

Ontology usage in Omics Standards Initiatives: Pros and Cons of enriching XML data formats with controlled vocabulary terms

Daniel Schober^{1*}, Michael Wilson², Daniel Jacob³, Annick Moing³, Gerhard Mayer⁴, Martin Eisenacher⁴, Reza M Salek⁵, Steffen Neumann¹

¹Leibniz Institute of Plant Biochemistry, Dept. of Stress and Developmental Biology, Weinberg 3, 06120 Halle, Germany

²Department of Computing/Biological Sciences, University of Alberta, Edmonton, Canada

³INRA, Univ. Bordeaux, UMR1332 Fruit Biology and Pathology, Metabolome Facility of Bordeaux Functional Genomics Center, MetaboHUB, IBVM, Centre INRA Bordeaux, 71 av Edouard Bourlaux, F-33140 Villenave d'Ornon, France

⁴Medizinisches Proteom Center (MPC), Ruhr-Universität Bochum, D-44801 Bochum, Germany

⁵European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

EMAIL-ADDRESSES

| | |
|------------------|--|
| Daniel Schober: | dschober@ipb-halle.de |
| Michael Wilson: | michael.wilson@ualberta.ca |
| Daniel Jacob: | djacob65@gmail.com |
| Annick Moing: | moing@bordeaux.inra.fr |
| Gerhard Mayer: | mayerg97@rub.de |
| Reza M Salek: | reza.salek@ebi.ac.uk |
| Steffen Neumann: | sneumann@ipb-halle.de |

ABSTRACT

Motivation:

We here review a method of XML data enrichment with controlled vocabularies (CV) in light of end-user compliance. We outline the reasons that made major standard initiatives in proteomics and metabolomics use this data enrichment scheme on omics data in favor of more formal approaches, e.g. description logics (DL) knowledge bases. We show that in comparison to other knowledge representation formalisms, the list of prerequisite skills on the user-side and the learning threshold is significantly lower, making the approach feasible for bioinformaticians with average skill levels, i.e. basic XML knowledge. Additionally our approach allows to source out the 'business logics' from the terminology into external rules. This enables the successive and encapsulated addition of semantics in a flexible way.

We feel our approach contributes to increase the amount of potential users, enabling them to participate in a peer-produced standards development process.

1 INTRODUCTION

After the very successful introduction of the Gene Ontology serving the genomics community with a convergent terminology, recent years saw the successful launch of omics standardisation initiatives such as the Proteomics Standards Initiative (PSI) and the Metabolomics Standards Initiative (MSI, Sansone 2007). Open data standards are needed in these domains, as an ever growing mountain of data is piling up due to abundant usage of high throughput data generators, i.e. Mass Spectrometry (MS) and Nuclear Magnetic Resonance spectroscopy (NMR).

The COSMOS (COordination of Standards in MetabOlo-mics) EU consortium¹ was tasked to foster the creation of an

* To whom correspondence should be addressed.

¹ <http://www.cosmos-fp7.eu/>

MSI-approved open exchange and storage standard² for metabolomics NMR data. It decided to leverage on a particular set-up proven already successful in the PSI community. To improve compliance, PSI and MSI consciously refrain from employing DL in their CV directly. Instead, they leverage on a user-friendly and pragmatic method of solely using the taxonomic backbones of ontologies in whatever format available to augment values in simple XML files. This allows shifting semantic constraints into the XML Schema definition (XSD) and easy to read XML rules. Although this set-up and methodology has already been described (Mayer 2013), the reasons that led the OMICS standardization bodies to prefer this set-up over a plethora of alternative knowledge representation (KR) formats have so far been largely implicit.

After briefly recapitulating the basic set-up, we elaborate on the reasons for this pragmatic CV referencing approach by highlighting the requirements with end-user compliance in mind. We discuss the pros and cons of alternative KR formalisms such as Description Logics (DL) and Frames to fulfill the same requirements.

2 MATERIAL AND METHODS

We discuss CV usage as occurring in the PSI and MSI XML data standards in light of different knowledge representation schools³, namely pragmatic life science users employing GO style taxonomies, their arguments probably best summarized in (Bada 2004) and academic formal logics ontologists creating DL-based axiomatised ontologies, their arguments probably best summarized in (Schulz 2013). We derive requirements from our use cases and compare key features needed to fulfil these requirements for our XML+CV approach with OWL-DL and Frames.

3 RESULTS

The COSMOS Standards workpackage developing the nmrML data standard and nmrCV, had several overarching goals guiding decision making. Based on the use cases and derived requirements, we here list the distinct features of the three compared KR approaches with respect to the defined requirements (Tab 1).

3.1 CV Referencing from within XML data files

The nmrML XML format, inspired by the PSI mzML format (Mayer 2013), consists of an XML Schema Definition (XSD) that is instantiated and accompanied by CV annotations in a concrete XML data file. The XSD defines the allowed XML elements, their attributes, cardinalities and mandatoryness etc. The requirement and modality for a CV

term occurrence in an XML instance is specified in the XSD by *reference elements/types*. The general approach can be seen as adding outsourced semantics to XML⁴. At certain locations specified in the XSD, the user is allowed to describe his data by <CVParam> tags, where attributes reference the standardized CV terms. The CV provides the terminology to describe the data in detail and provide standardized values for the XML tags. For example, the XSD defines a ValueWithUnitType reference to hold a value and a description of the unit the value is recorded in by means of a Unit Ontology CV term. An example XML code specifying the temperature of 30 °C of a given NMR sample XML looks like this:

```
<sampleAcquisitionTemperature unit-
Name="degree Celsius" unitCvRef="UO" val-
ue="30" unitAccession="UO:0000027"/>
```

In areas where the terminology is likely to change faster than the nmrML XSD can be updated and aligned, branching out from the XSD to CV usage can compensate for such dynamics in a flexible way, as the CV can be maintained externally and even in a decentralized and peer-produced⁵ manner. For example, new NMR probe types can be represented in an nmrML file by requesting and adding a new CV term for the unchanged XML element, without the need for any XSD revision, which in turn would require to also update programs using nmrML. The XSD ‘branches out’ into CV-usage where:

- Terms are unstable & dynamically evolving, or need to be changed and updated often, such as Hard- & software names/versions etc.
- Terms are lexically variant and need convergence via synonym equivalence detection
- Terms describe contextual metadata, rather than concrete NMR raw data, i.e. for cases where the terminology is already extensively defined in existing ontologies or CVs, e.g. the unit ontology.
- Terms represent important search attributes for data querying; this will ease large scale database-integration in an open linked data fashion.
- Terms should be accessible to rule-based or external DL reasoning techniques for ontology audit, validation and querying, e.g. to profit from subsumption to generalize over query attributes and increase result recall and precision.

3.2 Data Validation

The XSD+CV set-up allows for multiple data validation levels to be established in an onion layered approach, contributing to data consistency, completeness and overall quality assurance.

² <http://nmrml.org>

³ <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2447479/pdf/CFG-05-623.pdf>

⁴ <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.3250>

⁵ http://en.wikipedia.org/wiki/Peer_production

Table 1. For each Cosmos use case the format dependent requirements are listed and accompanied with pros and cons of potential implementations in the different KR set-ups.

| Use case | Requirement | XML+CV | Owl-DL | Frames |
|---|---|--|---|--|
| Extendability of Format to cover new domain developments | Update and versioning flexibility, timeliness of required additions. | Fast as no modularization and external coordination for e.g. term requests needed. Full namespace control, but not orthogonal to existing artefacts. Flexible, as no schema updates needed for new CV term addition. | Ontology extension slow due to complex axiomatisation and scattered term requests to multiple external sometimes slow reacting authorities. | Very fast, as all entities live within one KB that already comes with Knowledge Acquisition (KA) forms. |
| Increase data persistence and traceability | Robust storage format for long-term archiving yet flexibility for encoding new information yet keeping old data valid ⁶ . | Good, as only stable formats at basis of semantic web stack are used e.g. XML and OWL expressivities that are robust as well. Even RDFS would be sufficient here. | Not so good, as OWL specifications are still dynamic, i.e. recent introduction of OWL 2 with change from OWL 1 flavors to OWL 2 profiles. | Very good, as Frames are robust and Protégé 3 is the oldest, still widely supported KB editor in existence. |
| Allow high throughput generation of large data sets, i.e. from existing databases | Must be able to work without expensive human intervention. | Possible via available parsers employing mapping specifications. Good support for datatypes. | Possible on T-Box through ODPs and Quick Term Templates, but again need human intervention. Technologies still emergent. | Easy via Plugins like Data-master ⁷ . Easier due to classic Object Oriented approach & good support for datatypes. |
| Quality assurance | Allow automatic data validation on semantics/content correctness and completeness, i.e. ensuring Journal policies or MIBBI information standards (Taylor 2008). | Very good, incl. data level, as constraints can be checked at multiple levels with established tools, e.g. XML parser and XSD compliance checks with good access to datatypes. Semantic rules to check CV-based annotations. | Good, but restricted to small T-Boxes as tableau reasoners for consistency checking are slow. Top down rule based reasoners ⁸ are an option as they only evaluate local query-relevant parts of the T-Box. | Good and fast via P3 build in real-time constraint checks and PAL constraints ⁹ . All readily available in one Editor. |
| User guidance & error prevention for data acquisition | Self-explanatory, suggested datatypes, data entry constraints. | Good, as XML mapping rules, semantic validators as well as ISA specifications can be exploited to drive entry recommendations/constraints. Biportal support tools serve precoordinated terms. | Possible, but labor intensive and requiring full DL skillset. Postcoordinated anonymous classes hardly accessible, due to slow DL reasoning performance. | Very good due to ability to specify and check KA forms and default and allowed slot fillers/ own datatypes. |
| Build up large user community | Easy to parse and integrate with available tools. Intelligible to bioinformaticians & computers. Allow re-use of existing best practice guidelines. | As solely widely established and well investigated formats and technologies are used, there is abundant skill available and existing parsers can be re-used. Each KR involved is understood by average bioinformatician. | Compliance issues due to need to learn formal logics and set theory to understand DL semantics. Recent analysis showed difficulty in measuring success of a week of specialist DL training (Boeker 2013). | Plethora of robust tools available. Quasi standard since 30 years. Very large user community, but comparatively small compared to XML community. |
| Easy i/o to Public repositories, to share and support integrated analysis | Formats need good database connectivity, i.e. data should be easy to transit from tabular/relational to KR form. | Easy, as DTB to XML transitions are well understood and plethora of parsers exist. | Less easy, as DLs set theoretic nature and OWA make transition from RDB data to DL KB more difficult. | Transition fairly easy due to Frames' CWA, non-monotony and unique name assumption. |

Table 2. Comparison of different KR Methods

| Issue | XSD+CV | OWL-DL | Frames |
|----------------|---|---|------------------------------|
| Editor | XML Editor | OWL Ontology Editor (Protege4) | Ontology Editor (Protege3.x) |
| Syntax | XML | RDF, OWL | PINS & PONT |
| Semantics | Term hierarchy/DAG | DL, set theory | Frames, OOM, CLIPS |
| Constraints | XSD, Rules | Axioms (constructors, domain, range) | Facets, PAL constraint |
| Main KR Idioms | Entity, Attribute, CV Term, Unit, Value | Class, Individual, object property, datatypes | Concepts, slots, datatypes |

⁶ There is a strong desire from companies that develop software to keep the data format /XSD stable over time.

⁷ <http://protegewiki.stanford.edu/wiki/DataMaster>

⁸ <http://protegewiki.stanford.edu/wiki/NoHR>

⁹ http://protegewiki.stanford.edu/wiki/Protege_Axiom_Language_%28PAL%29_Tabs

In OWL-DL it is difficult to set distinct stringency-levels on data consistency checks. Although the expressivity regime dictates the types of reasoning on the data, an OWL ontology is either totally consistent or inconsistent¹⁰.

Here, XML syntax and structural validity of XML instances, e.g. XML element and attribute positions, order and cardinality, can be validated by an XML parser against the XSD. Additional mapping rule files are used to enforce semantic validity by specifying which CV terms are allowed for an element as well as their order and cardinality. A dedicated semantic validator¹¹ checks that the criteria outlined by the mapping file are being met by a given XML instance. It enforces simple IF-THEN rules not only making sure that the terms are actually found in a specified CV, but also that the correct terms are used in correct locations (XML Tags) in the XML document and the required terms are present the correct number of times. E.g. that there are two filler values “Kelvin” or “Degree Celsius” allowed for the SampleTemperature-Element, and that these must come from the Unit Ontologies’ “temperature unit” subtree. This allows greater flexibility in the schema, but enforces order in how the CV terms are used. This will require the discipline of using the semantic validator exploiting validation rules for checking an annotation prior to storage and submission.

The result is that new technologies or information can be accommodated with adjustments to the controlled vocabulary and validator, not to the schema which hence can stay stable.

Validating the above XML data snippet with the following validation rule (HTML version)

```
Identifier:sampleAcquisitionTemperature_must
Element:/nmrML/acquisition/acquisition1D/
acquisitionParameter-
Set/sampleAcquisitionTemperature@unitAccessi
on
Requirement level: MUST
Term: UO:0000012 ! kelvin
```

hence would result in a validation error, as the temperature is specified in degree Celsius rather than Kelvin. The mapping file combined with the CV can also be used for intelligent support in data acquisition, i.e. when creating an interface that records NMR experiment information it can populate a drop down menu or an autocomplete box with plausible entries.

3.3 Comparison of skills required by approaches

Table 2 summarizes the features leveraged on by the biomedical data curator applying the different KR approaches for data annotation and storage. In amendment to Tab 1, it also highlights some of the skill levels required on the end-user side in comparison.

¹⁰ often violating whole branches, even though subsets might be correct.

¹¹ <http://nmrml.org/validator/>

4 DISCUSSION

A large part of current discussions in the biomedical ontology field is devoted to OWL and description logics (DL) semantics. Beside its computational performance issues for tableau reasoning in big data set-ups, DL semantics is rather complex and requires users to get acquainted with its set theoretic basis. These characteristics have hindered widespread application of DL expressivity in high throughput data exchange and annotation. Although strategies for simplifying DL complexity were discussed in an earlier paper (Schober 2010), DL expertise is still not as abundant as needed for larger peer-produced formal ontology creation and exploitation. Although through its recommendation by the W3C, OWL became a major player even in the semantic web domain, many people do not leverage on DL expressivity¹² in OWL ontologies.

DL was never good at data-capture (Schulz 2010) and no easy-to-use knowledge acquisition (KA) forms that constrain user entries to allowed datatypes are available, e.g. like in Frames. Such tools¹³ are only now emerging.

Moreover, one can often find DL axiomatised class definitions which put the burden of capturing many properties of an individual or class which serve proper classification and reasoning, but are trivial and uninteresting for the scientific life science domain end user. We doubt that axioms that are good for consistency checking and untangling, are necessarily good for providing usable search attributes. We found that information retrieval is in fact often not well supported by DL axioms, e.g. in the following OBI example

```
Manufacturer SubclassOf MaterialEntity and
equivalentTo
('Homo sapiens' or organization)and
('has role' some 'manufacturer role')
```

the word-cloud that comes intuitively to the end users mind when mentioning ‘manufacturer’ is not well matched, e.g. the axiomatisation misses out on important aspects like product lists, contact details, location, etc. to be expected in a data centric environment. Instead we are informed that a manufacturer has a role manufacturer role, a statement which will seem trivial, if not tautological, to the end user. Such DL class definitions, exhibit dependencies to classes for which real world owl:individuals are hard to imagine e.g. what is the instance of a role ? Such individuals, entered into a knowledge base solely for the sake of fulfilling design patterns for DL reasoning will rather cause performance problems than supporting terminological end user requirements.

¹² A contributing factor might be that the OWL specification for DL was less stable than XML and underwent transitions from OWL 1 flavors, over OWL 2 ‘profiles’ to expressivity regimes.

¹³ <http://www.isa-tools.org/tools.html>

Another example of modeling against the users needs, not meeting his requirements is

```
Decapitated organism equivalentTo
'material entity' and
(is_specified_output_of some decapitation)
```

, which misses out on the fact that the organism is now consisting of two entities, cut along the head-neck joint, a separated rump with extremities and a head. Also the fact that it is now a dead organism with all its practical and ethical consequences, is missing in this reductionist definition. In DL, such definitions rely on this information – although being important – being drawn from sometimes distant related entities in the ontology by the reasoner, e.g. the fact that this definition applies to a living organism is entailed in the decapitation process, whereas the definition itself subclasses material entity. Leveraging only on the local axioms in the definition, it would not exclude a decapitated cigar or teddy bear, as some of this essential information is factored out elsewhere, i.e. here into superclasses of the decapitation process. This makes the classes very dependent on the contextual model and DL reasoning and hence renders it less usable when wanting to exploit the self-standing local class definition, i.e. when annotating a database entry.

To reach domain coverage in an acceptable time in our CV we simply pre-coordinate the terms according to appearance in our use cases. In formal DL ontologies, modularization policies demand that for each simple class addition of e.g. “Library based computed concentration” its refactored semantic Lego blocks are requesting in external artefacts, i.e. Unit Ontology (concentration), Evidence Ontology (computed), and maybe Information Artefact Ontology (Library). So, orthogonality assurance, besides domain-borders being difficult to determine, often requires time-costly requests for term additions of unavailable terms in external artefacts, an obstacle we avoid in our pragmatic CV setup.

Frames have a long and successful history in AI and with good support of configurable data types and KA forms they are ideally suited for data capture and *drag and drop* conceptual annotation (Schober 2005). Its intuitive object-oriented top down inheritance, e.g. of default slots with values allows for many of the requirements outlined above to be met in one monolithic and coherent format¹⁴. Frames and our rule based approach are also closer to the general procedure of human knowledge generation, as due to its non-monotonic reasoning ability, it allows a KB to be expanded slowly by adding new knowledge with effect on the truth value of already inferred statements. Such ability is only now emerging in the OWL-DL field¹⁵. The Protégé 3

editor is stable and besides data capture and editing, provides a plethora of plugins for processing, validation, querying and visualization. However this virtue can also be seen as burden, as although one single editing environment can handle all needed KR tasks, the GUI becomes rather complex. Non-programmers have difficulties to understand the object oriented modeling principle and users need to learn OKBC-CLIPS, a non-W3C supported KR language.

Compared to DL-based approaches, our approach relies on simple taxonomic CVs, preferring intuitiveness over formal rigidity and making no assumptions on any end user skill other than XML and what an *is-a* hierarchy of terms is.

Our approach has the advantage of being an addition that complements parallel DL approaches applied in the background i.e. for T-Box-reasoning-enabled ontology maintenance tasks, rather than validation through A-Box reasoning on data. In an additional reverse approach, lexical patterns in term names can still be exploited to support DL axiomatisation of CVs. However, efforts to axiomatise the GO taxonomy were perceived as a very time consuming endeavor (Wroe 2003). As DL tools become more mature and computational power increases, classification, consistency checks via T-Box tableau reasoning can later still be done on such axiomatised DL ontologies.

The approach to outsource completeness and consistency checks into XML rules suits large data sets better, as rule systems are faster than the slow DL tableau reasoning approaches. Besides performance, they are also better coping with unfinished and incomplete/missing data as rules act rather locally, whereas one distant missing axiom can render a whole DL KB branch invalid.

On another line, the separation of CV term creation and their formal semantic definition and validation can be seen as outsourcing the ‘business logic’ from the data layer, enabling more freedom on the end user side. Conflating the validation layer with the terminology layer makes the overall artefact not only hard to grasp for end users, but it also decreases shell-wise processing as all KR idioms are inherently dependent on and interrelated with each other.

Our approach is also rather flexible with datatype specifications, i.e. allows capturing value-unit pairs.

We propose to let the pragmatic life-science users and academic formal logics ontologists do what they are best in, and leverage on the best of both worlds, i.e. by connecting them in a two-tiered hybrid approach along what was already proposed in Cornet (2005). This way, CVs in our simple set-up can leverage from DL axiomatised top level resources, even when our nmrML itself stays unaxiomatised, e.g. we imported and subclassed an axiomatised top level ontology (TLO). Running a DL reasoner on nmrCV.owl, a class ‘chemical compound formula’ was detected as inconsistent, because ‘chemical compound formula’ was asserted as ‘information object’ AND ‘chemical compound attribute’ (a subclass of ‘quality’), but the TLO

¹⁴ It is interesting to note that this formalism is gaining attraction again, e.g. see this conferences OntoStudyEdit submission by A. Uciteli and H. Herre

¹⁵ http://www.aifb.kit.edu/web/Epistemic_Reasoning_in_OWL_2_DL

declared 'quality' and 'information object' as disjoint¹⁶. But we ask again, would any such loss of distinction of being a quality vs an information content entity really be a source of erroneous scientific interpretation with impact on our use case fulfilment? We feel these questions need a more thorough investigation.

5 CONCLUSIONS

The omics communities need efficient processing tools and results on available data today and cannot afford to wait until mature DL ontologies are available in sufficient coverage and granularity. These just take too much time to build and it is not clear that they capture what the end-users really need¹⁷. We have outlined justifications for a pragmatic approach for life science data standards, which shifts semantics from ontologies to rules and exemplified them by indicating how they contribute to end-user compliance. Our simple taxonomies/CVs can be built by the life scientists themselves in a peer production approach, ensuring fast growth and sufficient timely domain coverage. The nmrML standard is already sanctioned by the metabolomics standards initiative and is accepted by major open-source NMR data processing tools. It will serve the MetaboLights repository (Haug 2013), and other metabolomics data repositories, with a stable storage format. Also tools and parsers are easy to build as we exclusively leverage on established methods and formats. Many aspects of Minimum Information on an NMR experiment (MI NMR, Rubtsov 2007) can already be captured and validated with our set-up. The outsourcing of semantics into rule sets allows us to be flexible, e.g. having settings to validate either the basic nmrML raw data level or at a more comprehensive MI NMR compliance level, depending on the needs of the user. Specifying such stringency levels in a DL KB would be more difficult due to the inherently non-local behavior of the reasoning algorithms.

Positive feedback by vendors at the Metabolomics 2014 conference and the offer to support nmrML parsers in the ChenomX NMR suite hint for a good 'buy in' even from commercial companies, which often look at OWL-DL formalisms with suspicion.

To conclude, the XML/XSD based CV usage is well suited to meet the COSMOS requirements for an easy to understand data standard and to capture our experimental data, whereas OWL-DL better suited to formalize background knowledge, and aid in ontology maintenance and audit.

ACKNOWLEDGEMENTS

This work was financed via the EU FP7 project COSMOS grant EC312941. GM is funded by the Deutsche Gesetzliche Unfallversicherung (DGUV) project DGUV-Lunge (617.0 FP 339A). We thank Carol Goble and the Capulets from the PSI ontology working groups for inspirations.

REFERENCES

- Bada M, Stevens R, Goble CA, et al (2004) A short study on the success of the Gene Ontology. *J Web Semant* 1: 235–240
- Boeker M, Jansen L, Grewe N, Röhl J, Schober D, et al. (2013): Effects of Guideline-Based Training on the Quality of Formal Ontologies: A Randomized Controlled Trial. *PLoS ONE* 8 (5): e61425.
- Cornet R, Abu-Hanna A (2005): Description logic-based methods for auditing frame-based medical terminological systems. *Artif Int in Med*, 34(3): 201-217
- Haug K, Salek, RM, Conesa P, et al. (2013) MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucl acids res*, 41(Database issue), D781-786, doi:10.1093/nar/gks1004.
- Mayer G, Jones AR, Binz PA et.al. (2013) Controlled vocabularies and ontologies in proteomics: Overview, principles and practice. *Biochim Biophys Acta*. Feb 19. pii: S1570-9639(13)00080-0. doi: 10.1016/j.bbapap.2013.02.017.
- Rubtsov DV, Jenkins H, Ludwig C, Easton J, et al. (2007) Proposed reporting requirements for the description of NMR-based metabolomics experiments. *Metabolomics*, 3, 223-229.
- Sansone SA, Fan L, Goodacre R et al. (2007) The metabolomics standards initiative. *Nat Biotechnol*, 25, 846–848.
- Schober D, Boeker M (2010) Ontology Simplification: new buzzword or real need? *2nd Workshop of Ontologies in Biomedicine and Life Sciences (OBML)*, Mannheim, 9.-10.9. 2010). <http://www.bioontology.org/node/628>
- Schober D, Leser U, Zenke M, Reich JG (2005) GandrKB--ontological microarray annotation and visualization. *Bioinformatics* 21(11): 2785-2786
- Schulz S, Schober D, Daniel C, et al. (2010) Bridging the semantics gap between terminologies, ontologies, and information models. *Stud Health Technol Inform*; 160 (Pt 2): 1000-1004.
- Schulz S; Balkanyi L; Cornet, R; Bodenreider, O (2013) From Concept Representations to Ontologies: A Paradigm Shift in Health Informatics? *Health Inform Res.*; 19(4):235-242
- Taylor CF, Field D, Sansone SA, et al. (2008) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project, *Nat Biotechnol.*; 26(8):889-96. doi: 10.1038/nbt.1411. , PMID:18688244
- Wroe CJ, Stevens R, Goble CA, Ashburner M. (2003) A methodology to migrate the Gene Ontology to a description logic environment using DAML+OIL. *Pac Symp Biocomput* 8: 624–635.

¹⁶ <https://github.com/nmrML/nmrML/issues/62> The erroneous assertion of object attributes being a quality has now been removed and is now found under information objects.

¹⁷ The Nobel laureate Manfred Eigen states that 'a theory has only the possibility of being right or wrong. A model has a third possibility; it may be right but irrelevant.'