments. Similarly, symmetry properties of relations derive naturally from allowing for multiply instantiable role categories in the context of a role base. That means, a relation may be instantiated by relators that have several individual roles instantiating the same role category.

Notably, it is also symmetry of this kind that produces doubts on the treatment of concentration in [13, sect. 3.2]. Hastings et al. present a fairly detailed analysis of substance mixtures (among other topics) which we can follow to a large extent. This analysis is aimed at formalizing the notion of concentration in description logics. In this connection and transferred to the original EQ model (see sect. 1, 2),the consistency of domain modeling is achieved – for concentration only – by simply declaring that concentrations inhere in the entity, say `calcium`, that is concentrated in another, say `blood`. This likely means for EQ that the concentration is linked to that other entity by means of `towards`, and thus `concentration that inheresIn some calcium and towards some blood` is the

Feature descriptions, followed by feature matrix:

A	role information	E	max. nr. of relevant vocabulary elements (fixed / per n-ary relation)
B	unlimited arity of relations	F	add. characterization of relations
C	variable arity of relations		
D	straight-forward database support		

Feature	EQ	RP	RC	RQ
A	no	yes	yes	yes
B	no (yes)	yes	yes	yes
C	no	yes	yes	yes
D	yes	no (?)	no (?)	no (?)
E	2 / 0	0 / $n+1$	2 / $n+1$	$X+1$ / $Y+1$
F	no	no	no	yes

Table 1. Summary of the main features of the discussed approaches (EQ: entity-quality, RP: roles-as-properties, RC: roles-as-classes, RQ: relational-quality). Entry (B,EQ) reflects the discussed extensibility of EQ. X, Y stand for the respective numbers of the RP or RC columns, depending on the relation model combined with RQ.

[6] However, one may adopt linguistic principles in some cases. E.g., for binary relations that can be appropriately named by verbs, participles can be used as rolenames in many cases. E.g., if *concentration of* (PATO:0000033) is traced back to *to concentrate*, the role(name)s of the *concentrated* and the *concentrating* may be formed.

[7] Admittedly, the roles-as-classes approach is closer to the ontological view of GFO, whereas roles-as-properties is a mainly technical simplification of the former. But this is not the technical nature critized in [10].

preferred formalization, cf. sect. 2.2.[8] In their analysis, however, this choice itself is not explained. Considering other relational properties than concentration, an analogous decision would have to be made for each relational property (and established among modelers), which appears less attractive than finding more general rules. Closing the circle to symmetric relations, for these it is not possible to distinguish one of the arguments (at least, not based on their roles only). For instance, for a phenotype like *increased distance of the eyes*, it appears completely implausible to select one eye in which a distance `inheresIn`, whereas it is `towards` the other eye. Especially the relator-based-quality approach, despite its own unresolved choices, see sect. 3.3.3, avoids such arbitrary fixing.

A practical deficiency of all three approaches that might be of potential importance is that the increased complexity prevents a straight-forward integration of corresponding annotations into the relational schemas of annotating databases. However, we have not yet explored alternatives in this connection, and this problem may re-occur due to an in-principle incompatibility of various aims, including the provision for n-ary relations vs. simple database implementation.

5 CONCLUSION

In this paper we report on the (work-in-progress) state of our analyses and improvement proposals concerning the Entity-Quality (EQ) model. A simple general modification in the understanding of qualities in PATO is argued to be necessary. Moreover, three variants of extended support for relations / relational qualities are presented.

Much work remains to be done or completed. The approaches detailed herein rely on theoretical analyses thus far. For further assessment, an experimental evaluation should be conducted, e.g. exploring the efficiency of reasoning over ontologies which rely on one or another approach. Despite our (preliminary) decision to minimize changes to the EQ interpretation to the greatest possible extent, we still see many interesting open theoretical issues in the EQ model, respective ontologies, and phenotype understanding and representation in general. For instance, we are convinced that not all concepts of PATO should be regarded ontologically properly as qualities. The not yet elaborated connections between `hasPheno` and `hasPhene` in [19] are named above. Accordingly, further alternatives, which possibly involve larger re-interpretation of existing resources, should be studied and compared. On that basis EQ syntax extensions and possibly changes to phenotype ontologies can be devised.

ACKNOWLEDGEMENT

We are grateful to the reviewers for constructive criticism and for additional pointers to the literature.

REFERENCES

[1]Nary relationship. `http://www.gong.manchester.ac.uk/odp/html/Nary_Relationship.html`, 2009.

[2]Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, and Gavin Sherlock. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29, May 2000.

[3]Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, Cambridge (UK), January 2003.

[4]James P. Balhoff, Wasila M. Dahdul, Cartik R. Kothari, Hilmar Lapp, John G. Lundberg, Paula Mabee, Peter E. Midford, Monte Westerfield, and Todd J. Vision. Phenex: Ontological annotation of phenotypic diversity. *PLoS ONE*, 5(5):e10500.1–10, 2010.

[5]Tim Beck, Hugh Morgan, Andrew Blake, Sara Wells, John Hancock, and Ann-Marie Mallon. Practical application of ontologies to annotate and analyse large scale raw mouse phenotype data. *BMC Bioinformatics*, 10(Suppl 5):S2, 2009.

[6]Judith A. Blake, Carol J. Bult, James A. Kadin, Joel E. Richardson, Janan T. Eppig, and the Mouse Genome Database Group. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Research*, 39(suppl 1):D842–D848, 2011.

[7]C. J. Bult, J. T. Eppig, J. A. Kadin, J. E. Richardson, and J. A. and Blake. The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic acids research*, 36(Database issue), January 2008.

[8]Francis S. Collins, Richard H. Finnell, Janet Rossant, and Wolfgang Wurst. A new partner for the international knockout mouse consortium. *Cell*, 129(2):235, 2007.

[9]Georgios V. Gkoutos, Eain CJ Green, Ann-Marie Mallon, John M. Hancock, and Duncan Davidson. Using ontologies to describe mouse phenotypes. *Genome Biology*, 6(1):R8, 2004.

[10]Niels Grewe. A generic reification strategy for n-ary relations in DL. In Herre et al. [17], pages N.1–5.

[11]Giancarlo Guizzardi. *Ontological Foundations for Structural Conceptual Models*, volume 015 of *Telematica Instituut Fundamental Research Series*. Telematica Instituut, Enschede (Netherlands), 2005. also: CTIT PhD Series No. 05-74.

[12]Giancarlo Guizzardi. On the representation of quantities and their parts in conceptual modeling. In Anthony Galton and Riichiro Mizoguchi, editors, *Formal Ontology in Information Systems: Proceedings of the Sixth International Conference, FOIS 2010, Toronto, Canada, May 11-14*, volume 209 of *Frontiers in Artificial Intelligence and Applications*, pages 103–116, Amsterdam, 2010. IOS Press.

[13]Janna Hastings, Christoph Steinbeck, Ludger Jansen, and Stefan Schulz. Substance concentrations as conditions for the realization of dispositions. In Ronald Cornet and Stefan Schulz, editors, *Semantic Applications in Life Sciences: Proceedings of the 4th International Workshop on Formal Biomedical Knowledge Representation, KR-MED 2010, hosted by Bio-Ontologies 2010, Boston, Massachussetts, USA, Jul 9-10*, volume 754 of *CEUR Workshop Proceedings*, Aachen, Germany, 2010. CEUR-WS.org.

[14]Terry F. Hayamizu, Mary Mangan, John P. Corradi, James A. Kadin, and Martin Ringwald. The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome*

[8] At least, this is what we read from sect. 3.2 in [13]. It is not actually stated whether and how the concentration relates directly with the mixture, e.g. blood. One can also find more informal statements which suggest different interpretations, e.g. in sect. 2.2 of [13].

Biology, 6(3):R29, 2005.

[15]Heinrich Herre. General Formal Ontology (GFO): A foundational ontology for conceptual modelling. In Roberto Poli, Michael Healy, and Achilles Kameas, editors, *Theory and Applications of Ontology: Computer Applications*, chapter 14, pages 297–345. Springer, Heidelberg, 2010.

[16]Heinrich Herre, Barbara Heller, Patryk Burek, Robert Hoehndorf, Frank Loebe, and Hannes Michalek. General Formal Ontology (GFO) – A foundational ontology integrating objects and processes [Version 1.0.1]. Draft, Research Group Ontologies in Medicine, Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, 2007.

[17]Heinrich Herre, Robert Hoehndorf, Janet Kelso, and Stefan Schulz. Proceedings of the 2nd workshop of the gi-fachgruppe "ontologien in biomedizin und lebenswissenschaften" (obml): Mannheim, germany, sep 9-10, 2010. IMISE-Report 2/2010, IMISE, University of Leipzig, Leipzig, Germany, Sep 2010.

[18]Robert Hoehndorf, Michel Dumontier, Anika Oellrich, Dietrich Rebholz-Schuhmann, and Georgios V. Schofield, Paul N. an Gkoutos. Interoperability between biomedical ontologies through relation expansion, upper-level ontologies and automatic reasoning. *PLoS ONE*, in press, 2011.

[19]Robert Hoehndorf, Anika Oellrich, and Dietrich Rebholz-Schuhmann. Interoperability between phenotype and anatomy ontologies. *Bioinformatics*, 26(24):3112–3118, 2010.

[20]Robert Hoehndorf, Paul N. Schofield, and Georgios V. Gkoutos. PhenomeNET: a whole-phenome approach to disease gene discovery. *Nucleic Acids Research*, 2011. in press (advance access: `http://dx.doi.org/10.1093/nar/gkr538`).

[21]Degtyarenko K., P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcantara, M. Darsow, M. Guedj, and Michael Ashburner. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, 36(Suppl 1):D344–D350, 2008.

[22]Frank Loebe. An analysis of roles: Towards ontology-based modelling. Onto-Med Report 6, Research Group Ontologies in Medicine, University of Leipzig, 2003. Master's Thesis.

[23]Frank Loebe. Abstract vs. social roles – Towards a general theoretical account of roles. *Applied Ontology*, 2(2):127–158, 2007.

[24]Martin Mahner and Michael Kary. What exactly are genomes, genotypes and phenotypes? and what about phenomes? *Journal of Theoretical Biology*, 186(1):55–63, 1997.

[25]Chris Mungall, Georgios Gkoutos, Nicole Washington, and Suzanna Lewis. Representing phenotypes in OWL. In Christine Golbreich, Aditya Kalyanpur, and Bijan Parsia, editors, *Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions, Innsbruck, Austria, Jun 6-7*, volume 258 of *CEUR Workshop Proceedings*, Aachen, Germany, 2007. CEUR-WS.org.

[26]Christopher Mungall, Georgios Gkoutos, Cynthia Smith, Melissa Haendel, Suzanna Lewis, and Michael Ashburner. Integrating phenotype ontologies across multiple species. *Genome Biology*, 11(1):R2.1–R2.16, 2010.

[27]Fabian Neuhaus, Pierre Grenon, and Barry Smith. A formal theory of substances, qualities and universals. In Achille C. Varzi and Laure Vieu, editors, *Formal Ontology in Information Systems: Proceedings of the Third International Conference (FOIS-2004)*, volume 114 of *Frontiers in Artificial Intelligence and Applications*, pages 49–59, Amsterdam, 2004. IOS Press.

[28]Natasha Noy and Alan Rector. Defining N-ary relations on the Semantic Web. W3C working group note, World Wide Web Consortium (W3C), April, 12 2006. `http://www.w3.org/TR/2006/NOTE-swbp-n-aryRelations-20060412/`.

[29]P. N. Robinson, S. Koehler, S. Bauer, D. Seelow, D. Horn, and S. Mundlos. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *American journal of human genetics*, 83(5):610–615, 2008.

[30]Nadia Rosenthal and Steve Brown. The mouse ascending: perspectives for human-disease models. *Nature Cell Biology*, 9(9):993–999, 2007.

[31]Gary Schindelman, Jolene Fernandes, Carol Bastiani, Karen Yook, and Paul Sternberg. Worm phenotype ontology: integrating phenotype data within and beyond the c. elegans community. *BMC Bioinformatics*, 12(1):32, 2011.

[32]Stefan Schulz, Daniel Schober, Djamila Raufie, and Martin Boeker. Pre- and postcoordination in biomedical ontologies. In Herre et al. [17], pages L1–L4.

[33]Cynthia L. Smith, Carroll-Ann W. Goldsmith, and Janan T. Eppig. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biology*, 6(1):R7.1–R7.9, 2004.

[34]Robert Stevens, Mikel Egana Aranguren, Katy Wolstencroft, Ulrike Sattler, Nick Drummond, Matthew Horridge, and Alan Rector. Using OWL to model biological knowledge. *International Journal of Human-Computer Studies*, 65(7):583–594, July 2007.

[35]Alexandr Uciteli, Silvia Groß, Sergej Kireyev, and Heinrich Herre. An ontologically founded architecture for information systems in clinical and epidemiological research. *Journal of Biomedical Semantics*, 2(Suppl 4):S1, 2011.

[36]W3C. OWL 2 Web Ontology Language Document Overview. W3C Recommendation, World Wide Web Consortium (W3C), Cambridge (Massachusetts), 2009. `http://www.w3.org/TR/2009/REC-owl2-overview-20091027/`.

[37]Nicole L. Washington, Melissa A. Haendel, Christopher J. Mungall, Michael Ashburner, Monte Westerfield, and Suzanna E. Lewis. Linking human diseases to animal models using ontology-based phenotype annotation. *PLoS Biology*, 7(11):e1000247.1–20, 2009.