

Philosophie und biomedizinische Forschung¹

Barry Smith

University at Buffalo/Universität des Saarlandes

Bert R. E. Klagges

Universität Leipzig

Erschienen in: *Allgemeine Zeitschrift für Philosophie*, 30: 1 (2005), 5–26.

Zusammenfassung

Die bahnbrechenden wissenschaftlichen Ergebnisse der letzten Jahre erzwingen eine neue philosophische Auseinandersetzung mit den Grundkategorien der Biologie und der benachbarten Disziplinen. Insbesondere die Anwendung neuer informationstechnischer Mittel in der biomedizinischen Forschung und die damit verbundene, kontinuierlich zunehmende Datenflut sowie die Notwendigkeit, ihrer Herr zu werden, erfordern ein konsequentes Nachdenken darüber, wie biologische Daten systematisiert und klassifiziert werden können. Dafür wiederum bedarf es robuster Theorien von Grundbegriffen wie Art, Spezies, Teil, Ganzes, Funktion, Prozess, Fragment, Sequenz, Expression, Grenze, Locus, Umwelt, System usw. Solche Begriffe gehören zum impliziten Wissen jedes Biologen. Sie spiegeln einerseits eine Dimension der biologischen Wirklichkeit wider, die auch vor dem Hintergrund der biologischen Evolution unverändert bleibt. Andererseits verlangt deren theoretische Behandlung nach zeitgemäßen Analoga der in der traditionellen aristotelischen Metaphysik entwickelten Methoden. Zugleich können so die explizit formulierten Theorien und Definitionen bereitgestellt werden, die für computergestützte Informationssysteme unabdingbar sind. Das Entwickeln derartiger Theorien und Definitionen ist eine Aufgabe der Philosophie, die in diesem Sinne herausgefordert ist, zwischen Biologie und Informatik zu vermitteln.

Summary

The pathbreaking scientific advances of recent years call for a new philosophical consideration of the fundamental categories of biology and its neighboring disciplines. Above all, the new information technologies used in biomedical research, and the necessity to master the continuously growing flood of data that is associated therewith, demand a profound and systematic reflection on the systematization and classification of biological data. This, however, demands robust theories of basic concepts such as kind, species, part, whole, function, process, fragment, sequence, expression, boundary, locus, environment, system, and so on. Concepts which belong to the implicit stock of knowledge of every biologist. They amount to a dimension of biological reality which remains constant in the course of biological evolution and whose theoretical treatment requires contemporary analogues of the tools developed in traditional Aristotelian metaphysics. To provide the necessary theories and definitions is a task for philosophy, which is thus called upon to play an important role as intermediary between biology and informatics.

¹ Diese Arbeit wurde durch die Alexander-von-Humboldt-Stiftung im Rahmen des Wolfgang-Paul-Preises unterstützt und stellt die Grundideen eines Forschungsprojekts dar, das im Rahmen des Förderprogramms „Schlüsselbegriffe der Geisteswissenschaften“ von der Volkswagenstiftung gefördert wird. Unser Dank gilt Fabian Neuhaus, Steffen Schulze-Kremer, Cornelius Rosse und Dirk Siebert für wertvolle Anregungen und Hinweise.

(1) Die neue angewandte Ontologie

In den letzten Jahren hat sich ein neuer Wissenschaftszweig entwickelt: die *angewandte Philosophie*, die sich vor allem im Bereich der Medizin- und Bioethik profiliert. Auch im deutschen Sprachraum gibt es inzwischen auf diesem Gebiet eine Reihe wichtiger Beiträge prominenter Philosophen. .

In der angewandten Philosophie wird eine neue Ebene der Interaktion mit den nicht-philosophischen Nachbardisziplinen erreicht. Beiträge wie Ralf Stoeckers *Der Hirntod: ein medizinethisches Problem und eine moralphilosophische Transformation*² sowie Nikolaus Knoepfflers *Forschung an menschlichen Embryonen*³ machen deutlich, dass die ernsthafte philosophische Behandlung bio- und medizinethischer Probleme eine echte Auseinandersetzung mit den relevanten biomedizinischen Fakten voraussetzt. Das bloße Aufstellen philosophischer Theorien und Argumente bietet keine ausreichende Grundlage für künftige, signifikante Arbeit auf diesem Gebiet.

In der Medizin- und Bioethik tätige Philosophen müssen sich nicht nur mit den Inhalten der (Bio-)Wissenschaften vertraut machen; sie müssen auch Wege finden, diese Inhalte in ihre philosophischen Theorien zu integrieren. Vor diesem Hintergrund erscheint es nur logisch, dass sich in jüngster Vergangenheit mit der *angewandten Ontologie* ein Zweig der Philosophie zur angewandten Ethik gesellt hat, der philosophische Ideen und Methoden aus dem Bereich der Ontologie anwendet, um Ergebnisse der wissenschaftlichen Forschung ontologisch aufzubereiten.

Selbst die Computer- und Informationswissenschaften haben in bewusster Anlehnung an die philosophische Verwendung des Wortes den Terminus „Ontologie“ übernommen.⁴ Bereits in den siebziger Jahren wurde im Zusammenhang mit den Bemühungen um die Datenintegration erkannt, dass die vielen verschiedenen, im Laufe der Zeit entwickelten Informationssysteme jeweils auf eigenen Richtlinien für Terminologie und Kategorisierung beruhten. Innerhalb der Informatik konnte sich so das „ontological engineering“ etablieren, das versprach den Weg zu einer gemeinsamen Kommunikationsbasis – einer Art *Esperanto für Datenbanken* – zu bahnen. Das Ziel bestand darin, robuste Kategoriensysteme für die Datenerfassung zu entwickeln, um für die Kompatibilität und Wiederverwendbarkeit elektronisch gespeicherter Information zu sorgen.

Das *Metaphysics Lab* der Stanford University, die *Laboratorien für angewandte Ontologie* in Trient und Rom sowie das Turiner *Centro Interuniversitario di Ontologia Teorica e Applicata* nutzen ontologische Ideen und Methoden in der Interaktion zwischen Philosophie und verschiedenen Bereichen der Informationswissenschaften. Die Ergebnisse dieser Forschung werden von Technologie-Firmen wie Ingenuity Systems (Mountain View, USA), Cycorp, Inc. (Austin, USA), Language and Computing (Zonnegem, Belgien, und Philadelphia, USA) und Ontology Works (Baltimore, USA) in Software-Applikationen umgesetzt.

Die auch die biomedizinische Forschung immer weiter durchdringende Informationstechnologie verlangt danach, die Biomedizin gleichermaßen ontologisch zu erschließen. So greift die Bioinformatik neuerdings auf die angewandte Ontologie zurück, um die

² Freiburg 1999.

³ Stuttgart 1999.

⁴ Barry Smith, „Ontology“, in: Luciano Floridi (Hg.), *Blackwell Guide to the Philosophy of Computing and Information*. Oxford 2003, 155-166.

theoretischen Grundlagen für die Navigation zwischen verschiedenen heterogenen und autonomen biologischen und medizinischen Informationsquellen zu schaffen.⁵

Gegenwärtig ist das im Jahr 2002 an der Universität Leipzig gegründete und jetzt in Saarbrücken angesiedelte *Institute for Formal Ontology and Medical Information Science* die einzige Forschungseinrichtung zur angewandten Ontologie im deutschen Sprachraum. Zugleich ist das IFOMIS die einzige Forschungsstätte, die sich speziell der Anwendung der Ontologie auf den Bereich der Biomedizin gewidmet hat.⁶ Seine Forscher kooperieren eng mit anderen Vertretern der angewandten Ontologie und mit Bioinformatikern. Ihre Arbeit stellt eine einmalige Kombination aus der Orientierung an klassischen philosophischen Fragestellungen und dem Fokussieren auf die Problematik der aktuellen wissenschaftlichen Forschung in der Biomedizin allgemein und der Bioinformatik im Besonderen dar.

(2) Der geschichtliche Hintergrund der angewandten Ontologie

Die Wurzeln der angewandten Ontologie reichen zurück bis zu Aristoteles. Sie greift seine Grundidee wieder auf, dass es möglich sei, bestimmte Strukturen der von den Wissenschaften untersuchten Wirklichkeit philosophisch zu erfassen.

Wie kann aber diese alte Idee in der heutigen Welt mit neuem Leben erfüllt werden? Um diese Frage zu beantworten, müssen wir kurz einen Blick zurück in die Geschichte der Philosophie werfen, bevor wir zur aktuellen Problematik der Bioontologie zurückkehren.

Eine Ontologie kann der Einfachheit halber als eine Art Katalog der Objekte, Attribute, Prozesse und Relationen in einem vorgegebenen Gebiet angesehen werden. (Diese Auffassung entspricht ziemlich genau der Verwendung des Terminus unter Informatikern.) Eine Ontologie unterteilt die Welt in *Klassen* oder *Arten* (in klassischer philosophischer Terminologie: *Universalien*); in komplexeren Anwendungsgebieten sind mehrere Ebenen hierarchisch angeordneter Klassen notwendig. Die von Linné aufgestellten Taxonomien der Organismen sind Beispiele von Ontologien in diesem Sinn; auch in der Medizin wandte Linné die aristotelische Methodik hierarchischer Kategorien in seiner Klassifikation von Krankheiten an.

Aristoteles selbst war der Auffassung, dass die gesamte Wirklichkeit mit einem einzigen System von Kategorien zu erfassen wäre. Unter dem Einfluss Descartes und Kants wurde die so lange dominante Metaphysik aristotelischer Prägung in ihrer Rolle als zentrale Disziplin der Philosophie durch die Epistemologie ersetzt. Die darwinistische Revolution verdrängte die von Aristoteles inspirierte Auffassung von „Kategorien“ oder „Spezies“ als Teil einer für immer festgelegten Ordnung auch aus der Biologie; diese Entwicklung wurde in der ersten Hälfte des zwanzigsten Jahrhunderts unter dem Einfluss der logischen Positivisten vorangetrieben, die die Metaphysikkritik Kants weiterführten.

Gegen Ende des zwanzigsten Jahrhunderts änderte sich die Lage allerdings sowohl in der Philosophie als auch in der Biologie erneut. In der Philosophie ist es Denkern wie Saul Kripke, Hilary Putnam, David Armstrong, Roderick Chisholm, David Lewis und Ruth Milikan gelungen, der ontologischen bzw. metaphysischen Art des Philosophierens als „analytischer Metaphysik“ wieder die ihr gebührende Anerkennung zu verschaffen – in einer Entwicklung, die auch bestimmte Elemente der aristotelischen Kategorienlehre (als Lehre von den „Universalien“ oder „natural kinds“) neu aufleben ließ. Gleichzeitig erhielt mit der gewachsenen Bedeutung der neuen Bioethik gerade auch die Philosophie der biologischen

⁵ Domenico M. Pisanelli (Hg.), *Ontologies in Medicine: Proceedings of the Workshop on Medical Ontologies, Rome, October 2003*, Amsterdam 2004.

⁶ Barry Smith, Werner Ceusters und Dirk Siebert, „Was die philosophische Ontologie zur biomedizinischen Informatik beitragen kann“, in: *Information: Wissenschaft und Praxis*, 55 (2004), 143–146.

Wissenschaften eine neuartige, ontologische Prägung. In der Biologie selbst gewannen viele der zwischenzeitlich für obsolet erachteten, klassischen Ideen zu Kategorienhierarchien wieder an Aktualität; beispielsweise die bereits durch Aristoteles in den *Categoriae* und *De Interpretatione* skizzierte Methodik und Definitionstechnik. Diese Ideen wurden zwar auch durch die Darwinsche Neuinterpretation, die die Kategorien unter Berücksichtigung der Evolution als zeitlich befristet begriff, nie komplett aus der Biologie verdrängt. Die wachsende Bedeutung von Taxonomie- und Terminologiesystemen vor dem Hintergrund der heutigen Forschung in der Genomik und in der heutigen Bioinformatik ließen diese Ideen jedoch neu erblühen.

(3) Der ontologische Perspektivismus

Die aristotelische Auffassung, man könne die gesamte Wirklichkeit mit einem einzigen System von Kategorien erfassen, erwies sich im Laufe der dargestellten Entwicklungen als überholt. Statt dessen setzte sich die Erkenntnis durch, dass mehrere verschiedene, jeweils partielle Kategoriensysteme benötigt werden, um die durch die Forschung der verschiedenen Wissenschaften erschlossene Wirklichkeit erfassen zu können.

Durch ein Kategoriensystem wird ein Gegenstandsbereich in Kategorien, Klassen, Typen, Gruppierungen oder Arten aufgeteilt. Das System gibt Aufschluss darüber, wie sich die entsprechenden Kategorien von Objekten zueinander verhalten. Beispielsweise liefern die Taxonomien Linnés Aufteilungen der Welt der Organismen mit ihren verschiedenen Ober- und Unterkategorien (Reiche, Stämme, Klassen, Ordnungen, Familien, Gattungen usw.), die nunmehr im *International Code of Zoological Nomenclature*, im *International Code of Botanical Nomenclature* und im *International Code of Nomenclature of Bacteria* definiert sind.

Nun kann ein und derselbe Ausschnitt der Wirklichkeit oft durch verschiedene, nicht immer überschneidungsfreie Aufteilungen erfasst werden: Das Periodensystem der Elemente ist eine Aufteilung (nahezu) der gesamten materiellen Wirklichkeit in chemische Grundbestandteile. Auch der astronomische Kategorienbaum, eine Taxonomie der im All vorkommenden Gebilde (Sonnensysteme, Planeten, Monde, Asteroiden usw.), ist eine Aufteilung (nahezu) der gesamten materiellen Wirklichkeit – nur aus einer anderen Perspektive und in einer anderen Granularität.

Die Auffassung, dass mehrere solcher, sich überschneidende Aufteilungen der Wirklichkeit gleichermaßen verwendet werden können und müssen, werden wir im Folgenden „ontologischer Perspektivismus“ nennen. Im Gegensatz zu manchen perspektivistischen Positionen in der Geschichte der Philosophie (etwa bei Nietzsche oder Foucault) ist diese ontologische Variante des Perspektivismus mit dem wissenschaftlichen Weltbild durchaus vereinbar. Der ontologische Perspektivist akzeptiert, dass es alternative Sichten auf die eine Wirklichkeit gibt und dass dieselbe Wirklichkeit folglich unterschiedlich präsentiert werden kann. Dasselbe Stück Welt kann entweder durch ein Fernrohr, mit bloßem Auge oder durch ein Mikroskop betrachtet werden. Analog lassen sich die Gegenstände der wissenschaftlichen Forschung statt mit einem optischen Instrument mit einer Taxonomie, Theorie oder Sprache mehr oder weniger adäquat widerspiegeln.

Der ontologische Perspektivist sieht sich aber mit einem schwerwiegenden Problem konfrontiert: Wie können diese verschiedenen Perspektiven zueinander kompatibel gemacht werden? Wie können wissenschaftliche Disziplinen miteinander kommunizieren, wenn sie mit Aufteilungen unterschiedlicher Körnigkeit oder Granularität arbeiten? Welche Wissenschaftsdisziplin kann die benötigte integrierende Plattform bereitstellen? Im Folgenden versuche ich zu zeigen, dass eine philosophisch fundierte und auf rigorosen formalen

Prinzipien aufgebaute Ontologie für die angestrebte Integration durch nichts zu ersetzen ist. Meine Aufgabe ist dementsprechend vor allem pragmatischer Natur: Selbst wenn man zum Schluss gelangt, dass eine rigorose formale Ontologie sowohl prinzipiell und als auch z. B. unter Berücksichtigung des ständigen Erkenntnisgewinns in der Naturwissenschaft die Wirklichkeit immer nur unvollständig wiedergeben kann, ist eine solche Ontologie doch zumindest für die aus Sicht der aktuellen Biologie und Biomedizin unverzichtbare Datenintegration notwendig.

(4) Die modulare Struktur des Biologischen

Zu den für uns wichtigen Perspektiven gehören vor allem jene, die wissenschaftliche Erklärungen stützen. Dies sind häufig Perspektiven einer feinen Körnigkeit, die uns z. B. über die Anzahl und Anordnung der Gene auf einem Chromosom oder über die chemischen Reaktionen innerhalb eines Reaktionspfads Aufschluss geben. Wenn allerdings die wissenschaftliche Erfassung dieser Strukturen für die Zwecke der Medizin eine Bedeutung haben soll, muss sie durch Aufteilungen einer ganz anderen, grobkörnigeren Art ergänzt werden; nämlich um Aufteilungen auf der Ebene der alltäglichen Erfahrung, wie beispielsweise Aufteilungen der Krankheiten und ihrer Symptome, der menschlichen Verhaltensweisen und der dafür relevanten Umwelten.

Wie schon Leibniz konstatierte, erweisen sich die mit dem bloßen Auge wahrnehmbaren Entitäten der natürlichen Welt bei näherem Hinsehen als etwas Zusammengesetztes: Beispielsweise ist ein Embryo eine hierarchische Verschachtelung von Organen, Zellen, Molekülen, Atomen und subatomaren Teilen. „Eine Einheit im Mittelbereich einer solchen verschachtelten Struktur ist gleichzeitig umschlossen und umschließend, Ganzes und Teil, Entität und Umwelt. Ein Organ, z. B. die Leber, ist ein *Ganzes* hinsichtlich der sie konstituierenden Zellverbände und ein *Teil* in bezug auf den umliegenden Organismus, den es mit anderen Organen bildet. Ein Organ bildet die Umgebung seiner Zellen und ist selbst wiederum von einem Organismus umgeben.“⁷

Die biologische Wirklichkeit erscheint als eine komplexe Hierarchie ineinander verschachtelter Ebenen. Moleküle sind Teil von Molekülsammlungen, die wir „Zellen“ nennen; Zellen sind in Blätter eingebettet, Blätter in Bäume, Bäume in Wälder und so weiter. Wie unsere Wahrnehmung und unser Handeln mehr oder weniger perfekt auf die Ebene der alltäglichen Erfahrung abgestimmt ist,⁸ so sind auch die verschiedenen biologischen Wissenschaften auf verschiedene Ebenen innerhalb dieser komplexen Hierarchie abgestimmt. Es gibt z. B. nicht nur die klinische Physiologie, sondern auch die Zell- und Molekularphysiologie. Neben der makroskopischen Anatomie mit ihren Nebenzweigen, wie klinischer, chirurgischer und radiologischer Anatomie, gibt es auch die mikroskopische Anatomie mit Unterdisziplinen, wie Histologie und Cytologie.

Der ontologische Perspektivismus soll einen metaphysischen Rahmen bieten, in den die Gegenstandsbereiche der verschiedenen Disziplinen eingeordnet werden können, so dass diese Bereiche in einem synoptischen Metasystem sowohl miteinander als auch mit einer Ontologie auf der Ebene der grobkörnigen Gegenstände und Prozesse der menschlichen Lebenswelt verbunden werden können.

⁷ Roger G. Barker, *Ecological Psychology. Concepts and Methods for Studying the Environment of Human Behavior*, Stanford 1968, S. 154.

⁸ J. J. Gibson, *The Ecological Approach to Visual Perception*, Boston, 1979.

(5) Kommunizieren zwischen Perspektiven

Die zentrale Frage bleibt damit: Wie verhalten sich die grobkörnigen Teile und Strukturen der Wirklichkeit, auf die unser unmittelbares Wahrnehmen und Handeln gerichtet ist, zu solchen Teilen, Dimensionen und Strukturen der Wirklichkeit, die uns durch die speziellen Erfassungsmethoden der verschiedenen Wissenschaften zugänglich gemacht werden? Diese Frage erinnert an das durch Kant inspirierte Projekt des amerikanischen Philosophen Wilfrid Sellars zur Entwicklung einer sogenannten „stereoskopischen Betrachtungsweise“, die sowohl die Inhalte unseres alltäglichen Denkens und Sprechens als auch die maßgeblichen Erkenntnisse der Naturwissenschaften in einer einzigen synoptischen Auffassung von Personen und Welt zusammenführen sollte.⁹ Diese stereoskopische Betrachtungsweise sollte sowohl dem modernen „scientific image“ wie auch dem „manifest image“ des gesunden Menschenverstands gerecht werden und die Kommunikation zwischen ihnen ermöglichen.

Welche ist die wahre Sonne? Die des Bauern oder die des Astronomen? Dem ontologischen Perspektivismus zufolge müssen wir uns nicht entweder für die eine oder die andere Sicht der Dinge entscheiden, da sowohl Alltagswissen als auch die etablierten Wissenschaften auf Aufteilungen beruhen, die wir, unter Beachtung ihrer jeweiligen Funktion, gleichzeitig akzeptieren können.

Der kommunikative Bezugsrahmen, den wir benötigen, um zwischen diesen verschiedenen Aufteilungen navigieren zu können, soll zugleich die theoretische Grundlage für die Behandlung eines der wichtigsten Probleme der heutigen Biomedizin liefern: Wie integrieren wir das Wissen, das wir über Gegenstände und Prozesse auf der Ebene von Genen und Proteinen erlangt haben, mit unserem Wissen von Krankheiten und Verhaltensweisen von Menschen bis hin zu möglichen zukünftigen gentechnologischen Pharmazeutika und Therapien? Selbstverständlich können wir diese Frage hier nicht beantworten. Wir wollen aber zeigen, dass ein solcher Integrationsrahmen vor dem Hintergrund entwickelt werden kann, dass Biologie und Bioinformatik in den letzten Jahren bestimmte theoretische und methodologische Elemente der philosophischen Ontologie wieder aktuell werden ließen, die sich paradigmatisch um die (von Aristoteles initiierte) Lehre von hierarchischen Taxonomien drehen.

Gerade für die Biologie und die Bioinformatik haben philosophische Ideen zu Kategorien und Taxonomien – und, wie wir sehen werden, zu einer Reihe anderer traditioneller philosophischer Begriffe – eine neue Aktualität gewonnen. Die verschiedenen Zweige der wissenschaftlichen Biologie verwenden nach wie vor taxonomische Hierarchien als Fundament ihrer Forschung. [This is a repeat.] Ihre Taxonomien beziehen sich nicht nur auf Gattungen und Arten von Organismen, sondern schließen auch genomische und proteomische Taxonomien, sowie Taxonomien von Zellen und Zellkonstituenten, einzelnen biochemischen Reaktionen und ganzen Reaktionsketten ein. Gleichzeitig werden diese Taxonomien in Form ungeheuer großer Datenbanken wie die Gene Ontology (GO), Swiss-Prot, SCOP oder die Protein Data Bank (PDB) zum Instrument einer neuen Form der biologischen Forschung.¹⁰ Damit können neben praktischen Experimenten auch die Datenverarbeitung und die Informationsextraktion zu neuen wissenschaftlichen Ergebnissen führen. Die sinnvolle und fruchtbare Anwendung dieser neuen Techniken setzt aber voraus, dass das Problem der Kommunikation zwischen verschiedenen Kategoriensystemen gelöst wird.

Wir vertreten die Hypothese, dass uns der Einsatz der neuen Methoden der angewandten Ontologie einer Lösung dieses Problems näher bringt und dass es dadurch möglich ist, eine echte Zusammenarbeit zu schaffen – von Biologen, Informatikern und den zwischen ihnen vermittelnden Philosophen.

⁹ Wilfrid F. Sellars, “Philosophy and the Scientific Image of Man”, in Sellars: *Science, Perception and Reality*, London 1963.

¹⁰ <http://www.cs.man.ac.uk/~stevensr/ontology.html>

(6) Ontologie und Biomedizin

Mit der Ausweitung der elektronischen Informationsverarbeitung und der enormen Zunahme der Informationsvolumens und der Informationsarten [What are these?] wachsen die Hürden, die die Kommunikation überwinden muss, um zwischen Informationssystemen zu vermitteln, die auf unterschiedlichen Begriffssystemen beruhen. Die informationstechnische Unterstützung der biomedizinischen Forschung bietet dafür augenfällige Beispiele: von der Entschlüsselung des Humangenoms über Genexpressionsstudien bis hin zu einem besseren Verständnis von Proteinstrukturen – alles Resultate des Strebens nach einem besseren Verständnis der Rolle erblicher Faktoren und der Umwelt für die Gesundheit und den Verlauf von Krankheiten des Menschen beziehungsweise der Suche nach neuen Pharmazeutika.

Die heutige Bioinformatik ist mehr als gut gerüstet, jene rechenintensiven Bereiche der biomedizinischen Forschung zu unterstützen, die beispielsweise mittels statistisch basierter Mustererkennungsmethoden *quantitative* Korrelationen untersuchen. Dagegen wird von verschiedenen Seiten¹¹ eine geeignete *qualitative* Grundlage für diese Forschung vermisst. Um biologisch relevante quantitative Korrelationen erkennen und die dadurch gewonnenen Informationen sinnvoll verarbeiten zu können, sind entsprechende Begriffe und Kategorien notwendig, die im Kontext einer qualitativen Theorie zu behandeln sind – einer synoptischen Theorie, die unter anderem mit sehr allgemeinen Begriffen wie Art, Spezies, Teil, Ganzes, Funktion und Prozess arbeiten müsste.

Biologen selbst haben mitunter ein etwas vages Verständnis von der Bedeutung dieser Termini, die aber für die Bedürfnisse ihrer wissenschaftlichen Arbeit völlig ausreicht. Missverständnisse können in der Kommunikation zwischen Wissenschaftlern unter anderem dadurch vermieden werden, dass relevante Gegenstände und Prozesse im Labor zur Veranschaulichung genutzt werden. Die informationstechnische Verarbeitung verlangt aber nach expliziten formalen Definitionen. Solche Definitionen wiederum sind nur durch eine umfassende formale Theorie der entsprechenden Kategorien und Relationen verfügbar zu machen. Wie wir gesehen haben, wird eine solch umfassende Theorie in der Informatik eine „Ontologie“ genannt. Das Entwickeln der benötigten Ontologie bedeutet, dass die im Laufe der Zeit hervorgebrachte Menge an Terminologie- und Kategoriensystemen in ein einziges übersichtliches Ganzes zu verschmelzen ist. In den 90er Jahren waren umfangreiche Bemühungen zur Vereinheitlichung der Fachtermini darauf gerichtet, sie letztlich auf eine solche einheitliche Basis zu stellen. Die biomedizinische Informatik brachte Ansätze wie MeSH, die NCBI Taxonomy, SNOMED-CT oder die Gene Ontology sowie – als übergreifende Integrationsplattform – das Unified Medical Language System (UMLS) hervor.¹² Gestützt auf diese Systeme wurden die jeweiligen Gebiete nach und nach durch robuste und allgemein akzeptierte *kontrollierte Vokabulare* indiziert und von Experten in einer Weise annotiert, die auf Dauer die Kompatibilität und Wiederverwendbarkeit elektronisch gespeicherter Information sichern sollte. Diese kontrollierten Vokabulare trugen viel dazu bei, eine neue Phase der terminologischen Präzision und Ordnung in der biomedizinischen Forschung einzuläuten, so dass die erhoffte Integration biologischer Informationen greifbar schien.

¹¹ Vgl. z. B. Christos A. Ouzounis, u. a., “Classification Schemes for Protein Structure and Function”, in: *Nature Reviews Genetics*, 4 (2003), 508-19; Cornelius Rosse und José L. V. Mejino, “A Reference Ontology for Bioinformatics: The Foundational Model of Anatomy”, in: *Journal of Biomedical Informatics*, 36, (2003), 478-500.

¹² <http://www.nlm.nih.gov/research/umls>.

Diese Anstrengungen beschränkten sich aber zu sehr auf die Terminologien und die mit Terminologien arbeitenden Verfahren des Computers. Die bloße syntaktische Exaktheit der Termini, wie sie in kontrollierten Vokabularen gesammelt und geordnet sind, stand zu sehr im Vordergrund; die eigentliche ontologische, d. h. begriffliche und semantische, Klärung wurde vernachlässigt.

Definitionen wurden zwar mit Termini assoziiert. Diese stammten aber aus den medizinischen Wörterbüchern einer früheren Zeit – sie wurden für Menschen, nicht für Computer aufbereitet. Sie sind daher informeller Natur, oft zirkulär und inkonsistent. Die existierenden Terminologiesysteme beruhen daher in aller Regel auf ungenau formulierten Begriffen und unscharfen Regeln zur Klassifikation. Diese Regeln liefern bei der Anwendung durch Menschen mit entsprechendem praktischem Wissen akzeptable Ergebnisse (für wiederum entsprechend erfahrene Kollegen); sie sorgen aber zunehmend für Schwierigkeiten bei der elektronischen Verarbeitung – sofern sie sich überhaupt dafür eignen. Dadurch aber geht entscheidendes Potential verloren. Denn logisch strukturierte Definitionen sind notwendige Voraussetzung für das konsistente (und „intelligente“) Navigieren zwischen Informationsinhalten durch automatische Deduktions-Systeme. Während entsprechend qualifizierte, interessierte und motivierte Menschen mit ungenau gefassten Inhalten umgehen können, sind für elektronische Informationsverarbeitungssysteme exakte und gut strukturierte Definitionen unverzichtbar.¹³

Die Zusammenarbeit zwischen Informatikern und Biologen ist all zu oft durch eine Variante der obersten Direktive in *Star Trek* geprägt: *Die Sternenflotte soll sich nicht in fremde Kulturen einmischen*. Die „fremden Kulturen“ sind in unserem Zusammenhang die verschiedenen Zweige der Biologie. „Nicht einmischen“ heißt, dass die Informatiker die durch die Biologen bereitgestellten Informationsressourcen als etwas Vorgegebenes und Unantastbares akzeptieren sollen und beispielsweise Anwendungen entwickeln, die eine Navigation durch diese Informationen ermöglichen. Es steht den Informatikern sozusagen nicht zu, die Biologen zur Bereinigung fehlerhafter Konstruktionen zu ermuntern, obwohl der Nutzen der Informationsressourcen wachsen würde, wenn ihr Aufbau bestimmten logischen Grundprinzipien folgte.

(7) Die Rolle der Philosophie

Bislang ist es den Bio- und Medizininformatikern selbst nicht gelungen, ein ontologisch fundiertes Instrumentarium für die Integration ihrer Daten bereitzustellen. Bisherige Ansätze, z. B. das „Semantic Network“ des UMLS,¹⁴ lassen immer deutlicher Probleme zutage treten, die aus der Vernachlässigung in der Philosophie, der Logik und speziell in der Definitionstheorie allgemein anerkannter Prinzipien für die Entwicklung ontologischer Theorien erwachsen.¹⁵ Termini werden mit Konzepten verwechselt, diese wiederum mit den durch die Wörter bezeichneten Sachen selbst bzw. mit den Prozeduren, die wir anwenden, um Wissen über diese Sachen zu erhalten. Blutdruck wird z. B. mit dem Messen von Blutdruck identifiziert. Körperliche Systeme, wie das Kreislaufsystem, werden als „begriffliche Entitäten“ klassifiziert, ihre Teile aber, zum Beispiel das Herz, als „physikalische Entitäten“. Grundlegende philosophische Unterschiede finden keine angemessene Beachtung: So fehlt ein klares Verständnis des Verhältnisses zwischen Funktion und Prozess; beide werden etwa

¹³ Barry Smith, Jakob Köhler und Anand Kumar, „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)*, Berlin 2004, 79–94.

¹⁴ Alexa McCray, „An Upper Level Ontology for the Biomedical Domain“, in: *Comparative and Functional Genomics*, 4 (2003), 80–84.

¹⁵ Barry Smith, „Beyond Concepts“, im Druck.

(z. B. im Rahmen der Gene Ontology) mit „Tätigkeiten“ gleichgesetzt – und das ohne diesen Begriff weiter zu erläutern.

Da die Kommunikation zwischen Computerprogrammen jedoch wesentlich fehleranfälliger ist als die zwischen Menschen, verlangt der Einsatz computergestützter Systeme in Biologie und Medizin zwangsläufig nach maximaler Klarheit und Genauigkeit – insbesondere der grundlegenden und in allen Systemen verwendeten Konzepte und Relationen wie *ist_Teil_von* oder *ist_lokalisiert_in*. Allein eine auf philosophischen Prinzipien beruhende ontologische Theorie der entsprechenden allgemeinen Kategorien und Relationen erscheint geeignet, diese Klarheit und Genauigkeit zu gewährleisten.

Eine solche ontologische Theorie könnte nicht nur die sinnvolle Interpretation der vom Computer gelieferten Ergebnisse, sondern darüber hinaus auch eine bessere Kommunikation zwischen Wissenschaftlern verschiedener Disziplinen ermöglichen, die mit unterschiedlichem implizitem Hintergrundwissen arbeiten und daher oft Schwierigkeiten haben, gegenseitiges Verständnis zu erreichen.

Ein Instrument für diese bessere Kommunikation ist das durch das Department of Biological Structure der University of Washington in Seattle entwickelte „Foundational Model of Anatomy“ (FMA), das im Vergleich mit sonstigen bioinformatischen Systemen neue Maßstäbe setzt. Das FMA bildet den strukturellen Aufbau des menschlichen Körpers von der makromolekularen bis zur makroskopischen Ebene ab und liefert ein robustes und konsistentes Schema für das Klassifizieren anatomischer Einheiten auf der Basis expliziter Definitionen. Dieses Schema dient zugleich als Grundlage für den „Digital Anatomist“, eine computergestützte Visualisierung des menschlichen Körpers, sowie als Muster für zukünftige Systeme, die die exakte Darstellung der Pathologie, der physiologischen Funktionen und der Genotyp-Phänotyp-Beziehungen ermöglichen sollen. Von entscheidender Bedeutung ist, dass die Anatomie-Informatik des FMA explizit auf von Aristoteles übernommenen Ideen zur richtigen Struktur von Definitionen beruht.¹⁶ Die Definition einer Klasse im FMA, z. B. eine Definition für „Herz“ oder „Organ“, bestimmt dementsprechend das gemeinsame Wesen, das die entsprechenden Instanzen teilen. Dies geschieht durch die Spezifikation (a) eines *genus*, d. h. einer umfassenderen Klasse, die die gegebene Klasse einschließt, zusammen mit (b) den *differentiae*, die diese Instanzen innerhalb der breiteren einschließenden Klasse charakterisieren. Die modulare Struktur der Definitionen des FMA erleichtern sowohl das effiziente Verarbeiten und Überprüfen auf Fehler als auch die konsistente Erweiterung des Gesamtsystems. Zugleich garantiert diese modulare Struktur, dass die Klassen des FMA einen echten Kategorienbaum im Sinne sowohl der alten aristotelischen als auch der von Aristoteles beeinflussten Linnéschen (Bio-)Ontologie bilden.

Diese Umsetzung der aristotelischen Lehre des „*definitio fit per genus proximum et differentias specificas*“ wird mit dem aktuellen Erkenntnisstand der Biologie kombiniert. Welche Klassen in der konkreten Domäne der wissenschaftlichen Anatomie vorkommen, wird mittlerweile nicht mehr auf der Basis der sichtbaren Körperteile entschieden, sondern ist ein Gegenstand empirischer Forschung im Rahmen der Genetik. Phänomenologisch erkennbare anatomische Strukturen werden erst nach dem Nachweis struktureller Gene, deren koordinierter Ausdruck (*expression*) in der Entwicklung von Organismen des jeweiligen Typs die entsprechenden Instanzen hervorruft, als echte Klassen in das FMA aufgenommen.

¹⁶ J. Michael, J., José L. V. Mejino, Cornelius Rosse, “The Role of Definitions in Biomedical Concept Representation”, in: *Proceedings of the American Medical Informatics Association Symposium*, 2001, 463-467.

(8) Die Vielfältigkeit der Lebensformen

Die rasche Entwicklung der biologischen Forschung bringt ein neues Verständnis sowohl der Vielfalt der Ausprägungen bei gleichzeitig erstaunlich geringer Variationsbreite der Grundphänomene des Lebens mit sich. Es gibt einerseits eine Vielfalt der *substantiellen Formen des Lebens*, z. B. Mitochondrien, Zellen, Organe, Organsysteme, ein- und mehrzellige Organismen, Arten, Familien, Gesellschaften, Populationen sowie Embryonen und andere Lebensformen in verschiedenen Phasen der Entwicklung. Andererseits existieren bestimmte Grundbausteine der *prozessualen Formen des Lebens* in vielfältigen Variationen, z. B. Stoffwechsel und Abwehr von Pathogenen, pränatale Entwicklung, Kindheit, Adoleszenz, Altern, Verhalten, Wachsen, Wahrnehmen, Reproduktion, Gehen, Sterben, Handeln, Kommunizieren, Lernen, Lehren, Sozialverhalten und Evolution. Schließlich finden sich bestimmte Prozesstypen wie z. B. Zellteilung und der Transport von Molekülen zwischen Zellen in jeder Phase der biologischen Entwicklung.

Um ein konsistentes System ontologischer Kategorien und damit verbundener Prinzipien zu entwickeln, das sowohl die Spezifität dieser verschiedenen Formen des Lebens als auch die verbindenden Zusammenhänge zwischen ihnen verständlich machen kann, ist eine Reihe von Fragestellungen zu betrachten. Ihre sorgfältige Behandlung ist um so wichtiger, weil die bestehenden biomedizinischen Informationssysteme sie oft ignoriert oder nur unzureichend berücksichtigt haben. Sie sind aber auch von intrinsisch philosophischem Interesse und zeigen, wie philosophische Forschung an der Grenze zwischen Informatik und empirischer Biologie eine bisher nicht beachtete praktische Relevanz gewinnen kann.¹⁷

Zur Verdeutlichung dieser benötigen wir unter anderem:

- 1) *Eine Analyse der Existenzweisen der verschiedenen Formen des Lebens in der Zeit*: Substanzen (Zellen, Organismen, ...) unterscheiden sich mit Blick auf Existenz und Zeit grundlegend von Prozessen. Substanzen existieren als Ganzes zu jedem Zeitpunkt ihrer Existenz. Sie bewahren ihre Identität über die Zeit – und dieses Bewahren der Identität, dieses Ausharren ist für die Definition von „Leben“ und „Lebewesen“ selbst von zentraler Bedeutung. Prozesse dagegen existieren in ihren zeitlichen Teilen; sie entfalten sich im Zeitablauf und sind nie in einem einzelnen Augenblick als Ganzes existent.¹⁸

Wir unterscheiden dementsprechend zwischen dem kontinuierlich Existierenden („continuants“) und dem sich in der Zeit Ereignenden („occurents“). Nicht nur die Substanzen selbst existieren kontinuierlich, sondern auch die von ihnen getragenen Zustände, Dispositionen, Funktionen und Qualitäten. All diese Entitäten stehen in besonderen Beziehungen einerseits zu ihren substanziellen Trägern und andererseits zu den entsprechenden Prozessen. So werden Funktionen generell in Prozessen *verwirklicht*. Wie zum *Organismus* ein *Leben* gehört, gehört zu einer *Disposition* die Möglichkeit ihrer *Verwirklichung* und zu einem *Zustand* (etwa einer Krankheit) sein *Verlauf*, seine *Geschichte* (die in einem Krankenblatt dargestellt werden könnte).

- 2) *Eine Analyse des Begriffs Funktion in der Biologie*: Nicht nur Gene haben für das Leben von Organismen wichtige Funktionen. Dies trifft auch auf Organe und Organsysteme sowie auf Zellen und gewisse Zellbestandteile wie Mitochondrien oder Chloroplasten zu. Eine Funktion inhäriert z. B. einem Körperteil – und sie kommt in einem Prozess des *Funktionierens* zum Tragen; so besteht die Funktion des Herzens darin, Blut zu pumpen.

¹⁷ Barry, Smith, Jenniffer Williams und Steffen Schulze-Kremer, “The Ontology of the Gene Ontology”, in: *Proceedings of the Annual Symposium of the American Medical Informatics Association*, 2003, 609–613.

¹⁸ Ingvar Johansson, *Ontological Investigations*, London, 1989; Pierre Grenon und Barry Smith, SNAP and SPAN: Towards Dynamic Geospatial Ontologies, *Spatial Cognition and Computation*, 4 (2004), 69–103.

Was aber bedeutet das Wort „Funktion“ in diesem Zusammenhang? Naturwissenschaftler wie Wissenschaftstheoretiker des 20. Jahrhunderts sparten das Thema „Funktion“ – und den Faktor Teleologie überhaupt – aus ihren Theorien aus, da es mit dem heutigen naturwissenschaftlichen Kausalitätsverständnis nicht in Einklang zu bringen schien. Funktionen sind aber für das Weltbild (die Ontologie) des klinischen Mediziners unerlässlich. Sein Leben besteht darin, die Fehlfunktionen des Körpers möglichst weitgehend wieder in Funktionen umzuwandeln.

Langsam beginnen auch Philosophen, sich Gedanken darüber zu machen, was Funktionen sind. So begreift z. B. die amerikanische Philosophin Ruth Millikan in ihrer richtungweisenden Theorie der „proper functions“¹⁹ eine organische Funktion als eine durch Evolution zustande gekommene Disposition der Entitäten eines bestimmten Typs, deren Ausübung durch die biologischen Vorgänger (zumindest teilweise) für die Existenz dieser Entität verantwortlich ist. Eine Entität besitzt eine Funktion aber nur im Rahmen eines sie umfassenden biologischen Systems – und das verlangt zwangsläufig auch nach einer Analyse des Begriffs *System*. Den existierenden philosophischen Theorien sowohl von „Funktion“ als auch von „System“ mangelt es aber an der notwendigen Präzision und Allgemeinheit.²⁰

- 3) *Eine Analyse der Bestandteile und des Aufbaus eines Lebewesens*: In welchem Verhältnis stehen ein Organismus und seine körperlichen Bestandteile? Hier begegnet uns das alte Problem von *Form* und *Stoff* in neuem Gewand als das Problem der Beziehungen zwischen dem Organismus als organisiertem Ganzen und seinen verschiedenen materiellen Trägern (Nukleinsäuren, Proteine, Lipide, Zucker usw.).

Sowohl Einzeller als auch mehrzellige Organismen zeigen einen spezifischen Aufbau, eine modulare Struktur, so dass sich Teile von Lebewesen auf verschiedenen Ebenen der Granularität erkennen lassen. Je nachdem, ob molekulare oder zelluläre Strukturen, Gewebe, Organe, Organsysteme oder komplette Organismen im Mittelpunkt der Aufmerksamkeit stehen, ergeben sich unterschiedliche Einteilungen. Weil ein Lebewesen mehr als die Summe seiner Teile ist, ist diese Pluralität von transgranularen Perspektiven für das Erfassen dieser Teile von zentraler Bedeutung. Die Erklärung dafür, wie sich die Entitäten auf diesen verschiedenen Ebenen zueinander verhalten, wurde in der Literatur häufig unter dem Stichwort „Emergenz“ behandelt, aber selten klar gefasst.

Eine gewisse Modularität und dementsprechende Grade der Granularität sind auch in der zeitlichen Dimension zu erkennen, so dass sich verschiedene Aufteilungen der prozessualen Lebensformen – z. B. einzelne chemische Reaktionen, biochemische Reaktionspfade, Lebenszyklen einzelner Organismen, Generationen, evolutionäre Epochen – in Abhängigkeit davon ergeben, ob wir unser Augenmerk auf Sekunden, Jahre oder Millennien richten.

- 4) *Eine Theorie der biologischen Arten (Typen, Universalien)*: Eine den aktuellen Bedürfnissen gerecht werdende Theorie der biologischen Arten muss selbstverständlich – auch im Gegensatz zur Kategorienlehre und Biologie des Aristoteles – Raum für die *Evolution der Arten* bieten. Die meisten zeitgenössischen Ansätze für eine solche Theorie stützen sich mit mehr oder weniger formaler Strenge auf die mathematische Mengentheorie. Arten sind aber nicht gleich der Menge ihrer Instanzen – die Identität von Mengen im mathematischen Sinn ist von deren Elementen abhängig. Die Mengen selber sind aber zugleich zeitlos. Dagegen existieren Arten in der Zeit – und sie bleiben

¹⁹ Ruth Millikan, *Language, Thought, and Other Biological Categories: New Foundations for Realism*, Cambridge, MA 1988.

²⁰ Barry Smith, Igor Papakin und Katherine Munn, „Bodily Systems and the Spatial-Functional Structure of the Human Body“, in: D. M. Pisanelli, a.a.O. (Anm 5.)

selbst dann erhalten, wenn sich die Gesamtheit ihrer Instanzen ändert. Arten haben dementsprechend gewisse Eigenschaften eines Individuums – ein Aspekt ihrer Ontologie, der in der Bioinformatik bislang zu wenig berücksichtigt wurde.²¹

Existierende bioinformatische Systeme konzentrieren sich auf allgemeine, in taxonomischen Hierarchien organisierte Begriffe und befassen sich dementsprechend mit der biologischen Wirklichkeit ausschließlich auf der Ebene der Klassen (Arten, Universalien). Individuelle Lebewesen, die in diesen Taxonomien erfassten Klassen instantiieren, sowie die zeitlichen und räumlichen Aspekte dieser Instantiierung, werden außer Acht gelassen. Dies ist teilweise darauf zurückzuführen, dass die medizinischen Terminologien unserer Tage überwiegend auf den medizinischen Wörterbüchern der Vergangenheit beruhen und die jetzigen bioinformatischen Ontologien der Wissensrepräsentation entstammen. Sowohl die Autoren von Wörterbüchern als auch die im Bereich der Wissensrepräsentation Tätigen interessieren sich prinzipiell aber nur für das Generelle. Eine adäquate Ontologie des Biologischen muss dagegen sowohl Arten als auch Individuen (Instanzen, Partikularien) berücksichtigen. Sie muss z. B. der Tatsache Rechnung tragen, dass biologische Arten immer sowohl typische Instanzen als auch die stets sich verschiebende Penumbra von Grenzfällen, deren Existenz die biologische Evolution überhaupt erst möglich macht, umfassen. Wie wir im Weiteren zeigen werden, ist die Berücksichtigung der Rolle von Instanzen bei der Strukturierung des Gebiets des Biologischen unerlässlich, wenn wir gewisse Unzulänglichkeiten der bisherigen Systeme zur Repräsentation biologischen Wissens vermeiden wollen.

- 5) *Eine Theorie der synchronen und diachronen Identität*: Synchronische Identität bezieht sich darauf, ob x dasselbe Individuum (oder Protein, Gen, dieselbe Art, oder derselbe Organismus) wie y ist; diachronische Identität darauf, ob x heute dasselbe Individuum (oder Protein, Gen, dieselbe Art, oder derselbe Organismus) ist wie x gestern oder vor tausend Jahren war. Die logische Analyse der verschiedenen Genidentitätsbegriffe durch den Gestaltpsychologen Kurt Lewin²² ist ein wichtiger Orientierungspunkt für die Arbeit zu diesem Thema. Lewin unterscheidet zwischen physikalischer, biologischer und evolutionstheoretischer Genidentität, d. h. zwischen den Modi des zeitlichen Beharrens eines Molekülkomplexes, eines Organismus oder einer Art. Auch zeitgenössischen analytischen Philosophen, wie Olson oder Wilson, ist es durch das Einbeziehen biologischer Perspektiven gelungen, alte Fragen, z. B. nach persönlicher Identität und Individuation, mit neuer ontologischer Schärfe zu behandeln.²³ Auch die Frage danach, wann im Prozess der embryonalen Entwicklung ein Mensch als Individuum zu existieren beginnt, wurde vor dem Hintergrund einer den biologischen Tatsachen adäquaten Ontologie neu untersucht.²⁴
- 6) *Eine Theorie der Umgebungen biologischer Systeme*: Gene existieren und verwirklichen sich nur in ganz bestimmten, molekularen Kontexten oder Umgebungen, ihre konkrete Realisierung ist von der Natur eben dieses Zusammenhangs abhängig. Analog leben Organismen in jeweils besonderen Nischen oder Umwelten; ihre jeweiligen Umwelt bestimmt ihre Weiterexistenz.

²¹ Vgl. die wichtigen Beiträge: David Hull, "Are Species Really Individuals?" in: *Systematic Zoology*, 25 (1976), und Michael Ghiselin, *Metaphysics and the Origin of Species*, Albany, 1997.

²² *Der Begriff Genese in Physik, Biologie und Entwicklungsgeschichte. Eine Untersuchung zur vergleichenden Wissenschaftslehre*, Berlin 1922.

²³ Vgl. Eric T Olson, *The Human Animal. Personal Identity without Psychology*, Oxford 1999, Jack Wilson, *Biological Individuality: The Identity and Persistence of Living Entities*, Cambridge 1999.

²⁴ Barry Smith and Berit Brogaard, "Sixteen Days", *The Journal of Medicine and Philosophy*, 28 (2003), 45–78.

Die philosophische Literatur seit Aristoteles hat aber zur Ontologie der Umwelten wenig beigetragen; sie hat in der Regel den Substanzen und deren Akzidenzien (Qualitäten, Eigenschaften) viel mehr Bedeutung beigemessen als den diese Substanzen beherbergenden Milieus. Was aber *sind* Nischen oder Umwelten, und wie sind die Abhängigkeitsverhältnisse zwischen Lebewesen und Umwelt ontologisch zu begreifen?²⁵ Die Relevanz dieser Fragestellung liegt nicht nur für die Entwicklungsbiologie, sondern auch etwa für Ökologie und Umweltethik auf der Hand.

(9) Fallstudie: Zur Analyse der bioinformatischen Klassen- und Teil-Begriffe

Am Beispiel der Analyse zweier Relationen, nämlich der Klassenzugehörigkeit (*is_a*) und Teil_von (*part_of*), die in Ausdrücken wie „*Mensch is_a Säugetier*“ und „*Herz part_of Mensch*“ zur Anwendung kommen, sollen die zu lösenden Probleme vorgestellt werden. Diesen zwei Relationen wird in der aktuellen Bioinformatik sowie generell in der Literatur zur künstlichen Intelligenz und Wissensrepräsentation eine zentrale Rolle zugewiesen.

Die großen für die klinische Medizin entwickelten Terminologie-Systeme sowie die jetzt im biologischen (genomischen, proteomischen) Bereich entstehenden Datenbanken müssen aufeinander abgestimmt werden. Dieses Abstimmen kann nur gelingen, wenn die zwei genannten ontologischen Relationen in den verschiedenen Systemen kohärent behandelt werden. Aneinander angrenzende und sich überschneidende Ontologie-Systeme können nur dann automatisch miteinander verknüpft werden, wenn gleiche Termini mit gleichen Begriffe assoziiert sind. Aber schon mit Blick auf diese ontologischen Grundrelationen, die intern für die Konsistenz und Stabilität dieser Systeme von tragender Bedeutung sind, ist offenkundig, dass gleiche Ausdrücke nicht zwangsläufig für äquivalente Beziehungen stehen. Teilweise wird zwischen diesen beiden Relationen überhaupt nicht unterschieden; beispielsweise gilt in UMLS „*Pflanzenblätter is_a Pflanze*“, in SNOMED „*beide Utera is_a Uterus*“.

Wir verwenden das Wort „Entität“ im Folgenden als universalen ontologischen Terminus, der alles einschließt, was in irgendeiner Weise existiert (alle Objekte, Prozesse, Funktionen, Strukturen, Zeiten, Orte usw.).²⁶ Entitäten lassen sich ganz allgemein in *Klassen* und *Instanzen* unterscheiden. Instanzen sind (sozusagen) natürliche Individuen; sie sind örtlich und zeitlich gebunden, sie existieren nur einmal in einer (in der Regel topologisch zusammenhängenden) Region von Raum und Zeit. Während Instanzen konkret sind, sind Klassen abstrakt. Sie existieren mehrfach – nämlich in ihren jeweiligen Instanzen.

Um diese Begriffe zu präzisieren, führen wir zwei primitive Relationen ein: *inst* und *part*. *Inst* ist das Verhältnis der Instantiierung einer Klasse durch eine Instanz; z. B.: „*Bruno ist eine Instanz von Mensch*“. *Part* ist die Teil-Ganzes-Relation zwischen Individuen; z. B.: „*Brunos Kopf ist ein Teil von Brunos Körper*“.

Für die Relation *inst* gelten insbesondere folgende Axiome:

- (1) Sie bezieht sich immer auf eine Instanz und eine Klasse in dieser Reihenfolge.
- (2) Keine Entität kann zugleich Instanz und Klasse sein.

Für die Relation *part* gelten folgende Axiome:²⁷

- (1) Sie ist *irreflexiv*. (Kein Individuum ist ein Teil von sich selbst.)
- (2) Sie ist *asymmetrisch*. (Wenn *x* in der *part*-Relation zu *y* steht, dann steht *y* nicht in der

²⁵ Barry Smith und Achille C. Varzi, „The Niche“, *Noûs*, 33 (1999), 198-222 und „Surrounding Space: The Ontology of Organism-Environment Relations“, in: *Theorie in den Biowissenschaften*, 121 (2002), 139-162.

²⁶ Vgl. Barry Smith, „The Logic of Biological Classification and the Foundations of Biomedical Ontologies“, in: Dag Westerstahl (Hg.), *Invited Papers from the 10th International Conference in Logic Methodology and Philosophy of Science*, Oviedo, Spain, 2003 (im Druck).

²⁷ Peter M. Simons, *Parts. An Essay in Ontology*. Oxford 1987.

part-Relation zu x .)

- (3) Sie ist *transitiv*. (Wenn x zu y und y zu z jeweils in der *part*-Relation stehen, dann steht x in der *part*-Relation zu z .)
- (4) Es gilt ein Summierungsprinzip, das die Existenz gewisser Typen von Summen oder Ganzheiten garantiert.
- (5) Es gilt ein Differenzierungsprinzip: Wenn x in der *part*-Relation zu y steht, dann existiert ein weiterer Teil z von y , der mit x keine gemeinsamen Teile hat.

Auf dieser Grundlage ist die Klasseneinordnung unmittelbar wie folgt zu definieren:

D1 $A \text{ is_a } B \stackrel{\text{def}}{=} \text{Für jedes Individuum } x, \text{ das in der Relation } \mathit{inst} \text{ zu } A \text{ steht, gilt, dass } x \text{ in derselben Relation auch zu } B \text{ steht.}$

Mit anderen Worten: Alle Instanzen von A sind zugleich Instanzen von B .

Die Definition von *part_of* als eine Relation zwischen Klassen ist aufwändiger. Wir definieren zunächst:²⁸

D2 $A \text{ part_for } B \stackrel{\text{def}}{=} \text{Für jedes Individuum } x, \text{ das in der Relation } \mathit{inst} \text{ zu } A \text{ steht, gibt es ein Individuum } y \text{ so, dass } y \text{ in } \mathit{inst} \text{ zu } B \text{ und } x \text{ in der } \mathit{part}\text{-Relation zu } y \text{ steht.}$

Diese Definition beschränkt die A s. Sie besagt, dass A s ausschließlich als Teile von B s existieren.

D3 $B \text{ has_part } A \stackrel{\text{def}}{=} \text{Für jedes Individuum } y, \text{ das in der Relation } \mathit{inst} \text{ zu } B \text{ steht, gibt es ein Individuum } x, \text{ das in } \mathit{inst} \text{ zu } A \text{ und in der } \mathit{part}\text{-Relation zu } y \text{ steht.}$

Diese Definition beschränkt die B s. Sie besagt, dass jedes B mindestens ein A enthält.

Wir definieren nun eine restriktivere *part_of* Relation, indem wir D2 und D3 kombinieren:

D4 $A \text{ part_of } B \stackrel{\text{def}}{=} A \text{ part_for } B \text{ und } B \text{ has_part } A.$

$A \text{ part_of } B$ gilt also genau dann wenn, (a) es für eine beliebige Instanz x von A immer irgendeine Instanz y von B so gibt, dass x zu y in der *part* Relation steht, *und umgekehrt*: (b) es für eine beliebige Instanz y von B immer irgendeine Instanz x von A so gibt, dass x zu y in derselben Relation steht.

Um die Bedeutung dieser Erkenntnis für die aktuelle bioinformatische Forschung anzudeuten, merken wir nur an, dass eine Unterscheidung analog zu der zwischen D2, D3 und D4 nicht nur für die Definition von *part_of*, sondern auch für viele andere in der Strukturierung biomedizinischer Ontologien verwendete Grundrelationen unerlässlich ist. Betrachten wir z. B. die Behandlung der Kausalitätsrelation im UMLS Semantic Network.²⁹ Dort finden wir unsinnige und nicht weiter kommentierte Behauptungen, wie z. B.:

Antibiotic causes Experimental Model of Disease

Biomedical or Dental Material causes Mental or Behavioral Dysfunction

²⁸ Vgl. Stefan Schulz und Udo Hahn, "Necessary Parts and Wholes in Bioontologies", in: *Principles of Knowledge Representation and Reasoning. Proceedings of KR 2002*, 387-394.

²⁹ <http://www.nlm.nih.gov/research/umls/META3.HTML>.

Manufactured Object causes Disease or Syndrome

Vitamin causes Injury or Poisoning

Bacterium causes Pathologic Function

die nahe legen, dass es den Autoren dieses Netzwerks an einer klaren Vorstellung von den verwendeten Grundrelationen mangelt (sofern es sich denn nicht um Flüchtigkeitsfehler handelt).

Durch D4 wird eine starke strukturelle Bindung zwischen den Klassen *A* und *B* gesichert: Alle *As* existieren ausschließlich als Teile von *Bs*; alle *Bs* enthalten mindestens ein *A*.

Leider werden in der gängigen Literatur zur Wissensrepräsentation³⁰ sowohl *is_a* als auch *part_of* ohne weitere Spezifizierung als einfache Relationen zwischen Klassen eingeführt, ohne der Tatsache Rechnung zu tragen, dass es oft nicht trivial ist, für allgemeine Termini festzustellen, ob besagte Relationen zwischen den entsprechenden Klassen tatsächlich bestehen. Weil es sowohl weibliche als auch männliche Menschen gibt, können wir wohl behaupten, dass „*menschlicher Hoden part_for Mensch*“, nicht aber, dass „*Mensch has_part menschlicher Hoden*“ gilt. Weil es nicht-menschliche Säugetiere gibt, die auch Herzen haben, können wir behaupten: „*Mensch has_part Herz*“, nicht aber: „*Herz part_for Mensch*“.

Gilt *Erwachsener is_a Kind*, weil jede Instanz der Klasse *Erwachsene* die Klasse *Kind* (am Anfang seiner Existenz) instanziiert hat? Gilt *Säugetier is_a Tier_das_weniger_als_2000_kg_wiegt*? Um diese Fragen beantworten zu können, müssen wir in der Definition D1 einen zeitlichen Faktor einführen, etwa in:

D1* $A \text{ is_a } B =_{\text{def}}$	Für jedes Individuum x und jeden Zeitpunkt t gilt: Wenn x zum Zeitpunkt t in der Relation <i>inst</i> zu A steht, dann steht x zum Zeitpunkt t in derselben Relation auch zu B .
---	--

Um Fälle wie *Alle_Tiere_die_dem_Kaiser_gehören is_a Säugetier* auszuschließen, muss in allen angeführten Definitionen weiter explizit zum Ausdruck kommen, dass sie nicht zufällige, sondern gesetzmäßige Verhältnisse zwischen biologischen Klassen wiedergeben sollen. Auch hier sind die Philosophen (bzw. Logiker) gefordert, die entsprechenden Instrumentarien zu entwickeln.

Insbesondere *part_of* darf nicht als triviales Inklusionsverhältnis auf der Ebene von Klassen gedeutet werden. Vielmehr folgt aus der formalen Struktur der Definition D4 unmittelbar, dass *part_of* eine in D2 und D3 erfasste, tieferliegende Beziehung der involvierten Instanzen zugrunde liegt

(10) Granularität und Idealisierung

Ein konsistentes System von Definitionen und Axiomen, das sich für das Abstimmen der verschiedenen biomedizinischen Ontologien eignet, muss auch die Relationen zwischen den Begriffen *Klasse*, *Ganzes* und *Menge* klarstellen. Auch hier ist der Verweis auf Instanzen unvermeidbar. Denn für Klassen (wie Mengen) gilt, dass ihre Instanzen unter Berücksichtigung des Faktors Körnigkeit erfasst werden. Das heißt, sie unterteilen die entsprechende Domäne in Einheiten oder Glieder („Elemente“), deren innere Teile und inneren Strukturen quasi verdeckt sind. Eine Summe ist nicht granulös in diesem Sinn. Denn die Summe der Menschen erfasst auch alle *Menschenteile*; einschließlich der Organe, Zellen, Moleküle usw. Die Klasse der Menschen dagegen erfasst nur Menschen als jeweiliges

³⁰ So in der einflussreichen lexikalischen Datenbank WordNet. Vgl. C. Fellbaum (Hg.), *WordNet. An Electronic Lexical Database*, Cambridge, MA: 1998.

Ganzes.

Biologische Klassen unterscheiden sich allerdings von Mengen dadurch, dass sie genau wie ihre Instanzen *in der Zeit* existieren. Die Summe der Instanzen einer gegebenen Klasse verändert sich zwar mit dem Zeitablauf, doch die Klasse selbst muss von dieser Änderung nicht betroffen sein. Eine Klasse ist dementsprechend nicht allein durch ihre Instanzen determiniert, genau so wie ein Staat nicht vollständig durch die Gesamtheit seiner jeweiligen Bürger bestimmt ist. Biologische Klassen erfahren eine ständige biologische Erneuerung, wenn sie den Umschlag all ihrer Instanzen überleben. Einzelne Lebewesen sind gleichfalls einer ständigen biologischen Erneuerung unterworfen, solange sie die Erneuerung der sie konstituierenden Zellen oder Moleküle überleben.

(11) Die Gene Ontology

Verdeutlichen wir uns den Nutzen der von uns angestrebten philosophischen Begriffsklärung am Beispiel der Gene Ontology™ (<http://www.geneontology.org>).

Das Human Genome Project und andere Projekte zur Entschlüsselung des Erbguts verschiedener Arten erfordern die Zusammenarbeit zwischen Biochemie und Genetik, die aufgrund der großen Datenvolumina in erheblichem Maße durch automatisierten Datenaustausch unterstützt werden muss. Die von Biologen entwickelte und äußerst erfolgreiche Gene Ontology (GO)³¹ soll die Grundlage für diesen automatisierten Austausch bieten und gleichzeitig einen Beitrag zur Integration der gesamten Biologie im Sinne einer verstärkt auf hierarchische Taxonomien und Datenbanken zurückgreifenden Wissenschaft leisten. Sie beruht nicht nur auf Daten zum menschlichen Erbgut, sondern auch auf Untersuchungen des Erbguts anderer Spezies, vor allem der Fruchtfliege, der Hefe, des Zebrafisches und der Maus.

Die Termini und Definitionen der GO sollen so weit wie möglich von einzelnen Spezies unabhängig sein. Das Interesse gilt vor allem den zentralen biologischen Kategorien, d. h. jenen genetischen und biochemischen Strukturen, Funktionen und Prozessen – wie etwa *Zelle* oder *Sterben* – die allen Lebewesen gemein sind. Es ist eine nicht triviale Leistung der GO, dass sie das Augenmerk systematisch auf solche high-level Kategorien der biologischen Welt gerichtet hat, die trotz der Evolution konstant geblieben sind. Die Bedeutung der systematischen Behandlung dieser Elemente im Rahmen einer umfassenden Kategorisierung für die heutige Bioinformatik zeigt nicht zuletzt, dass gewisse, längst für überholt betrachtete Elemente der Methodologie der aristotelischen Metaphysik auch in der postdarwinschen Zeit für die Wissenschaft von praktischer Bedeutung sind.

Die GO wurde 1998 konzipiert und kann bereits auf große Erfolge verweisen. Andere, denselben Prinzipien gehorchende Projekte folgten; z. B. die Fungal Ontology, die Plant Ontology und die Disease Ontology. Inzwischen wurde das „Open Biological Ontologies Consortium“ (OBO)³² als Dachorganisation eingerichtet, die sich der Standardisierung und Weiterentwicklung von Ontologien auf der Basis der GO-Methodologie widmet.

Die GO umfasst drei kontrollierte Vokabulare bzw. Ontologien – *Cellular Component*, *Biological Process* und *Molecular Function* – mit insgesamt circa 19.000 häufig verwendeten biologischen Termini. Leider sind sowohl die Grundrelationen von GO als auch die Relationen zwischen diesen drei Vokabularen nicht klar spezifiziert, so dass bestimmte charakteristische Kodierungsfehler auftreten.³³

³¹ <http://www.geneontology.org>

³² <http://obo.sourceforge.net/>

³³ Smith, Williams, Schulze-Kremer, a.a.O. (Anm. 16), Smith, Köhler, Kumar, a.a.O. (Anm. 12)

Die Gene Ontology selbst ist keine Integration von Datenbanken, sondern eher ein zentrales Wörterbuch für das Indizieren und das Annotieren von Genen und Genprodukten. Mittlerweile sind mehrere mächtige Werkzeuge für das Durchsuchen und Manipulieren des GO-Wortschatzes wie AmiGO und QuickGO verfügbar, die aber alle mit Blick auf die ursprüngliche Struktur von GO der oben erwähnten „obersten Direktive“ gehorchen. Darüber hinaus existieren Instrumente wie GOAT³⁴ für das Finden von Informationen über Gene und Genprodukte, die mit GO-Termini annotiert wurden – Informationen, die nicht nur für das theoretische Verständnis biologischer Prozesse, sondern auch für die klinische Medizin oder die Pharmakologie praktisch relevant sein können.

Es ist insbesondere die hierarchische Organisation der drei Vokabulare der GO, die durch schwerwiegende Ungereimtheiten geprägt ist, denn die Struktur wurde übereilt mit Hilfe der beiden Relationen *is_a* und *part_of* definiert, ohne den damit verbundenen, oben dargestellten Problemen Rechnung zu tragen. Dabei erwachsen hinsichtlich *is_a* vor allem daraus Probleme, dass andere Relationen als *is_a* Relationen kodiert werden. Dies gilt sowohl für ontologische Relationen, wie *ist_lokalisiert_in*, als auch für epistemische Relationen, die sich in Ausdrücken wie *is_a unbekannte Molekularfunktion* niederschlagen.

Die Verwendung von *part_of* ist wenigstens in dreierlei Hinsicht problematisch:

- Erstens wird *part_of* in Sätzen wie „Cell Component *part_of* Gene Ontology“ verwendet, also um die Inklusionsrelation zwischen den Vokabularen selbst abzubilden. Dies ist aber eigentlich keine Teil-Ganzes-Relation zwischen Klassen, sondern ein Inklusionsverhältnis zwischen Listen von Termini. Dies ist eine klassische Verwechslung des Verwendens mit dem Bezeichnen: Ein GO-Terminus bezeichnet sowohl eine bestimmte Klasse biologischer Phänomene als auch sich selbst.
- Zum Zweiten bezeichnet *part_of* die oben in D4 definierte Beziehung: *A part_of B* heißt dementsprechend: *As* kommen nur in *Bs* vor, z. B. in *Nukleus part_of Zelle* (ohne dass jede Zelle einen Nukleus besitzt).
- Drittens wird *part_of* für Beziehungen verwendet, die eher das aneinander-Angrenzen von Entitäten repräsentieren, z. B. in:

extrinsic to membrane part_of membrane

oder in:

synaptonemal complex part_of chromosome

Dabei ist ein synaptonemaler Komplex etwas, das *zwischen* zwei homologen Chromosomen gebildet wird.

Wir haben uns hier darauf beschränkt stellvertretend einige ausgewählte Probleme in der Handhabung bestimmter Grundrelationen der (Bio-)Ontologie zu diskutieren. Diese Beispiele erscheinen uns deshalb so bezeichnend, weil sie auf eine gewisse logische Inkonsequenz und generelle philosophische Undifferenziertheit hinweisen, die der Mehrzahl der Bemühungen innerhalb der aktuellen Bioinformatik eigen sind. Aus unserer Sicht sind die Ungereimtheiten in der Gene Ontology und ähnlichen Systemen nur durch präzise Definitionen und die damit einhergehenden philosophischen Analysen zu überwinden. Die große Tragweite der Gene Ontologie für die aktuelle Bioinformatik hätte bei Verwendung präzis definierter Grundrelationen der Informationsverarbeitung einen Quantensprung bescheren können. Wenn es sich aber in dieser Weise zeigen ließe lässt, dass mittels einer philosophisch fundierten Methodologie die Gene Ontology verbessert werden kann, wäre zu folgern, dass sich hier für das Handwerk(szeug) des Philosophen ein *praktisches* Anwendungsfeld bietet.

³⁴ <http://www.cs.man.ac.uk/~stevensr/ontology.html>