

[About EJBI](#)[Editorial Board](#)[Instructions to Authors](#)[Browse EJBI](#)[Special Issues](#)[Sponsorship & Ads](#)[Downloads](#)[Contact](#)

ISSN 1801-5603

German English

## Biomedizinische Ontologien für die Praxis

Brochhausen M.<sup>1</sup>, Smith B.<sup>1</sup>

1. Institute of Formal Ontology and Medical Information Science, Saarland University, Germany

2. Department of Philosophy and New York State Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, USA

### Abstract

*Hintergrund:* Biomedizinische Ontologien existieren unter anderem zur Integration von klinischen und experimentellen Daten. Um dies zu erreichen ist es erforderlich, dass die fraglichen Ontologien von einer großen Zahl von Benutzern zur Annotation von Daten verwendet werden. Wie können Ontologien das erforderliche Maß an Benutzerfreundlichkeit, Zuverlässigkeit, Kosteneffektivität und Domänenabdeckung erreichen, um weitreichende Akzeptanz herbeizuführen?

*Material und Methoden:* Wir konzentrieren uns auf zwei unterschiedliche Strategien, die zurzeit hierbei verfolgt werden. Eine davon wird von SNOMED CT im Bereich der Medizin vertreten, die andere im Bereich der Biologie und Biomedizin von der OBO Foundry. Es soll aufgezeigt werden, wie die Verpflichtung zu speziellen Kriterien der Ontologieentwicklung die Nützlichkeit und Effektivität der Ontologien positiv beeinflusst, indem die Pflege der terminologischen Systeme und ihre Interoperabilität vereinfacht werden.

*Ergebnisse:* SNOMED CT und die OBO Foundry unterscheiden sich grundlegend in ihren Ansätzen und Zielen. Unabhängig davon kann jedoch ein allgemeiner Trend zur strengeren Formalisierung und Fokussierung auf Interoperabilität zwischen unterschiedlichen Domänen und ihren Repräsentationen beobachtet werden.

Keywords: Biomedizinische Ontologien, Harmonisierung, Qualitätssicherung, SNOMED CT

### Einleitung

In der Entwicklung terminologischer Ressourcen für Anwendungen in der Biomedizin können wir vier wesentliche Einflussfaktoren unterscheiden, die hier in grober chronologischer Folge aufgelistet werden:

1. Der Einfluss von Bibliothekswissenschaften und elektronischen Terminologiesystemen, der sich deutlich zum Beispiel bei MeSH, einer Indizierungsressource, die von der National Library of Medicine gepflegt wird [1], zeigt.
2. Der Einfluss aus der Datenbankgestaltung und dem *conceptual modeling*, der durch HL7 illustriert wird [2].
3. Der Einfluss der Molekularbiologie, der sich am deutlichsten in der Gene Ontology (GO) [3] und anderen Ontologien der Open Biomedical Ontologies (OBO) Foundry [4], [5] nachverfolgen lässt.
4. Der Einfluss einer Orientierung hin zu stärkerer logischer Formalisierung. Dieser Trend zeigt sich bei den momentanen Entwicklungen in SNOMED CT und im Bereich des Semantic Web [6].

Im Folgenden befassen wir uns mit den Fortschritten, die im Hinblick auf die Punkte 3) und 4) gemacht wurden. Mehr und mehr können Ontologien und verwandte terminologische Ressourcen von ihren Vorgängern durch folgende Faktoren unterschieden werden:

- Die Beschäftigung mit Aspekten der Interoperabilität von Ontologien, die für inhaltlich zusammenhängende, doch unterschiedliche Domänen geschaffen werden.
- Die Bereitschaft, kohärente Verfahren zur Qualitätssicherung von Ontologien zu entwickeln, die auf dem Feedback durch Nutzer und auf empirischen Tests beruhen.
- Die Bereitstellung von Techniken zur Pflege von Ontologien, die die Evolution von Ontologien gemäß dem wissenschaftlichen Fortschritt garantieren.
- Ein zunehmender Grad an Formalisierung, der durch Fortschritte im Bereich der Entwicklung von Softwareanwendungen zur Pflege, Validierung und Verknüpfung von Ontologien beschleunigt wird.
- Eine gesteigerte Sorgfalt bei der akkuraten Darstellung biologischer Sachverhalte und damit der Realität insgesamt, die durch die terminologischen Komponenten des repräsentationalen Artefakts, z. B. der Ontologie, abgebildet werden.

Damit ist das Folgende aufs Engste verbunden:

- Die Bereitschaft genauer zwischen Daten und Informationen auf der einen Seite und Objekten in der klinischen Realität, die durch Daten und Informationen abgebildet werden, auf der anderen zu unterscheiden.



[About EJBI](#)

[Editorial Board](#)

[Instructions to Authors](#)

[Browse EJBI](#)

[Special Issues](#)

[Sponsorship & Ads](#)

[Downloads](#)

[Contact](#)

ISSN 1801-5603

... diese beiden Strategien werden als Beispiele ONOMED CT und die OBO Foundry angesehen.

Vor allem existieren biomedizinische Ontologien, um der Integration von klinischen und experimentellen Daten zu dienen, sind die Kriterien für deren Erfolg im Grunde die gleichen, die für ein Telefonnetzwerk gelten: Wichtig ist die Menge der Nutzer. Ein Netzwerk, das nur wenige Kunden hat, ist ein Misserfolg, unabhängig davon wie gut die verwendeten Technologien sind. In diesem Artikel werden die beiden oben genannten Strategien auf ihre Anlagen hin untersucht, die Nutzer für die terminologischen Ressourcen gewinnen und nachhaltig daran binden zu können.

## 1. Anwachsen der formalen Strenge biomedizinischer Ontologien

Eine wichtige Tendenz bei der Entwicklung von Ontologien ist der gesteigerte Wert, der formallogischen Verfeinerungen von biologischen Terminologien und verwandten Artefakten beigemessen wird. Sicherlich gibt es immer noch Zweifel in einigen Kreisen, wenn es um die Frage geht, ob verstärkte formale Strenge ein Gewinn für medizinische Ontologien ist. Manche Parteien behaupten, dass medizinisches Wissen zu intuitiv ist (Medizin als *Kunst*, nicht als *Wissenschaft*) und außerdem in einem zu großen Maße von subjektiven Erfahrungen und örtlichen Traditionen abhängt, um wissenschaftsbasierte, einheitliche Terminologien oder Ontologien schaffen zu können.

Aus der Entwicklung der molekularen und genetischen Medizin und den damit verbundenen Technologien entsteht die Notwendigkeit, Daten aus verschiedenen Disziplinen, die sich auf unterschiedliche Granularitätsebenen beziehen, zu integrieren. Diese Entwicklungen werden sicherlich dazu beitragen, dass solche Argumente langsam aber sicher als überholt erkannt werden.

Die Entwicklung medizinischer Ontologien und Terminologien sieht sich zunehmend mit Fragen über formale Methoden konfrontiert, was z.B. in der Arbeit der Semantic Web Health Care and Life Sciences Interest Group [8] deutlich wird, aber ebenso bei der Entwicklung einer *description logic* Infrastruktur für Vokabularien wie GALEN [9], den National Cancer Institute Thesaurus [10] und – von grundsätzlicher Bedeutung für diese Arbeit – den neusten Versionen des SNOMED Vocabulary [11].

Auf Seiten der Biologie haben die zunehmende Bedeutung der Bioinformatik und die enorme Steigerung der Masse der den Forschern zur Verfügung stehenden Daten zu bemerkenswerten Anstrengungen darin geführt, formal stärkere Ontologieressourcen zu entwickeln, die jetzt in der Arbeit des National Center for Biomedical Ontology [12] zusammengefasst werden.

Der Erfolg der Gene Ontology führte dazu, dass viele Biologen das OBO Format (ehemals das GO Format) [13] als Repräsentations-sprache für die Ontologieentwicklung bevorzugten. Das OBO Format ist hinsichtlich formaler Möglichkeiten ebenfalls wachsender Aufmerksamkeit unterworfen. Das Ziel lautet, Wege für neue Möglichkeiten algorithmischen *Reasonings* zu finden, die sowohl die biomedizinische Forschung als auch die klinische Versorgung unterstützen. Die OBO Foundry stellt 53 Ontologien für die biologischen und biomedizinischen Disziplinen zur Verfügung, von denen 49 im OBO Format vorliegen, zusammen mit einer großen Menge an Daten, die mit ihrer Hilfe annotiert wurden, inklusive über 11 Millionen frei zugänglicher Annotationen, die Produkte von Genen (Proteine und funktionale RNA) mit Ausdrücken in der GO in Verbindung setzen [14]. Außerhalb der Biologie ist die *Web Ontology Language* (OWL) die bevorzugte Repräsentations-sprache der Ontology Community. OWL ist ein W3C Standard, und seine algorithmischen Eigenschaften machen es zu einem attraktiven Werkzeug für die Softwareentwicklung. Allerdings ist der ontologische Gehalt, der im ursprünglichen OWL Format repräsentiert werden kann mitunter sehr unausgereift. Dank der Entwicklung eines OBO-OWL Konverters sind ontologische Inhalte von OBO jetzt ebenfalls in auf OWL basierenden Anwendungen möglich. Die OBO Ontologien und die damit verbundenen Annotationen dienen also jetzt als wichtiges Medium für die Ausweitung des Semantic Web in den Gebieten der Lebenswissenschaften.

## 2. Wachsende Konzentration auf Biologische Realität

Während Informatiker sich früher vornehmlich mit Daten beschäftigten, ist ein zweiter Trend bei der jetzigen Entwicklung von Ontologien die steigende Beschäftigung mit der durch diese Daten abgebildete Wirklichkeit. Ontologien werden nach wie vor häufig von Personen aus dem *Knowledge Management* erstellt. Dadurch wurden Ontologien zunächst hauptsächlich als Möglichkeit zur Abbildung von Information, mitunter unabhängig von einer zugrundeliegenden Wirklichkeit interpretiert. Jetzt aber, wo Ontologien zunehmend von Biologen und Kliniken für ihre eigenen fachspezifischen Bedürfnisse geschaffen werden, tritt die Frage nach der genauen Abbildung der Wirklichkeit immer stärker in den Vordergrund.

Diese neue Fokussierung auf die biologische Realität seitens der Bioontologen wird unter anderem durch die letzten Revisionen des SNOMED CT Vokabulars [11], [18], [19] und durch die Entwicklung der OBO Foundry deutlich. Diese stehen für eine Verschiebung von der früheren Beschäftigung mit Daten und Informationen um ihrer selbst Willen hin zu etwas, was wir das realistische Paradigma nennen.

Letzteres basiert auf einer Trennung zwischen drei Ebenen der Realität:

- Mentale Repräsentationen (Ideen oder Gedanken in unserem Bewusstsein, die bestimmte Teile der Realität abbilden).



[About EJBI](#)

[Editorial Board](#)

[Instructions to Authors](#)

[Browse EJBI](#)

[Special Issues](#)

[Sponsorship & Ads](#)

[Downloads](#)

[Contact](#)

**ISSN 1801-5603**

Für einen Realisten hängt der Erfolg bei der Entwicklung einer Ontologie davon ab, die Trennung dieser drei Ebenen einzuhalten [20]. Diese Trennung der Ebenen impliziert, dass die Realität, für deren Abbildung unsere Repräsentationen entwickelt werden, unabhängig von diesen Repräsentationen existiert. Der konzeptualistische Standpunkt, der unter Informationswissenschaftlern immer noch verbreitet ist, sieht Ontologien als Repräsentationen von Begriffen, die, kurz gesagt, Einheiten von Wissen (oder Bedeutungen) im Bewusstsein von Menschen darstellen. Realisten definieren eine Ontologie im Gegensatz dazu folgendermaßen: Repräsentation der Typen von Entitäten, die in einem bestimmten Feld der Wirklichkeit existieren und deren Beziehungen zueinander.

Typen sind die Invarianten oder Muster in der Welt, die von den einzelnen Wissenschaften durch Experimente (auch klinische Experimente) ausgemacht werden. Die Beziehung zwischen Typ und Einzelding ist die der Instantiierung. Typen werden an unterschiedlichen Orten und zu unterschiedlichen Zeiten in unterschiedlichen Partikularien oder Einzeldingen instantiiert und normalerweise mit allgemeinen Ausdrücken wie „Hund“ oder „Diabetes“ bezeichnet [20]. „Hund“ ist der Name des Typs, der von meinem Hund Fido und von deinem Hund Rover instantiiert wird. Da Ontologien für Realisten repräsentationale Artefakte sind, die entwickelt werden, um wissenschaftliche Untersuchungen oder ähnliche Unternehmungen zu unterstützen, sind diese mehr auf allgemeine Gesetzmäßigkeiten und daher eher auf Typen als auf Instanzen bezogen.

*SNOMED CT*: Die Tendenz hin zu stärkerem Realismus zeigt sich bei SNOMED CT an der aktuellen Deaktivierung von Konzepten, die als „not otherwise specified“ (NOS, nicht weiter spezifiziert) qualifiziert werden, zum Beispiel „Brain injury NOS (disorder)“ (26286008.) Bereits Cimino [21] hat herausgestellt, dass solche Qualifikationen wie „NOS“ Probleme bereiten müssen. Von einem realistischen Standpunkt aus können diese Probleme darauf zurückgeführt werden, dass es keine solche Entität wie eine „nicht weiter spezifizierte Gehirnverletzung“ gibt; und zwar weder als Typ noch als Instanz. Wenn es zum Zeitpunkt  $t$  eine solche Entität geben würde und es würde zu einem späteren Zeitpunkt  $t'$  eine Spezifikation dieser Entität erfolgen, so müssten wir entweder davon ausgehen, dass die ursprüngliche Entität durch den schlichten Akt der Aufnahme dieser neuen Information zerstört würde, oder dass diese Entität mit der geplanten neuen Nachfolgeentität ungleich ist.

Wie bereits in [22] ausgeführt, referieren „NOS“ und ähnliche Terme nicht auf eine bewußtseins-unabhängige *Realität*, vielmehr handelt es sich um konfus formulierte Repräsentationen unseres *Wissens* über die Realität. Selbstverständlich ist es für biomedizinische Informationssysteme unerlässlich die fortschreitenden Veränderungen, denen unser Wissensstand, etwa bei einer medizinischen Untersuchung, unterworfen ist, aufzuzeichnen, zum Beispiel im Hinblick auf die Zuverlässigkeit der Information. Es müssen also beide Arten von Information gesammelt werden: Informationen über die Realität und Informationen über unser Wissen über die Realität. Wenn aber Kodierungsschemata algorithmisches *Reasoning* in dem Umfang, in dem es in der Biomedizin in Zukunft erforderlich sein wird, unterstützen sollen, dann ist es ebenso wichtig zwischen diesen beiden Informationsarten deutlich zu unterscheiden.

Ihre Entwickler haben inzwischen einige Schritte in die richtige Richtung unternommen, aber SNOMED CT enthält immer noch Konzepte wie „unknown living organism“ (SNOMED: 89088004) und „presumed viral agent“ (SNOMED: 106551006). Diese werden in SNOMED behandelt, als seien es spezifische Arten von Organismen bzw. Viren, tatsächlich handelt es sich aber in solchen Fällen um Repräsentationen eines Wissensstandes.

Zudem enthält SNOMED CT einige Terme, wie zum Beispiel „Abszess“, die zweifach eingetragen sind, zum einen als Befund, zum anderen als morphologische Anormalität:

128477000 Abscess (disorder),

44132006 Abscess (morphologic abnormality).

In solchen Fällen muss es zu Problemen kommen, wenn SNOMED CT als Kodierungsschema genutzt wird, denn es gibt hinsichtlich dieser beiden Terme keinen eindeutigen Unterschied in ihrer Referenz [23], [24].

*OBO Foundry*: Die Gene Ontology und ihre Schwesterontologien im OBO Repository wiesen eine Reihe antirealistischer Komponenten auf, die zu strukturalen Problemen in der GO führten, wie in [25], [26] dargestellt.

Die Gründung der OBO Foundry führte zu einer Reihe tiefgreifender Reformen dieser Ontologien. Zum Hauptziel der Initiative wurde eine akkurate Darstellung des biologischen Gegenstandsbereiches, und, damit verbunden, wurde ein realistischer Standpunkt übernommen.

Jede Ontologie, die Mitglied der OBO Foundry ist, enthält strukturierte Repräsentationen der Typen, die in einer spezifizierten Domäne der Realität existieren. Ziel ist es, die Typen gemäß der besten wissenschaftlichen Theorie zum gegenwärtigen Zeitpunkt abzubilden. In diesem Sinne ist jede Ontologie selbst analog zu einer wissenschaftlichen Theorie: sie hat einen vereinheitlichten Gegenstandsbereich, der aus Entitäten besteht, die unabhängig von der Ontologie existieren. Weiterhin wird durch die Ontologie an einer Optimierung der deskriptiven oder repräsentationalen Adäquatheit hinsichtlich dieses



About EJBI

Editorial Board

Instructions to Authors

Browse EJBI

Special Issues

Sponsorship & Ads

Downloads

Contact

ISSN 1801-5603

ist der Verzicht auf Aussagen wie:

1. biological process *is\_a* Gene Ontology,

die bislang gemeinsam mit biologisch zutreffenden Aussagen wie

2. cell development *is\_a* biological process

bestanden.

Aussagen wie 1. zeigen die Verwendung der *is\_a* Relation in Zusammenhängen, in denen sie ungefähr „ist eine terminologische Komponente von“ bedeuten soll. Dies ist jedoch inkonsistent mit der Hauptverwendung dieser Relation in der GO, die die Bedeutung „hat Untertyp (-klasse)“ hat. Aussage 2. kann in dem Moment zu Problemen führen, in dem die GO mit anderen Ontologien zusammengeführt wird, zum Beispiel mit solchen, die Informationsobjekte betreffen. Bei der Gültigkeit z.B. von:

3. Gene Ontology *has\_file\_format* RDF-XML

kann, bedingt durch die Transitivität von *is\_a*, aus 2. und 3. geschlossen werden, dass

4. cell development *has\_file\_format* RDF-XML.

Dies ergibt natürlich biologisch betrachtet keinen Sinn.

Um solche Probleme in der GO und ihren Schwesterontologien zu eliminieren, wird jetzt das Prinzip angewandt, dass alle Aussagen in den Ontologien Relationen verwenden, die eindeutig definiert sind, wie dies in den Definitionen der Relationen in der OBO Relation Ontology musterhaft zu sehen ist [27], [28]. Die Anwendung dieses Prinzips stellt sicher, dass die Bedeutung der Begriffe in Relationen in allen Ontologien die gleiche ist, was die logisch schlüssige Integration solcher Ontologien für *Reasoning*-Anwendungen erleichtert.

Ein weiteres Anzeichen für die höhere Gewichtung, die dem Realismus zukommt, ist die Aufnahme des Foundational Model of Anatomy (FMA) in die OBO Foundry. Ziel des FMA war es vom Anfang seiner Entwicklung an, von einem realistischen Standpunkt aus die Repräsentationen von Typen und Relationen vorzunehmen, die notwendig sind, um die Strukturen des Säugetierkörpers, im Besonderen des menschlichen Körpers, computerisierbar zu hinterlegen, und zwar in einer Form, die auch für Menschen verständlich ist [29], [30].

Die Formulierung von Definitionen im FMA, die auch in der gesamten OBO Foundry angenommen wurde, folgt Prinzipien, die erstmalig von Aristoteles niedergelegt wurden. Die Definition einer repräsentationalen Einheit „A“ in einer Ontologie, die auf einen Typ A referiert, muss durch die Referenz auf die Oberklasse von A, nämlich B, in der gleichen Ontologie und eines unterscheidenden Kriteriums C, dass von allen und nur von A erfüllt wird, formuliert werden. Aus diesem Prinzip resultieren Definitionen der Form:

ein A = ein B, das C erfüllt.

Diese Methode ist in der Biologie jedem aus dem Bereich der Taxonomie bekannt, da sich hier nach aristotelischem Muster der Speziesname aus dem Genusnamen einer Differentia zusammensetzt, z.B. *Homo sapiens* als Bezeichnung der Spezies von Mitgliedern der Gattung *Homo*, die vernunftfähig sind.

Beispiele hierfür aus dem FMA sind:

a cell = def. an anatomical structure which consists of cytoplasm surrounded by a plasma membrane  
 a plasma membrane = def. a cardinal cell part that surrounds the cytoplasm

Durch das Durchhalten dieser Strategie wird nicht nur die Konsistenz der Verwendung von Begriffen unterstützt, sondern sie trägt auch zur Schlüssigkeit der *is\_a* Hierarchie bei.

### 3. Modularität biomedizinischer Ontologien

SNOMED CT und die OBO Foundry verfolgen zwei unterschiedliche Strategien gerade auch im Hinblick auf Modularisierung. Diese Strategien unterscheiden sich sowohl in ihren theoretischen, wie auch in ihren praktischen Konsequenzen.

*SNOMED CT*: Die ambitionierteste Initiative im Bereich der Vereinheitlichung biomedizinischer Begriffssysteme wurde gerade von der International Health Terminology Standards Development Organization (IHTSDO) übernommen [31]. Diese versucht SNOMED CT als internationale Hauptterminologie für den klinisch-medizinischen Bereich zu etablieren, die die Gesamtheit dieser Domäne in einer Vielzahl von Sprachen abdeckt. Die Basis für dieses Vorhaben ist die englische Version von SNOMED CT, die mehr als 315 000 Konzepte beinhaltet und bereits teilweise in andere Sprachen übertragen worden ist. Die Hauptstränge von SNOMED CT sind in *Description Logics* darstellbar [32], wenn auch nicht in einer der W3C Standard *description logics*. Weiterhin wird an *Mappings* für andere existierende Standards, zum Beispiel die International Classification Of Disease (ICD) 10 der WHO,



[About EJBI](#)

[Editorial Board](#)

[Instructions to Authors](#)

[Browse EJBI](#)

[Special Issues](#)

[Sponsorship & Ads](#)

[Downloads](#)

[Contact](#)

ISSN 1801-5603

Das Ziel all dieser Bemühungen ist es für den gesamten Bereich der klinischen Medizin eine einzige terminologische Ressource bereitzustellen. Die Verwendung einer solchen „Master“-Terminologie würde zahlreiche positive Effekte zeitigen, darunter weniger redundante Daten, bessere Durchführbarkeit von longitudinalen Studien und Meta-Analysen, sowie die Sicherstellung der Konsistenz von Patientendaten über die gesamte Lebensspanne hinweg. Die Existenz einer einzigen Ressource ermöglicht die Benutzung gemeinsamer Tools und die Durchführung gemeinsamen Trainings. Trotz all dieser Vorteile, die vor allem der klinischen Entscheidungsfindung und der Datensammlung im Rahmen von Public Health Programmen zu Gute kommen würden, ist es schwer die notwendigen finanziellen Mittel zu mobilisieren, zumal SNOMED CT gut dokumentierte Schwachstellen enthält [33].

Die Tatsache, dass immer noch wenige Institutionen des Gesundheitswesens SNOMED CT zum Kodieren klinischer Daten verwenden, deutet darauf hin, dass die Initiative noch weiter investieren muss, um die Vorteile auch tatsächlich einzufahren [34]. Es steht weiterhin die Frage im Raum, inwieweit die hohen Kosten, die die Implementierung und Nutzung einer hochentwickelten Terminologie wie SNOMED CT im Klinikbereich verursacht, zu amortisieren sind. SNOMED CT zum Beispiel ist nicht wie viele andere Systeme entwickelt worden, um die Rechnungsstellung zu simplifizieren. Letzteres ist jedoch ein zentraler Aspekt des Krankenhausmanagements, der aber von wesentlich weniger komplexen Terminologien abgedeckt werden kann und auch wird, die zudem schon sehr weit verbreitet sind.

Wie bereits erwähnt, ist auch die Existenz interner Fehler einer der Gründe für die langsame Ausbreitung von SNOMED CT. Es gibt Lücken in der Terminologie, Fehler in der Strukturierung und Probleme mit der Konsistenz, die aus der Anwendung unterschiedlicher Prinzipien bei der Entwicklung unterschiedlicher Zweige der Terminologie herrühren. In letzter Zeit wurden Beweise dafür, dass diese Schwächen effizientes Kodieren verhindern, publiziert [23], [24]. Diese Qualitätsprobleme werden inzwischen von den Kuratoren angegangen.

Ein Einwand ganz anderer Art zielt auf die Frage, ob die Strategie *einer* kompletten Abdeckung des Gegenstandsbereiches in einer Ressource richtig ist oder nicht.

OBO *Foundry*: Im Gegensatz zu SNOMED CT verfolgt OBO einen modularen Ansatz. Die Strategie, die der OBO Foundry [5] zugrunde liegt, ist eine Reihe von interoperablen Ontologien bereitzustellen, die unterschiedliche, nicht überlappende Domänen der biomedizinischen Realität auf der Basis von gemeinsamen Prinzipien der Ontologieentwicklung abbilden. Nur Ontologien, deren Entwickler diese Prinzipien anwenden, können in die OBO Foundry aufgenommen werden. Es gibt zwei Hauptziele der Prinzipien: zum einen sollen sie die Harmonisierung existierender Domänenontologien vorantreiben, zum anderen sollen sie als Leitfaden zur Entwicklung neuer Ontologien und als Beispiel guter wissenschaftlicher Praxis für Neulinge in der Ontologieentwicklung dienen. Immer mehr fallen im biomedizinischen Bereich Teilbereiche auf, für die völlig neue Ontologien entwickelt werden müssen. Leitlinien sind gerade deshalb besonders wichtig, um die Interoperabilität mit Ontologien benachbarter Bereiche sicherzustellen.

Das Ziel ist es, eine Situation herbeizuführen, in der die Abdeckung von Wissensbereichen durch terminologische Ressourcen schrittweise über wachsende Domänen der biomedizinischen Realität vergrößert wird. Diese Strategie ist selbstverständlich das exakte Gegenteil der Strategie, die von SNOMED CT verfolgt wird.

Aus Sicht der Foundry ist das Ziel von Ontologieentwicklung nicht, eine einzige große Ontologie bereitzustellen, die alles abdeckt, was überhaupt existiert. Vielmehr gilt es bessere Wege zu finden kleine Domänenontologien zunächst getrennt zu entwickeln, um so die Vorteile der Arbeitsteilung und die Bildung von domänenspezifischen Kompetenzteams zu fördern. Im Hinblick auf die Harmonisierung, die einen unabdingbaren zweiten Schritt darstellt, muss ein gewisses Maß an zentraler Organisation stattfinden, um die Einhaltung von allgemein anerkannten Leitlinien sicherzustellen. Die Tatsache, dass Versuche unternommen werden, Harmonisierungen *von oben* durchzusetzen, zeigt, dass Harmonisierungsbestrebungen evolutionär aufgesetzt werden müssen und auch die Leitlinien von Zeit zu Zeit einer Revision unterworfen werden sollten. Die Dokumentation der Leitlinien und ihrer Entwicklung ist jedoch eine wichtige Voraussetzung dafür, auf die Bedürfnisse unterschiedlicher Interessengruppen einzugehen zu können und somit ihre Unterstützung zu gewinnen.

Die Gruppen, die die OBO Foundry vorantreiben, stimmen darüber ein, dass eine minimale gemeinsame *Upper Ontology* ausreichend ist, um Harmonisierung zu ermöglichen. Eine solche Ontologie kann dazu beitragen von Beginn an Grundlagen einer *best practice* der Ontologieentwicklung zu realisieren. Diese Vorteile kann eine Upper Ontology allerdings nur bringen, wenn sie selbst nach gut fundierten Prinzipien entwickelt und hinreichend getestet wurde. Ist dies der Fall, dann tritt als weiterer positiver Effekt bei der Benutzung einer Upper Ontology erhebliche Zeit- und Ressourcenersparnis hinzu, denn es muss nicht mehr in die Repräsentation der Strukturen der Metaebene investiert werden, die für Fachwissenschaftler unerlässlich sind um ihre Ontologie zu organisieren. Zudem gehen Upper Ontologies in der Regel auf Prinzipien zurück, die außerhalb der Erfahrung von Fachwissenschaftlern liegt.

Die Basic Formal Ontology (BFO), die als Kern des Upper Levels der OBO Foundry Ontologien dient, basiert (ebenso wie SUO und DOLCE) auf der grundlegenden Unterscheidung von *continuants* und *occurents*. Erstere sind Entitäten, die über die Zeit hinweg existieren und Bestand haben. Sie bleiben mit sich selbst identisch, auch wenn sie sich auf unterschiedliche Weisen verändern. Letztere ereignen sich, was



[About EJBI](#)

[Editorial Board](#)

[Instructions to Authors](#)

[Browse EJBI](#)

[Special Issues](#)

[Sponsorship & Ads](#)

[Downloads](#)

[Contact](#)

ISSN 1801-5603

Upper Ontologien, sondern die gesamte Domäne abdecken, denn die abstrakte Domäne verliert sie kontinuierlich Teile, oder erhalten sie dazu im Rahmen von Entwicklung und Wachstum [35], [36].

Die Benutzung einer Upper Ontology fördert die Harmonisierung von Ontologien, indem sie eine einheitliche und kohärente Methodik zur Wirklichkeitsrepräsentation auf der höchsten Organisationsebene vorgibt. Selbstverständlich finden sich die Terme, die für die Praxis von vorrangiger Bedeutung sind, auf den unteren Ebenen. Kriterien, wie sie von einer Upper Ontology nahe gelegt werden, vereinfachen jedoch die Organisation dieser unteren Level immens [37], wobei zu berücksichtigen gilt, dass die Benutzung einer Upper Ontology die Nützlichkeit der Informationsressource insgesamt für den Endbenutzer nicht einschränken darf.

Neben diesen qualitätsbezogenen Aspekten widmet sich die OBO Foundry auch Fragen der Motivation. Ontologieentwicklung ist ein kostspieliges Unternehmen, zumal, um ein optimales Ergebnis zu erzielen, auch die Domänenexperten viel Zeit aufwenden müssen, um die Arbeit der Ontologen zu unterstützen. Die bisherigen Erfahrungen zeigen, dass Experten bereit sind, sich im Rahmen der Foundry zu engagieren, da sie, gerade vor dem Hintergrund der Arbeitsteiligkeit und Modularität, direkten Einfluss auf die Ressourcen haben, die sie in der Zukunft verwenden werden [38]. Insofern ähnelt die Strategie der OBO Foundry der in der *open source* Softwareentwicklung [39].

## Ergebnisse

Sowohl IHTSDO als auch die OBO Foundry zielen darauf, eine vollständige Repräsentation der klinischen Realität zu schaffen. IHTSDO hat den Vorteil einer großen, einheitlichen Organisation, die erhebliche finanzielle Ressourcen mobilisieren kann und einem systematischen Plan folgt. Die OBO Foundry, für die die klinische Realität nur eine Untermenge der biomedizinischen Realität ist, hat den Vorteil einer Strategie, der Mobilisierung von Fachexperten und einer großen, etablierten Benutzergruppe, letztere allerdings vor allem im Bereich der Biologie und noch nicht so sehr unter den Klinikern. Es ist noch zu früh endgültig zu beurteilen, welche der beiden Strategien die anstehenden dringlichen Probleme lösen wird. Aber diese Probleme – Benutzerfreundlichkeit, Reliabilität, Kosteneffektivität und größtmögliche Abdeckung der Domäne – müssen gelöst werden, wenn die effektive Integration biomedizinischer Daten gesichert werden soll.

## Danksagung

Barry Smiths Beitrag zu dieser Arbeit wurde teilweise durch das National Center for Biomedical Ontology, dass durch das National Institut of Health finanziert wird, im Rahmen der NIH Roadmap for Medical Research, Grant 1 U 54 HG004028 gefördert. Informationen über die National Centers for Biomedical Computing können unter <http://nihroadmap.nih.gov/biomedicine> abgerufen werden.

## Literatur

- [1] <http://www.ncbi.nlm.nih.gov/sites/entrez?db=mesh>
- [2] <http://www.hl7.org>
- [3] <http://www.geneontology.org>
- [4] <http://obofoundry.org>
- [5] Smith B., Ashburner M., Rosse C., et al.: The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*; 25 (11), 2007, pp. 1251-1255.
- [6] Ruttenberg A., Clark T., Bug W., et al.: Advancing translational research with the Semantic Web. *BMC Bioinformatics*, 8, 2007.
- [7] Smith B., Brochhausen M.: Establishing and Harmonizing Ontologies in an Interdisciplinary Health Care and Clinical Research Environment. In B. Blobel, P. Pharow, M. Nerlich, eds. "eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge, IOS Press, Amsterdam, 2008, pp: 219-234.
- [8] <http://www.w3.org/2001/sw/hcls>
- [9] <http://www.opengalen.org>
- [10] <http://nciterns.nci.nih.gov/NCIBrowser/Dictionary.do>
- [11] <http://www.ihtsdo.org/snomed-ct>
- [12] <http://bioontology.org>
- [13] <http://www.geneontology.org/GO.format.shtml>
- [14] <http://www.ebi.ac.uk/GOA>
- [15] <http://www.w3.org/2004/OWL>
- [16] <http://www.berkeleybop.org/ontologie>
- [17] Moreira D. A., Musen M. A.: OBO to OWL: a Protégé tab to read/save OBO ontologies. *Bioinformatics*, 2007 Jul; 23(14), pp. 1826-70.
- [18] Ceusters W., Spackman K.A., Smith B.: Would SNOMED CT benefit from Realism-Based Ontology Evolution? In American Medical Informatics Association 2007 Annual Symposium Proceedings,



[About EJBI](#)

[Editorial Board](#)

[Instructions to Authors](#)

[Browse EJBI](#)

[Special Issues](#)

[Sponsorship & Ads](#)

[Downloads](#)

[Contact](#)

**ISSN 1801-5603**

- Ontologists and Logicians' Health Check. *Int J Med Inform*, 2006, Sep. 10 [Epub ahead of print].
- [20] Smith B., Kusnierczyk W., Schober D., Ceusters W.: Towards a Reference Terminology for Ontological Research and Development in the Biomedical Domain. KR-MED 2006.
- [21] Cimino J.J.: Desiderata for controlled medical vocabularies in the Twenty-First Century. *Methods Inf Med*; 37(4-5), 1998, pp.394-403.
- [22] Bodenreider O., Smith B., Burgun A.: The ontology-epistemology divide: A case study in medical terminology. *FOIS (Formal Ontology and Information Systems)*, 2004, pp. 185-95.
- [23] Andrews J.E., Richesson R.L., Krischer J.: Variation of SNOMED CT coding of clinical research concepts among coding experts. *J Am Med Inform Assoc* 2007, 14, 4, p. 497-506.
- [24] Chiang M.F., Hwang J.C., Yu A.C., Casper D.S., Cimino J.J., Starren J.: Reliability of SNOMED-CT coding by three physicians using two terminology browsers. *AMIA 2006 Symposium Proceedings*, 2006, p. 131-135.
- [25] Smith B., Williams J., Schulze-Kremer S.: The Ontology of the Gene Ontology. In *Biomedical and Health Informatics: From Foundations to Applications, Proceedings of the Annual Symposium of the American Medical Informatics Association*, Washington DC, November 2003, pp. 609-613.
- [26] Smith B., Köhler J., Kumar A.: On the Application of Formal Principles to Life Science Data: A Case Study in the Gene Ontology. In *Proceedings of DILS 2004 (Data Integration in the Life Sciences)*, (Lecture Notes in Bioinformatics 2994), Berlin: Springer, 2004, pp. 79-94.
- [27] <http://www.obofoundry.org/ro>
- [28] Smith B., Ceusters W., Klagges B., et al.: Relations in Biomedical Ontologies. *Genome Biology*, 6:R46, 2005.
- [29] <http://sig.biostr.washington.edu/projects/fm>
- [30] Rosse C., Mejino J.V.L.: A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*. 2003, 36, pp. 478-500.
- [31] <http://www.ihtsdo.org>
- [32] Donnelly K.: Multilingual documentation and classification. In B. Blobel, P. Pharow, M. Nerlich, eds. "eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge", IOS Press, Amsterdam, 2008, pp: 235-44.
- [33] <http://www.hiww.org/smcs2006/talks/Rector.ppt>
- [34] <http://www.chcf.org/documents/chronicdisease/SantaBarbaraLessonsLearned.pdf>
- [35] <http://www.ifomis.org/bfo>
- [36] Grenon P., Smith B., Goldberg L.: Biodynamic Ontology: Applying BFO in the Biomedical Domain. In *Ontologies in Medicine*, D.M. Pisanelli, Ed. Amsterdam, IOS Press, 2004, pp. 20-38.
- [37] Smith B.: From Concepts to Clinical Reality: An Essay on the Benchmarking of Biomedical Terminologies. *Journal of Biomedical Informatics* 39(3), 2006, pp. 288-298.
- [38] Smith B.: Ontology (Science). In C. Eschenbach and M. Gruninger (eds.), *Formal Ontology in Information Systems. Proceedings of FOIS 2008 Amsterdam*: IOS Press, 21-35.
- [39] Weber S.: *The Success of Open Source*. Cambridge, MA: Harvard University Press, 2004.